Ethnicity, Migration and Conflict: Evidence from Contemporary South Africa

Francesco Amodio Giorgio Chiovelli*

July 4, 2013

Abstract

This paper explores the extent to which changes over time in ethnic distribution correlate with contemporaneous changes in conflict incidence. We focus on the history of contemporary South Africa. Implementation and repeal of apartheid segregation laws account for cross-sectional and time variation in districts' ethnic composition. Ethnolinguistic diversity within the black majority is shown to be strongly related to the incidence of armed confrontations between black-dominated organized groups. Results are confirmed when comparing neighboring localities, and robust to the implementation of an instrumental variable identification strategy where pairwise distance between districts is used to predict location decisions of internal migrants.

Keywords: conflict, ethnicity, apartheid, South Africa.

JEL Codes: D74, J15, N47, N97, O15, R23.

*Amodio: Department of Economics and Business, Universitat Pompeu Fabra, C/ Ramon Trias Fargas 25-27, Barcelona 08005 Spain (e-mail: francesco.amodio@upf.edu); Chiovelli: Department of Economics, Università di Bologna, Piazza Scaravilli 2, Bologna 40126 Italy (e-mail: giorgio.chiovelli2@unibo.it). We are indebted to Matteo Cervellati and Albrecht Glitz for their insightful guidance. We would like to thank the following people for helpful comments and discussion: Jean-Marie Baland, Marcelo Bérgolo, Bruno Caprettini, Rodrigo Deiana, Giovanni Di Mauro, Margherita Fort, Andrea Ichino, Laura Mayoral, Stelios Michalopoulos, Hannes Mueller, Giacomo Ponzetto, Giovanni Prarolo, Marta Reynal-Querol, Simon Roberts, Martín Rossi, Uwe Sunde, Alessandro Tarozzi, Mathias Thoenig, John Tsoukalis, Fabrizio Zilibotti and all seminar participants at Università di Bologna, Universitat Pompeu Fabra, Toulouse School of Economics, Aix-Marseille School of Economics, Universidad de la República in Montevideo and Universidad Torcuato Di Tella. We also thank Lynn Woolfrey from DataFirst and Ralph Sundberg from Uppsala University for their help with South Africa Census data and UCDP-GED data respectively. Errors remain our own. Amodio acknowledges financial support from the Spanish Ministry of Economy and Competitiveness. Chiovelli acknowledges financial support from the University of Bologna.

1 Introduction

Civil conflict is closely related with poor economic performance. Understanding its determinants is regarded as crucial in the challenge for world development (Blattman and Miguel 2010). Conflict commonly manifests itself through ethnic markers. Ethnic traits stand out as a salient technology for either or both the generation and expression of social tensions. Indeed, the empirical evidence shows the probability of conflict outbreak and conflict incidence to be strongly correlated with measures of historical ethnic distribution. However, it is still unknown whether migration flows involving ethnically diverse communities have any impact on conflict prevalence. In this paper, we study the extent to which changes over time in ethnic distribution correlate with contemporaneous changes in conflict incidence.

We focus on the history of contemporary South Africa. Starting in the 1950s, apartheid segregation laws regulated the ethnic composition of local districts. The repeal of apartheid legislation in the early 1990s restored free internal mobility of blacks within the country while democratization in 1994 ended the white minority rule in favor of the newly enfranchised majority. Throughout this period, black-dominated parties and unions were struggling violently amongst themselves to benefit from the transition and dominate the new institutional scenario. We relate the changes in local districts' ethnic composition to the concurrent changes in conflict incidence and find the number of conflict events to be positively correlated with contemporaneous measures of within-black ethnolinguistic polarization[1].

For this purpose, we combine geo-referenced information on conflict with South Africa Census data. As a preliminary analysis, we exploit cross-district variation in ethnic composition before the fall of apartheid in 1991 and find a positive correlation between ethnolinguistic polarization within the black majority and the number of recorded armed confrontations between black-dominated organized groups. We thus provide a qualification

[1]We use the polarization index used in Reynal-Querol (2002) and Montalvo and Reynal-Querol (2005, 2010), which falls within the class of measures proposed by Esteban and Ray (1994). The index formula is presented and discussed in Section 4.

of these conflicts as expressed through ethnic markers. Then, we exploit the most distinguished feature of this setting: time variability in the ethnic distribution at the local district level. Mainly driven by the migration flows following the repeal of apartheid segregation laws, substantial time variation allows us to combine data from 1991 and 1996 and show the change in polarization at the district level to be positively strongly and significantly correlated with the change in the number of conflict events per year. We implement a first-difference specification which rules out unobserved differences in time-invariant determinants of conflict incidence across districts. The point estimate of the coefficient of interest is stable across specifications. Evidence suggests an increase in the within-black polarization index of one cross-district standard deviation to be associated with an increase in the number of conflict events per district of more than the 1991 national average. Moreover, conditioning on the total number of conflict events in 1989-1990 still yields highly significant results.

The relationship we find using a first-difference approach can potentially be driven by unobserved time-varying factors which affect the evolution of both ethnic composition and conflict in the same way, generating a spurious correlation between the two. We examine the severity of this issue by looking at the relationship of interest within clusters of neighboring districts. In the spirit of matching methods, we control for cluster-specific unobservable trends, and still find evidence of a positive relationship between within-black ethnolinguistic polarization and conflict incidence. We then explore the extent to which the results could be driven by the endogeneity of individuals' migration choice and location decision to the evolution of conflict. In the presence of more than two groups, mapping the outcomes of individual decision problems into systematic changes in the polarization index is theoretically challenging[2]. We thus implement an instrumental variable strategy where we use pairwise distance between districts to predict the location decision of internal migrants. In a simulation exercise, we use predicted location probabilities to re-allocate an exogenously given fraction of migrants from each ethnolinguistic group from each district to the rest of

[2]Esteban and Ray (1994) provide a series of examples to intuitively show the absence of a partial order for increasing polarization. The same arguments prevent establishing even a partial map between migration flows and changes in the polarization measure.

South Africa, independently from conflict levels and changes. The resulting predicted change in the district-level polarization index is thus used as instrument for the actual one, ruling out potential endogeneity in internal migrants' location decisions. Instrumental variable estimates confirm the existence of a strong link between within-black ethnolinguistic polarization and number of violent confrontations between black-dominated organized groups.

The paper is organized as follows. In Section 2 we relate our work with the existing literature. Section 3 provides a short overview of the history of contemporary South Africa. Section 4 introduces the relevant concepts, variables and measures, coupled with the data we use. The empirical strategy and results are presented in Section 5. Section 6 concludes.

2 Related Literature

Our study builds on the previous literature and results in the exploration of the relationship between ethnic diversity and conflict. Horowitz (1985) insightfully noticed that conflicts seem to arise in societies where a large ethnic minority faces another ethnic majority[3]. Modeling social confrontation and its relationship with measures of heterogeneity in general, Esteban and Ray (1994, 1999) argue that polarized societies, characterized by internally homogeneous but externally very distant groups, are more prone to experience conflict. In the specific case of ethnicity, ethnic lines may become salient in conflict as they allow for separation of individuals into groups that are homogenous under one trait (ethnicity), but significantly different under economic characteristics. Indeed, Esteban and Ray (2008, 2011a) explore the role of intra-group synergies in conflict effort that arise when ethnic groups are characterized by high within-group economic inequality. Caselli and Coleman (2013) rationalize instead the salience of ethnic trait in conflicts framing it as a technological device which prevents indiscriminate access to the expected gains of the winning group. Finally,

[3]Collier (2001) and Collier and Hoeffler (2004) further develop and empirically test this hypothesis. Horowitz himself extensively studied South Africa as a case study of a society divided along ethnic lines.

Esteban and Ray (2011b) develop an extensive theory of conflict which highlights the role of different distribution measures. Indeed, together with Esteban et al. (2012), they show both theoretically and empirically the informativeness of the polarization measure with regard to conflict severity to be positively related to the *degree of publicness* of the disputed *prize*. We conceptualize the struggle within the black majority in South Africa through the end of apartheid as a dispute for political power in the new institutional scenario. Such prize is intrinsically public in nature. We thus refer to these last contributions as the theoretical foundation for our focus on the polarization measure[4].

The first empirical investigation of the relationship between ethnic polarization and conflict incidence is due to Montalvo and Reynal-Querol (2005). More recently, Desmet et al. (2012) explore the explanatory power of diversity measures as computed at different levels of the ethnolinguistic world tree. Most of the existing studies exploit variation at the cross-country level and make use of time-invariant distribution indices, drawn from historical and encyclopedic sources of ethnic diversity[5]. One contribution of this paper is that we employ contemporaneous ethnolinguistic group populations from Census data in the computation of our time-varying index of ethnic polarization at the local district level. In particular, we draw information from South Africa Censuses on languages spoken and validate them using the *Ethnologue* linguistic database[6] (Lewis 2009). We thus trade the possible drawbacks from the use of potentially endogenous administrative boundaries with the advantages of disposing of a time-variant measure of ethnic polarization to employ in our analysis at the sub-national level. We also use information on districts' socio-economic characteristics, drawn from Census data, in order to show the robustness of results in reduced form.

[4]For an overview of the debate over the use of alternative ethnolinguistic distribution measures in the empirical analysis see also Alesina et al. (2003) and Desmet et al. (2009).

[5]One exception is Novta (2013), who studies theoretically and empirically the relationship between municipal-level ethnic composition and the spread of civil conflict in Bosnia.

[6]Detailed information about this procedure is available in the data section. The *Ethnologue* database has been already used as source of ethnolinguistic information in Alesina et al. (2003), Desmet et al. (2009), Desmet et al. (2012) and Esteban et al. (2012).

Finally, we contribute to a growing body of the literature that studies conflict at a more disaggregated level, such as local communities and individuals themselves. Within these, Michalopoulos and Papaioannou (2011) use information on spatial distribution of African ethnicities before colonization and find contemporary civil conflict incidence to be concentrated in the historical homeland of partitioned ethnicities. Rohner et al. (2012) use individual, county and district-level data from Uganda to investigate the social and economic consequences of ethnic conflicts. Besley and Reynal-Querol (2013) look instead into conflict persistence in Africa having fine geographical grids as units of observation. Finally, Harari and La Ferrara (2012) study instead the impact of negative climate shocks on conflict incidence using within-year variation at the local level.

3 Historical Background: South Africa and the End of Apartheid

Since the end of World War II until 1994, South Africa was ruled under *apartheid* regime. Apartheid - meaning *apartness* in Afrikaans - was a form of government based on physical separation of blacks and whites achieved through racial discrimination and political disenfranchisement of the black majority. Divisions along racial lines were thought to be the fundamental organizing principle for the allocation of all resources and opportunities, the basis of all spatial demarcation, planning and development and the boundary for all social interactions (Posel 2001). Under apartheid, economic activity was based on the exploitation of black cheap labor force in order to ensure high returns to white-owned capital investments (Clark and Worger 2011). By the same token, white workers and farmers were given protection from the competition of their black counterparts.

An extensive legislation implementing racial segregation was put in place after the election of the National Party (NP) into government in 1948. The Population Registration Act (1950) formalized the racial classification system through identifying four different races living in the South Africa. Blacks were further classified into native ethnicities on the basis of the

first language they spoke. The Bantu Authorities Act (1951) and Bantu Resettlement Act (1954) established ten black ethnicity-based *homeland* reserve areas, known as *Bantustan*: Transkei and Ciskei (Xhosa ethnicity), Bophuthatswana (Tswana), Venda (Venda), Gazankulu (Tsonga), Lebowa (Sotho), Qwaqwa (Sotho), KwaZulu (Zulu), KaNgwane (Swazi) and KwaNdebele (Sotho). The long-term objective of the law was to let the homelands become independent territories[7]. Political separation was achieved with the Promotion of Bantu Self-Government Act (1959), which implied the *de iure* disenfranchisement of blacks from white South Africa. In less than thirty years, approximately 3.5 million blacks were obliged to move to the homelands the government assigned to the different ethnolinguistic groups (Clark and Worger 2011). Figure 1 presents the map of Bantustans in South Africa as defined by the apartheid legislation as of 1986.

[Figure 1]

The apartheid system began to collapse in the mid-1980s mainly due to the internal contradictions and conflicting effects of the same implemented policies. Exploitation and control of black labor force revealed itself as costly and hardly compoundable with the enforcement of physical separation of races and ethnicities. Moreover, twentyfive countries (USA and UK among them) set a trade embargo in the late 1980s, following the 1977 UN Resolution 418 on a mandatory arms embargo. At the same time, opposition parties started to organize. Together with the black Africa National Congress (ANC) - a multiethnic party established in 1912 - and the Pan-Africanist Congress (PAC) founded in 1953, the Zulu-based Inkatha Freedom Party (IFP) was formed in 1980s and the multiracial United Democratic Front (UDF) was created in 1983. Opposition to apartheid was not homogeneous and parties were not cohesive. Black political parties engaged in conflicts for two main reasons. First, there was no agreement on how to put an end

[7]Over time, the government granted various degrees of self-government to the Bantustans, ranging from independent state-nation - as in Transkei, Bophuthatswana, Venda, and Ciskei between 1977 and 1981 - to limited self-government - as in Kwazulu, Lebowa, Gazankulu, Qwaqwa, KaNgwane and KwaNdebele.

to the apartheid experience. Second, there was no consensus on how a new post-apartheid South Africa should be ruled. Confrontation between ANC and IFP was the harshest, also because of divergent views on the future role of Bantustans and the consequences of international embargoes[8].

In the late 1980s, the government decided to repeal the Pass Law and to free ANC leaders like Nelson Mandela and other black leaders with the aim of freezing the protests and starting to negotiate. Between 1990 and 1991, the Natives' Land Act, the Population Registration Act and the Group Area Act were repealed and free mobility of blacks within South Africa was restored. Nonetheless, the negotiations taking place in 1991-1992 were everything but smooth: while Mandela and de Klerk (NP) were signing official agreements, fightings kept going on. Free elections were held in 1994, with Mandela becoming the first black president of the Republic of South Africa. The struggles between black-dominated groups weakened but continued in the following years.

4 The Data: Concepts and Measurement

The database for the analysis is built through combining several different data sources. The fundamental geographical unit for the within-country empirical analysis is the Magisterial District (MD), a sub-provincial territorial unit defined by the judicial system under the administration of the Department of Justice and Constitutional Development[9]. The map of South Africa in Figure 2 shows the district boundaries. These are still informed by the pre-1994 demarcations of the self-governing states and the Republic of South Africa territory. This makes them particularly valuable as units of comparative analysis on a small-scale geographical basis, as all other ad-

[8]It is nowadays recognized the apartheid government took advantage of divisions between black ethnicities in its struggle to retain power. Indeed, the independentist goals of the IFP and its opposition to ANC were seconded if not supported by the apartheid establishment (Clark and Worger 2011). As shown in the Online Appendix, the fundamental results from our empirical analysis are qualitatively the same when we control for measures of the intensity of governmental repression.

[9]South Africa is currently divided into 9 provinces and 354 MDs. 298 of these were surveyed in the 1991 Census of the Republic of South Africa, as the remaining ones were part of independent Bantustans.

ministrative divisions have been subject to frequent re-demarcations after democratization in 1994. Partly because of this, MDs have been used as unit of analysis in the economics and natural science literature (Case and Deaton 1999; Hoffman and Todd 2000).

[Figure 2]

4.1 Ethnic Polarization

Ethnolinguistic information is drawn from the 1991 and 1996 Census of South Africa[10] (Statistics South Africa 1991, 1998). These allow for separate identification of individuals belonging to different ethnolinguistic groups according to the first language they speak. The native African groups in the database are Swazi, Xhosa, Zulu, Sotho, Tswana, Tsonga and Venda[11]. For each MD, we count the total number of individuals in each group and compute the relative share of each group within the black majority.

Throughout the analysis, we employ the *binary* version of the polarization index implemented in Reynal-Querol (2002) and Montalvo and Reynal-Querol (2005, 2010), which falls within the class of measures proposed by Esteban and Ray (1994). The ethnolinguistic polarization within the black majority is computed as

$$ELP_{Within-Black} = 1 - \sum_{i=1}^{N} \left(\frac{1/2 - \pi_i}{1/2} \right)^2 \pi_i \qquad (1)$$

[10]Access to the 10% sample Census data for both years was kindly provided by DataFirst Research Unit at the University of Cape Town.

[11]Desmet et al. (2012) show how to make use of the genealogical relationship between world languages in the construction of distribution indexes. In line with their approach, we use information contained in the *Ethnologue* linguistic database (Lewis 2009) assuming all languages at the same level to be as equally distant from the proto-languages of their respective families. We build our measure considering black ethnolinguistc groups in South Africa which correspond to level 11 in the world language tree as reported in Figure B.1 in the Online Appendix.

where π_i is the within-black share of group i and N is the number of groups. The index value ranges between 0 and 1. It combines information on both the number of groups and their relative size, returning a distribution measure linked to the generation of social tension between equally distant groups (Esteban and Ray 1994).

4.2 Conflict

The measure of conflict incidence is derived from the Geo-referenced Event Dataset of the Uppsala Conflict Data Program (UCDP-GED v1.5)[12]. Assembled by the Department of Peace and Conflict Research at Uppsala University, it provides geo-referenced information on organized violence in Africa between 1989 and 2010, detailing different categories - state-based conflict, non-state conflict and one-sided violence[13]. The data are disaggregated spatially and temporally down to the level of individual events of fatal violence. For each conflict event, information is given on date of the event, place of the event (with coordinates), actors participating and estimates of fatalities. A conflict event is recorded in the database if it caused at least 1 death and it involved actors engaged in a nationwide conflict which caused at least 25 deaths in a year in the period (1989-2010).

We measure the yearly incidence of armed confrontations between black-dominated organized groups in South Africa at the MD level by counting the number of related geo-referenced conflict events in each MD per year.

[12]Department of Peace and Conflict Research, Uppsala University. The dataset is available at http://www.ucdp.uu.se/ged/data.php. See Melander and Sundberg (2011) for the last data presentation. See Eck (2012) for a complete discussion of the UCDP-GED database and its comparison with the Armed Conflict Location Events Dataset (ACLED). The UCPD-GED geocoding and precision is there concluded to be far superior to ACLED's and found particularly suitable for the study of conflict at the sub-national level.

[13]*State-based* conflict is defined as a contested incompatibility that concerns government and/or territory where the use of armed force between two parties, of which at least one is the government of a state, results in at least 25 battle-related deaths in a year in the period (1989-2010). *Non-state* conflict refers to the use of armed force between two organised armed groups, neither of which is the government of a state, which results in at least 25 battle-related deaths in a year in the period. *One-sided violence* refers to the use of armed force by the government of a state or by a formally organised group against civilians which results in at least 25 deaths in a year in the period (Sundberg et al. 2010).

These amount to all non-state conflict events[14].

4.3 Socio-economic and Geographical Controls

We aggregate further information for the surveyed territories at the MD level from Census 1991 and 1996, checking for consistency across waves. Moreover, we use data on population, rural population, number of individuals reporting no education, number of unemployed individuals, number of individuals out of the labor force and number of South African citizens[15].

In addition to conflict and Census data, we use three additional data sources. Following Michalopoulos and Papaioannou (2012, 2013), we use NOOA (2012) night-time light satellite images data for 1992 (the first available year) and 1996 as a proxy for economic conditions in South Africa at the MD level[16]. Consistently with their approach, we average night-time light density across 30-second grid areas (approximately 1 square kilometer) within the same MD. In line with the existing literature, we also make use of geographical variables as controls. In our cross-sectional specification we include use a MD-level measure of terrain ruggedness(data from Nunn and Puga (2012)). The measure is computed by averaging 30-second grid area observations belonging to the same MD[17]. The terrain slope index from Global Aero-ecological Zones (GAEZ) data (IIASA/FAO 2012) is created by averaging 5-minute by 5-minute (approximately 9 km by 9 km) grid-cells observations within the same MD. Finally, area accessibility from

[14]In the study of the relationship between ethnic polarization and conflict, the literature has focused on PRIO conflict (Montalvo and Reynal-Querol 2005, 2010; Esteban et al. 2012). This is mainly due to the cross-countries analysis and large time span (1945 to 2010) these studies usually consider. Non-state and one-sided conflicts are there considered as outcomes to test robustness of results.

[15]In our regression specification we use the logarithm of these variables, augmenting all values by 0.01 when some of them are equal to zero. Results go through other variables specification and level shift.

[16]For an extensive discussion of these data and their validity as a proxy for economic conditions in the african territories see Michalopoulos and Papaioannou (2012, 2013). See also Doll et al. (2006) and Sutton et al. (2007). In our regression specifications, consistently with Michalopoulos and Papaioannou (2012, 2013), we augment the night-time light satellite measure by 0.01 before taking its logarithm. Results are unaltered with respect to other or no level shift.

[17]Data are available at http://diegopuga.org/data/rugged/.

GAEZ is computed as estimated travel time to nearest city with 50,000 or more inhabitants in year 2000.

5 Empirical Strategy and Results

5.1 Preliminary Analysis

More than 2,000 non-state conflict events are geo-referenced in MDs of South Africa in the UCDP-GED dataset in the period 1989-1996[18]. ANC and IFP confronted each other in more than 85% of events. As for the others, within the most numerous are those where the United Democratic Front (UDF) is involved against the IFP, and the ANC Greens faction against the ANC Reds faction. An average number of 1.2 non-state conflict events per MD is recorded in 1991, with cross-district variation being larger than four times the national mean. Overall, conflict prevalence decreases after democratization in 1994. The average is 0.1 in 1996, but with a cross-district variation of 0.7[19].

[Table 1]

Table 1 shows the population sizes of the African native ethnolinguistic groups in South Africa according to the 1991 and 1996 Census. Together with nationwide stocks, the table reports the differences in the population of each ethnolinguistic group between 1991 and 1996. MDs in those Bantustans which were already granted independence are not covered by the 1991 Census of the Republic of South Africa. The inclusion of former independent Bantustan territories in the 1996 Census generates a dramatic increase in the population stocks of the correspondent ethnolinguistic groups. In the fourth column, the same overall differences are computed looking only at

[18]Table A.1 in the appendix reports the total number of one-sided conflict events involving the Government and non-state conflict events per year in MDs in South Africa recorded in the UCDP-GED dataset, together with the estimated total number of deaths.

[19]Summary statistics for the derived sample are reported in Table A.2 of the appendix.

those 294 districts which are part of both the 1991 and 1996 Census. We find evidence of a substantial inflow into these territories of individuals belonging to ethnicities previously segregated in the independent Bantustans. In some cases population inflows are large. At the extreme, the Tswana group population increases by more than 30%.

On top of migration inflows from independent Bantustans, we find evidence of substantial internal mobility[20]. According to 1996 Census data, 2.5 millions blacks moved from one MD to another in between 1991 and 1996. Figure 3 shows the percentage change in the population size of each ethnolinguistic group at the district level in between 1991 and 1996, plotted against the share of black population in 1991. It is worth noticing that the most relevant changes do not seem to be concentrated in those districts where the share of blacks in 1991 was either negligible or close to one.

[Figure 3]

As a result, the negligible change in average within-black polarization between 1991 and 1996 hides substantial variability over time across districts. The time difference goes from a minimum of -0.89 to a maximum of one, spanning almost the entire support of admissible values. Remarkably, the standard deviation of the overtime chenge in the polarization index is as high as half of its 1991 cross sectional standard deviation.

About 38% of conflict events recorded in 1991 takes place in the 25% of districts with the highest level of within-black polarization. More importantly, the relationship appears to be stronger when within-province variability in conflict incidence is considered. The 25% of districts with the highest within-black polarization have on average 0.68 conflicts more than

[20]Using data from the South Africa Migration and Health Survey (SAMHS), Reed (2013) studies internal migration patterns amongst the black population of South Africa in the second half of the twentieth century. He reports non-negligible migration rates even before the repeal of the Pass Law in 1986. Nonetheless, migration rates spike in 1991 and after, and estimates are very similar to the one we obtain using data from Statistics South Africa (1998). Indeed, SAMHS data are recognized by Reed (2013) to compare favorably to Census data.

the province average, as opposed to the 0.55 less recorded in the 25% of districts with the lowest within-black polarization.

5.2 Cross-sectional Estimates, 1991

We start the regression analysis by exploiting cross-district variation in both within-black polarization and conflict in 1991. We adopt the following linear regression specification

$$conf_{ip91} = \gamma_p + \beta\ ELP_{WB\ ip91} + \mathbf{Z}'_{ip}\omega + \mathbf{X}'_{ip91}\varphi + u_{ip91} \tag{2}$$

where $conf_{ip91}$ is the number of non-state conflict events recorded in district i in province p in 1991. $ELP_{WB\ ip91}$, computed as discussed in Section 4, is the within-black polarization index in the same district in 1991, while Z_{ip} is a set of time-invariant district geographical characteristics (ruggedness, slope index and accessibility). X_{ip91} is the vector of time-varying demographic and economic controls, capturing time-variant district characteristics in 1991 (population, black population, rural population, number of individuals reporting no education, number of unemployed, number of individuals out of the labor force and number of South Africa citizens). γ_p capture province fixed effects. The analysis therefore explore within-province district variability. The residual u_{ip91} captures instead those unobserved factors which affect conflict incidence.

[Table 2]

Table 2 provides the corresponding results. Throughout all specifications, an Ordinary Least Squares (OLS) estimation with province fixed effects is run on the available 1991 sample. Given the possible endogeneity of within-black polarization to conflict intensity, we start by providing results from a regression with no additional controls in the first column. The results show a positive and significant relationship between the within-blacks polariza-

tion index and number of conflict events at the district level[21]. An increase of one cross-district standard deviation of the within-black polarization index is associated with 0.7 more conflict events per year in 1991, more than half of the national average. Columns (2) and (3) include additional controls such as population, black population, and the Night-time Lights measure as proxy for local economic activity (all in logs). Both the magnitude and significance of the coefficient of interest turn out to be almost unaffected. Column (4) reports estimation results after further including geographic controls such as indexes of ruggedness, terrain slope and accessibility. The coefficient of interest unchanged in magnitude and significance with respect to columns (2) and (3). Results in column (5) show the results that come from our fully constrained specification that controls for a detailed set of economic covariates, including the log of rural population, the number of unemployed individuals and the number of individuals reporting no education. In this specification, the coefficient of the within-blacks polarization is slightly reduced and almost significant. Overall, results are consistent with the hypothesis of the observed non-state conflict events to be qualifiable as expressed along ethnic lines. These findings are in line with the cross-country literature relating ethnic polarization to conflict as discussed in Section 2.

5.3 First-difference Estimates, 1991-1996

The restoration of free internal mobility of blacks after 1991 largely accounts for time variation in ethnolinguistic group population shares per district. Together with initial cross-district variation, it opens the way for the implementation of a first-difference strategy which looks at the relationship between within-black ethnolinguistic polarization and non-state conflict net of district-specific time-invariant characteristics. In other words, it is possible to test whether the observed change in within-black polarization

[21]We performed the same estimations using a Poisson model specification which takes into account the count nature of our dependent variable. Results are substantially unchanged. These are available in the Online Appendix. Finally, as a further check, we use the logarithm of the number of conflict events (augmented by 0.01) as outcome in all specifications and find highly consistent results.

significantly correlates with the change in non-state conflict incidence.

We combine the available information from both 1991 and 1996 and adopt the following specification

$$\Delta conf_{\,it} = \delta_t + \beta\ \Delta ELP_{WB\ it} + \Delta \mathbf{X}'_{it}\varphi + \Delta u_{it} \tag{3}$$

where $\Delta conf_{\,it}$ is the change in the number of recorded non-state conflict events in district i in between year 1991 and 1996, while $\Delta ELP_{WB\ it}$ is the correspondent change in the within-blacks polarization index. The proposed first-difference specification allows to cancel out both observable and unobservable time-invariant characteristics at the district level. The effect of nationwide events (such as democratization in 1994) and general time trends are instead captured by the constant term δ_t. As before, X_{it} is the vector of time-varying demographic and economic controls in year t (population, blacks, night-time light, rural, etc.). The difference residual Δu_{it} captures those unobserved changes and factors which affect the change in conflict incidence.

Results are reported in Table 3. All specifications are implemented over the sample of districts for which Census data are available for both 1991 and 1996. The first column provides the results from the unconstrained specification. The negative estimate of the constant term is consistent with the general decrease in conflict prevalence with the first democratic elections in 1994. The estimated coefficient of the ethnolinguistic polarization is highly significant and higher in magnitude than the one obtained from the 1991 cross-sectional analysis[22]. An increase in one cross-district standard deviation in the within-black ethnolinguistic polarization measure is now associated with a 1.4 increase in the number of conflict per district, which is again more than the 1991 national average. Column (2) and (3) show that the results are robust to the inclusion of time-variant economic controls. The coefficient of interest remains stable and significant.

[22]We produced estimates from a non linear specification. Results are still positive and significant using a fixed-effects Poisson model. These are available in the Online Appendix. As before, we also use the logarithm of the number of conflict events (augmented by 0.01) as outcome in all specifications and still find consistent results.

[Table 3]

Column (4) provides estimation results assuming heteroskedastic difference residuals and estimating Eicker-Huber-White robust standard errors (White 1980). This allows to take into account heterogeneity in the variability of the first-difference residuals in the computation of the standard errors used for inference. The coefficient of interest is still significant. In column (5), we take into account cross-sectional dependence of first-difference residuals and follow Conley (1999) in allowing for non-zero correlation when coordinate distance between districts' centroids is less than 1 degree latitude and/or 1 degree longitude (approximately 110 km). The estimate of our coefficient of interest retains significance at the 10% level.

Finally, column (6) reports the results from a specification which augments the first-difference one by including the number of non-state conflict events in the 1989 and 1990. Given the general decrease in the number of conflict events over the period, we want to test whether the systematic relationship we find between within-black polarization and conflict incidence is robust to conditioning on a measure of pre-1991 conflict incidence[23]. The coefficient of the latter is negative and highly significant. This means that the observed decrease in the number of conflict events in between 1991 and 1996 was higher in those districts with a high number of conflicts in 1989 and 1990. More importantly, we still find evidence of a systematic relationship between within-black polarization and conflict incidence. The estimated coefficient is highly significant. Together with the negative estimate of the pre-1991 conflict measure coefficient, this suggests a negative relationship across districts between the change in within-black polarization in between 1991 and 1996 and conflict incidence in 1989-1990.

[23]Given the adopted linear regression model specification, the estimator of the parameter in column (6) is inconsistent and not directly comparable with the ones from the simple first-difference specification. The inclusion of a lag as a regressor in the first-difference model does not allow to net out district-level time-invariant unobservable characteristics. We intentionally compromise on the consistency of our estimator in order to check whether our results are robust to a partial solution to the problem of mean-reversion and convergence.

5.4 First-difference with Cluster-specific Trends

One possible concern is that the results we find using a first-difference approach can be driven by the presence of unobserved time-varying factors which affect the evolution of both ethnic composition and conflict in the same way, generating a spurious correlation between the two. For example, those districts where polarization increased in 1991-1996 might be systematically different from others. These may belong to fully segregated areas with low within-black polarization in 1991. Unobserved characteristics such as the existence of natural resources might have been on their own responsible for the change in the incidence of conflict within the black majority, generating a spurious correlation between changes in polarization and conflict.

We study the extent to which these issues are likely to affect the results by looking at the evolution of within-black polarization and conflict within clusters of neighboring districts. Similarly to the neighbors-pair fixed-effects analysis in Acemoglu et al. (2013), our argument is that districts located next to each other are highly comparable in terms of both observable and unobservable characteristics. When looking at the relationship of interest within clusters of neighboring districts, this should help in netting out to unobserved sources of heterogeneity and time-varying omitted variables correlated with the evolution of both polarization and conflict incidence.

We define as *treated* those districts $g \in M$ where our polarization measure *decreased more than average* in 1991-1996. We then keep their neighboring *non-treated* districts $f \in N(g)$, and drop treated districts with no non-treated neighboring districts. We obtain a sub-sample of 227 out of the initial 294 districts[24]. For each treated district $g \in M$ and its non-treated neighbors $f \in N(g)$, we focus on 1991-1996 differences and consider the

[24]Following the proposed definition, the final subsample contains 105 treated and 122 non-treated districts.

following model

$$\begin{aligned} \Delta conf_{gt} &= \delta_t + \gamma_g + \beta\ \Delta ELP_{WB\ gt} + \Delta \mathbf{X}'_{gt}\varphi + \Delta\varepsilon_{gt} \qquad g \in M \\ \Delta conf_{ft} &= \delta_t + \gamma_g + \beta\ \Delta ELP_{WB\ ft} + \Delta \mathbf{X}'_{ft}\varphi + \Delta\varepsilon_{ft} \quad f \in N(g) \end{aligned} \tag{4}$$

where, as before, $\Delta conf_{it}$ is the change in the number of recorded non-state conflict events in district i in between year 1991 and 1996, while $\Delta ELP_{WB\ it}$ is the change in the within-blacks polarization index (with $i = g, f$). Time trends are still controlled for by δ_t. γ_g captures cluster-specific trends, controlling for unobservable determinants of evolution of conflict in 1991-1996, possibly related with change in polarization over the period. In other words, we introduce a dummy for each treated district which takes value 1 when the observation is the treated district itself or one of its non-treated neighbors. X_{it} is the vector of time-variant district characteristics in year t (population, blacks, night-time light, rural, etc.). The difference residual $\Delta\varepsilon_{it}$ captures those idiosyncratic unobserved changes and factors which affect the change in conflict incidence, net of cluster-specific trends.

In the final restricted sample, each non-treated district is possibly neighbor of more than one treated district. Thus, there exist several different ways to group districts into clusters. We implement a bootstrap-type procedure where we run a series of regressions, matching in each repetition each non-treated district to a single treated district. The results deliver an empirical distribution of parameter estimates $\hat{\beta}$ which can be used for inference[25].

[Table 4]

Results in Table 4 confirms the previous findings. Column (1) reports the

[25]The focus on clusters instead of pairs of neighboring districts and the implementation of the suggested bootstrap-type procedure differs from Acemoglu et al. (2013). Our approach avoids duplicating observations as required by their neighbors-pair fixed-effects strategy, and delivers standard errors for inference without building up and estimating the parameters of the variance-covariance matrix of residuals as required by their random-effects strategy.

estimate of the coefficient of interest when only the logarithm of total black population, total population and the night-time satellite light variables are included as controls. The full set of other time-variant economic controls is included in column (2). Results are substantially unchanged. A full set of third-degree polynomials of covariates is included in column (3). We do this in order to control for non-linear discontinuities at the border, which may themselves be correlated with the outcome variable (Acemoglu et al. 2013). The point estimate is still highly significant, even if lower in magnitude.

Given the restricted sample, we attach to these estimates only a local interpretation. The results from this section are largely consistent with the ones we found in the previous section. If anything, the presence of unobserved omitted factors which are systematically correlated with both the change in within-black polarization and conflict incidence seem to downward bias our initial first-difference estimate of the parameter of interest.

5.5 Endogeneity of Internal Migration Flows and Instrumental Variable Strategy

With the implementation of a first-difference identification strategy we compare districts which experienced different changes in the within-black polarization measure in between 1991 and 1996, and find heterogeneity along this dimension to be significantly correlated with the observed changes in the incidence of conflict within the black majority at the district level. Changes over time in the within-black polarization measure occur when group population sizes change at different rates: if all ethnolinguistic group populations were changing at the same rate in all districts, the polarization measure would remain unchanged everywhere. Therefore, one concern for the validity of our empirical exercise is that the disproportional changes in the size of ethnolinguistic groups may be driven by the change of conflict incidence.

Conflict can causes displacement of individuals and households. High non-state conflict incidence in 1991 could be a *push* factor that positively affects individuals' propensity to migrate out of the district after 1991. By the

same token, expectations about low non-state conflict incidence in 1996 can be conceptualized as a *pull* factor in the same framework. If agents' expectations were fulfilled, migration decisions would be endogenous to conflict incidence in both 1991 and 1996. Nonetheless, notice that this reasoning is not informative of the possible endogeneity of migration decisions to the *changes* in conflict incidence over the period. Furthermore, our identification strategy would be invalidated only if the endogeneity of migration decisions mapped into the endogeneity of the change in polarization to the change in conflict at the district level in a systematic way. The absence of a partial order for increasing polarization, as shown by Esteban and Ray (1994), prevents establishing such map between migration flows and changes in the polarization measure.

To investigate the issue, we implement an instrumental variable (IV) strategy where pairwise distance between districts is used to predict location decisions of internal migrants. We use information from the 1996 Census on surveyed individuals who declare to have moved in between 1991 and 1996 and to have been resident in a different district in 1991. Using the same data, Kok et al. (2003) show a negative non-linear relationship between the number of migrants moving between two districts and pairwise distance between them. We exploit this feature and estimate a conditional logit model (Cameron and Trivedi 2005) for the location decision of migrants in the sample of the form

$$p_{ij} = \frac{e^{\beta \ distance_{ij}}}{\sum_{j \in S} e^{\beta \ distance_{ij}}} \tag{5}$$

where $distance_{ij}$ is the distance (in km) between district i and j in the set of South Africa districts S. We thus estimate the probability of each individual leaving district i after 1991 to be observed in district j in 1996 as predicted by the values of pairwise distance only[26].

In a simulation exercise, we next assume that a fraction x of individuals from each ethnolinguistic group e leave each district $i \in S$ after 1991,

[26]Conditional Logit estimation results are provided in the Online Appendix to the paper.

and allocate them to districts $j \neq i$ using the estimated probabilities $\hat{p}_{ij}$ derived as above. The value of x is chosen by matching it with the total district-level outflow rate of blacks who located in any other district over the relevant period[27]. The estimated district-level share of black population that moved from one MD to another is equal to 8.66%. We thus compute predicted population stocks for each black ethnolinguistic group in each district i in 1996 as

$$\widehat{N}_{e,i,1996} = N_{e,i,1991}(1-x) + \sum_{j \in S} \hat{p}_{ji}\ xN_{e,j,1991} \tag{6}$$

The predicted population share are next used to construct a predicted within-black ethnolinguistic polarization index for 1996 using the same formula as in section 4. We then take the difference between our predicted ethnolinguistic polarization in 1996 and the actual one in 1991 in order to obtain the predicted change in polarization between 1991 and 1996, $\Delta\widetilde{ELP}_{WB,i}$. Notice that, by forcing x to be the same for all ethnic groups in all districts, we rely on $\hat{p}_{ji}$ only in the generation of predicted time variability in the change in the polarization index. Our predicted value for the change in within-black ethnolinguistic polarization is the one which would be observed in case a given exogenous fraction of individuals in each ethnic group in each district were to leave after 1991 and locate in another district according to what predicted by distance only. Indeed, individual reallocation probabilities are computed as independent from conflict levels and their changes at both origin and destination, and assigned to each fictitiously displaced individual. We use this prediction as a source of exogenous variation for the actual change in polarization observed in the data, ruling out the potential endogeneity of individuals' migration decisions.

[Table 5]

[27]Focusing on the 294 districts in our sample, we calculate the total number of individuals moving from one MD to another in between 1991 and 1996 as recorded in the 1996 Census. We then divide this number by the total number of individuals belonging to any black ethnolinguistic group living in any of the 294 districts surveyed in the 1991 Census of the Republic of South Africa.

Results from both the first and the second-stage regressions are reported in Table 5. Conditional on the other variables included as controls, the *F-statistics* for the significance test of the instrument in the first-stage regression are safely above 10 in all specifications. The instrument appears strong enough in producing a relevant shift in the actual value of the change in the within-black polarization index. For consistency with the previous analysis, we start by including only the total number of blacks, total population and night-time satellite light value in logs as controls in the first column. The point estimate of the coefficient of interest is significant and around three times higher in magnitude with respect to the one obtained in Table 3, suggesting a downward bias from endogenous migration decisions. An increase in within-black polarization of one 1991 cross-district standard deviation is associated with 4.4 more conflict events, more than three times the 1991 national average. However, results from a Hausman test do not allow to reject the hypothesis of both the first-difference and the IV estimators being consistent (Hausman 1978). In the second column, we allow first-difference residuals to be heteroskedastic and estimate Eicker-Huber-White robust standard errors (White 1980). The estimate of the within-black polarization coefficient are significant at the 5% level[28]. The same pattern of significance is observed when conditioning on the full set of economic controls in column (3) and (4), with the point estimate being somewhat lower in magnitude.

Overall, results from the implemented instrumental variable strategy confirm the existence of a positive relationship between within-black ethnolinguistic polarization and non-state conflict incidence in the data. Ruling out potential threats from the endogeneity of individual migration decisions produces results which are qualitatively similar than the ones previously obtained. Given the spatial heterogeneity of internal migration flows we

[28]Eicker-Huber-White robust standard errors may be lower than conventional ones. Angrist and Pischke (2010) show that this can be the case whenever lower variance of the residual is associated with covariate values far from the mean of the covariate distribution. In our case, given that the mean of the predicted change in polarization is close to zero, this would mean that those districts which are assigned by our instrument to experience the smallest change in polarization also have higher variability in the first-difference residuals of the change in conflict.

observe in the data, we cannot exclude differences in magnitude to be attributable to the local interpretation of the instrumental variable estimate of the parameter we wish to identify (Angrist et al. 1996).

6 Conclusions

This paper studies the extent to which conflict incidence correlates over time with contemporaneous measures of ethnic distribution. The history of contemporary South Africa in 1991 through 1996 carries with it substantial variation in the variables of interest. Combining Census data with georeferenced information on conflict, we show the incidence of violent struggles amongst black-dominated organized groups during the fall of apartheid to be positively correlated with within-black ethnolinguistic polarization. Time variability along both dimensions allows to investigate the relationship of interest after clearing out the impact of unobserved time-invariant characteristics at the local level. A one cross-district standard deviation increase in within-black polarization is found to be associated with an increase in the number of conflict events of more than the 1991 national average. Findings are robust to several additional checks. Comparing the evolution of within-black polarization and conflict within clusters of neighboring districts still yields highly significant results. The same holds when internal migration patterns and the resulting changes in ethnic distribution are simulated using location probabilities as predicted by pairwise distance between districts: instrumental variable estimates are consistent with first-difference ones[29].

The approach and empirical results in this paper contribute to disclose the potential of the use of micro-level data in the study of social conflict and its determinants. On one hand, migration-driven changes in the ethnic distribution at the local level are here revealed to be informative of conflict prevalence. The extent to which this result deserves further exploration on

[29]Results are confirmed when the estimated number of deaths in non-state conflicts per MD is used as measure of conflict incidence, and a measure of the intensity of governmental repression is controlled for. These findings are not for publication and can be found in the Online Appendix to the paper.

the theoretical ground or whether it can help to identify the relative importance of different conflict motives constitute fruitful avenues for future research. On the other hand, the specificities of the South Africa setting can be further investigated in order to study the effect of democratization on conflict. Evidence from this paper suggests the heterogeneity and divisions within the newly enfranchised majority to interact with nationwide institutional changes and still inform conflict incidence at the local level even after democratization in 1994. The need for a theoretical and empirical investigation of this argument motivates our future research agenda.

References

Acemoglu, D., C. García-Jimeno, and J. A. Robinson (2013). Finding Eldorado: Slavery and Long-run Development in Colombia. *Journal of Comparative Economics*. Forthcoming.

Alesina, A., A. Devleeschauwer, W. Easterly, S. Kurlat, and R. Wacziarg (2003). Fractionalization. *Journal of Economic Growth 8*(2), 155–94.

Angrist, J. D., G. W. Imbens, and D. B. Rubin (1996). Identification of Causal Effects Using Instrumental Variables. *Journal of the American Statistical Association 91*(434), pp. 444–455.

Angrist, J. D. and J. S. Pischke (2010). A Note on Bias in Conventional Standard Errors under Heteroskedasticity. Unpublished.

Besley, T. and M. Reynal-Querol (2013). The Legacy of Historical Conflict. Evidence from Africa. Unpublished.

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

Cameron, A. C. and P. K. Trivedi (2005). *Microeconometrics: Methods and Applications*. Cambridge University Press, New York.

Case, A. and A. Deaton (1999). School Inputs and Educational Outcomes in South Africa. *The Quarterly Journal of Economics 114*(3), pp. 1047–1084.

Caselli, F. and W. J. Coleman (2013). On the Theory of Ethnic Conflict. *Journal of the European Economic Association 11*, 161–192.

Clark, N. L. and W. H. Worger (2011). *South Africa. The Rise and Fall of Apartheid* (2nd ed.). Pearson.

Collier, P. (2001). Implications of Ethnic Diversity. *Economic Policy 16*(32), 127–166.

Collier, P., V. Elliott, H. Hegre, A. Hoeffler, M. Reynal-Querol, and N. Sambanis (2003). *Breaking the Conflict Trap: Civil War and Development Policy (Policy Research Reports)*. World Bank; Oxford University Press.

Collier, P. and A. Hoeffler (2004). Greed and Grievance in Civil War. *Oxford Economic Papers 56*(4), 563–595.

Conley, T. G. (1999). GMM Estimation with Cross Sectional Dependence. *Journal of Econometrics 92*(1), 1–45.

Desmet, K., I. Ortuño Ortín, and R. Wacziarg (2012). The Political Economy of Linguistic Cleavages. *Journal of Development Economics 97*(2), 322–338.

Desmet, K., I. Ortuño Ortín, and S. Weber (2009). Linguistic Diversity and Redistribution. *Journal of the European Economic Association 7*(6), 1291–1318.

Doll, C. N., J.-P. Muller, and J. G. Morley (2006). Mapping Regional Economic Activity from Night-time Light Satellite Imagery. *Ecological Economics 57*(1), 75 – 92.

Eck, K. (2012). In Data We Trust? A Comparison of UCDP GED and ACLED Conflict Events Datasets. *Cooperation and Conflict 47*(1), 124–141.

Esteban, J., L. Mayoral, and D. Ray (2012). Ethnicity and Conflict: An Empirical Study. *American Economic Review 102*(4), 1310–42.

Esteban, J. and D. Ray (1994). On the Measurement of Polarization. *Econometrica 62*(4), 819–51.

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

Esteban, J. and D. Ray (2008). On the Salience of Ethnic Conflict. *American Economic Review 98*(5), 2185–2202.

Esteban, J. and D. Ray (2011a). A Model Of Ethnic Conflict. *Journal of the European Economic Association 9*(3), 496–521.

Esteban, J. and D. Ray (2011b). Linking Conflict to Inequality and Polarization. *American Economic Review 101*(4), 1345–74.

Harari, M. and E. La Ferrara (2012). Conflict, Climate and Cells: A Dis-

aggregated Analysis. Working Papers 461, IGIER (Innocenzo Gasparini Institute for Economic Research), Bocconi University.

Hausman, J. A. (1978). Specification Tests in Econometrics. *Econometrica 46*(6), 1251–71.

Hoffman, T. M. and S. W. Todd (2000). A National Review of Land Degradation in South Africa: the Influence of Biophysical and Socioeconomic Factors. *Journal of Southern African Studies 26(4)*, 743–758.

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

IIASA/FAO (2012). *Global Agro-ecological Zones (GAEZ v3.0)*. Laxenburg, Austria and FAO, Rome, Italy: available at http://www.gaez.iiasa.ac.at/ (accessed April 20, 2012).

Kok, P., M. O'Donovan, O. Bouare, and J. Van Zyl (2003). *Post Apartheid Patterns of Internal Migration in South Africa*. Human Sciences Research Council.

Lewis, M. P. (2009). *Ethnologue: Languages of the World, Sixteenth edition*. SIL international Dallas, TX, available at http://www.ethnologue.com/ (accessed March 11, 2012).

Melander, E. and R. Sundberg (2011). Climate Change, Environmental Stress, and Violent Conflict: Tests Introducing the UCDP Georeferenced Event Dataset. Paper presented at the International Studies Association, March 16-19, Montreal, Canada.

Michalopoulos, S. and E. Papaioannou (2011). The Long-Run Effects of the Scramble for Africa. NBER Working Papers 17620, National Bureau of Economic Research, Inc.

Michalopoulos, S. and E. Papaioannou (2012). National Institutions and African Development: Evidence from Partitioned Ethnicities. NBER Working Papers 18275, National Bureau of Economic Research, Inc.

Michalopoulos, S. and E. Papaioannou (2013). Pre-colonial Ethnic Institutions and Contemporary African Development. *Econometrica 81(1)*, 113–152.

Montalvo, J. and M. Reynal-Querol (2005). Ethnic Polarization, Potential Conflict and Civil War. *American Economic Review 95*(3), 796–816.

Montalvo, J. and M. Reynal-Querol (2010). Ethnic Polarization and the Duration of Civil Wars. *Economics of Governance 11*(2), 123–143.

NOOA (2012). *Version 4 DMSP-OLS Nighttime Lights Time Series*. available at http://www.ngdc.noaa.gov/dmsp/downloadV4composites.html (accessed April 20, 2012).

Novta, N. (2013). Ethnic Diversity and the Spread of Civil War. Unpublished.

Nunn, N. and D. Puga (2012). Ruggedness: The Blessing of Bad Geography in Africa. *The Review of Economics and Statistics 94*(1), 20–36.

Posel, D. (2001). What's in a Name? Racial Categorisations Under Apartheid and Their Afterlife. *Transformation 47*, 50–74.

Posel, D. and D. Casale (2003). What Has Been Happening To Internal Labour Migration In South Africa, 1993-1999? *South African Journal of Economics 71*(3), 455–479.

Reed, H. E. (2013). Moving Across Boundaries: Migration in South Africa, 1950-2000. *Demography 50*(1), 71–95.

Reynal-Querol, M. (2002). Ethnicity, Political Systems, and Civil Wars. *Journal of Conflict Resolution 46*(1), 29–54.

Rohner, D., M. Thoenig, and F. Zilibotti (2012). Seeds of Distrust: Conflict in Uganda. OxCarre Working Papers 078, Oxford Centre for the Analysis of Resource Rich Economies, University of Oxford.

Statistics South Africa (1991). *South African Census 1991 [dataset]. Version 1.1.* Pretoria: Statistics South Africa [producer]. Cape Town: DataFirst [distributor]. http://www.datafirst.uct.ac.za/ (accessed February 3, 2012).

Statistics South Africa (1998). *South African Census 1996 [dataset]. Version 1.2.* Pretoria: Statistics South Africa [producer]. Cape Town:

DataFirst [distributor]. http://www.datafirst.uct.ac.za/ (accessed February 3, 2012).

Statistics South Africa (2007). *Using the 2001 Census: Approaches to Analysing Data*. Statistics South Africa, Pretoria.

Sundberg, R., M. Lindgren, and A. Padskocimaite (2010). *UCDP GED Codebook Version 1.0-2011*. Department of Peace and Conflict Research, Uppsala University.

Sutton, P. C., C. D. Elvidge, and T. Ghosh (2007). Estimation of Gross Domestic Product at Sub-national Scales Using Nighttime Satellite Imagery. *International Journal of Ecological Economics & Statistics 8*(S07), 5–21.

University of Texas at Austin (1986). *South Africa: Black Homelands 1986*. Perry-Castañeda Library Map Collection, http://en.wikipedia.org/wiki/File:Southafricanhomelandsmap.png (accessed June 8, 2012).

White, H. (1980). A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity. *Econometrica 48*(4), 817–38.

Tables and Figures

Table 1: Ethnic Group Population Sizes

	Total Population		Difference 1996-1991	
	1991 Census 298 MDs	1996 Census 354 MDs	Overall	Study Sample 294 MDs
Xhosa	2,493,382	7,206,005	4,712,623	77,370
Zulu	8,416,125	9,216,413	800,288	163,555
Sotho	6,414,684	7,378,473	963,789	234,038
Swazi	943,989	1,017,233	73,244	45,365
Tswana	1,437,660	3,299,902	1,862,242	463,044
Tsonga	1,450,874	1,757,589	306,715	82,791
Venda	116,533	876,546	760,013	2,296

Notes. The Table shows the total ethnic group population sizes in our sample in 1991 and 1996, together with the overall difference and the difference using only all districts for which information is available in both 1991 and 1996 samples (Sources: Statistics South Africa 1991, 1998).

Table 2: 1991 Cross-sectional Estimation

	Total Number of Non-state Conflict Events 1991				
	(1)	(2)	(3)	(4)	(5)
ELP_{WB}	2.399	2.464	2.444	2.408	1.896
	(1.07)	(1.09)	(1.10)	(1.10)	(1.16)
Blacks (log)		-0.046	-0.029	-0.271	-0.355
		(0.42)	(0.43)	(0.47)	(0.52)
Population (log)		0.367	0.312	0.536	0.211
		(0.52)	(0.55)	(0.59)	(3.89)
Night-time Lights (log)			0.038	0.022	-0.013
			(0.12)	(0.13)	(0.13)
Rural Population (log)					-0.001
					(0.13)
No Education (log)					0.349
					(1.64)
Unemployed (log)					1.215
					(0.73)
Constant	-0.425	-3.973	-3.414	-1.768	0.413
	(0.74)	(3.19)	(3.66)	(4.03)	(4.93)
Province Fixed Effects	Y	Y	Y	Y	Y
Geographic Controls	N	N	N	Y	Y
Other Economic Controls	N	N	N	N	Y
Observations	294	294	294	291	291
R^2	0.180	0.185	0.185	0.199	0.214

Notes. Standard errors in parenthesis. The table reports Ordinary Least Squares coefficients estimates from the 1991 cross-sectional specification. The unit of observation is a MD in South Africa for which information is available in 1991. The dependent variable is the total number of non-state conflict events coded in the MD in 1991 in the UCDP-GED dataset. ELP_{WB} is the district-level within-black polarization measure. Other controls are defined as in the data section (Sources: Statistics South Africa 1991, 1998; Nunn and Puga 2012; IIASA/FAO 2012; NOOA 2012).

Table 3: First-difference Estimation

	Change in Total No. of Non-state Conflict Events 1991-1996					
	(1)	(2)	(3)	(4)	(5)	(6)
ΔELP_{WB}	5.924	5.562	5.178	5.178	5.178	4.107
	(1.59)	(1.63)	(1.66)	(2.96)	(3.04)	(1.41)
Non-state Conf 89-90						-0.273
						(0.03)
Δ Blacks (log)		-0.276	-0.325	-0.325	-0.325	-0.228
		(0.43)	(0.45)	(0.27)	(0.26)	(0.38)
Δ Population (log)		0.019	5.366	5.366	5.366	-0.796
		(0.75)	(4.80)	(6.70)	(5.10)	(4.11)
Δ Night-time Lights (log)		-0.277	-0.239	-0.239	-0.239	-0.129
		(0.16)	(0.16)	(0.25)	(0.29)	(0.13)
Δ Rural Population (log)			-0.682	-0.682	-0.682	-0.378
			(0.22)	(0.26)	(0.29)	(0.19)
Δ No Education (log)			-1.011	-1.011	-1.011	0.540
			(1.75)	(1.94)	(1.43)	(1.49)
Δ Unemployed (log)			-0.113	-0.113	-0.113	-0.112
			(0.83)	(0.61)	(0.68)	(0.70)
Constant	-1.068	-0.882	-3.718	-3.718	-3.718	-1.290
	(0.28)	(0.30)	(1.91)	(1.80)	(1.89)	(1.63)
Other Economic Controls	N	N	Y	Y	Y	Y
Observations	294	294	294	294	294	294
R^2	0.045	0.058	0.096	0.096	0.096	0.354

Notes. Standard errors in parenthesis. The table reports first-difference coefficients estimates. The unit of observation is a MD in South Africa for which information is available in both periods. The dependent variable is the change in total number of non-state conflict events coded in the MD in between 1991 and 1996 in the UCDP-GED dataset. ELP_{WB} is the district-level within-black polarization measure. *Non-state Conf 89-90* is the total number of conflict events coded in the MD in 1989-1990. Other controls are defined as in the data section (Sources: Statistics South Africa 1991, 1998; NOOA 2012). (1), (2), (3) are first-difference estimates assuming homoskedastic difference residuals; those in (4) are first-difference estimates assuming heteroskedastic difference residuals; in (5) we allow for cross-sectional dependence in the structure of difference residuals (Conley 1999) allowing for non-zero correlation when coordinate distance between districts' centroids is less than 1 degree latitude and/or 1 degree longitude (approximately 110 km). Those in (6) are estimates from the first-difference model augmented with the measure of pre-1991 conflict incidence, assuming homoskedastic difference residuals.

Table 4: First-difference with Cluster-specific Trends

	Change in Total No. of Non-state Conflict Events		
	(1)	(2)	(3)
ΔELP_{WB}	6.08	6.36	5.26
	(1.83)	(1.98)	(1.51)
Other Economic Controls	N	Y	Y
Polynomial of Controls	N	N	Y
Observations	227	227	227
Repetitions	200	200	200

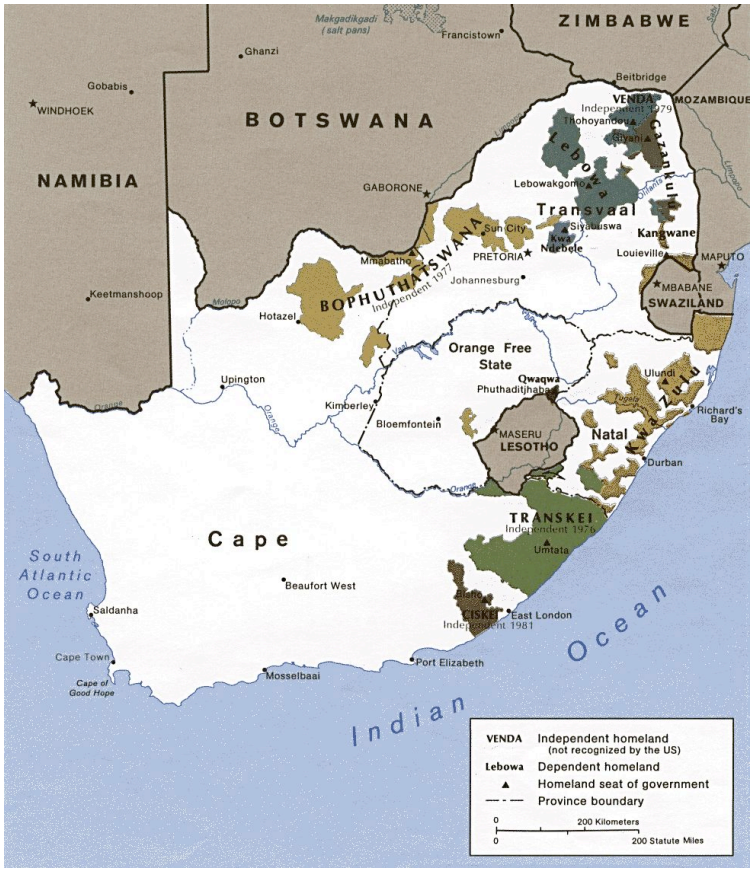
Notes. Empirical standard errors in parenthesis. The table reports coefficients estimates and standard errors from a first-difference specification with cluster-specific trends. A restricted sample of *treated* districts are kept from the initial sample, together with their neighboring *control* districts (see text for details). A bootstrap-type procedure is implemented, where at each repetition every control district is randomly matched to a single treated district and coefficients from the first-difference specification with cluster-specific trends are estimated. The total number of blacks and total population, and night-time satellite light value in logs are used as controls in (1), while all economic controls are included in (2). Specification (3) is augmented with a 3rd order polynomial of controls (Sources: UCDP-GED v1.5; Statistics South Africa 1991, 1998; NOOA 2012).

Table 5: Instrumental Variable Estimation

	(1)	(2)	(3)	(4)
1st Stage	Change in Polarization Measure ΔELP_{WB}			
$\Delta\widetilde{ELP}_{WB}$	0.244	0.244	0.249	0.25
	(0.06)	(0.06)	(0.06)	(0.06)
F-stat	17.13	17.13	17.60	17.60
robust *F-stat*		14.82		16.07
2nd Stage	Change in Total Number of Non-state Conflict Events			
$\Delta\widehat{ELP}_{WB}$	13.438	13.438	12.074	12.074
	(7.10)	(5.85)	(6.95)	(5.28)
Δ Blacks (log)	-0.121	-0.121	-0.117	-0.117
	(0.46)	(0.26)	(0.50)	(0.28)
Δ Population (log)	0.361	0.361	2.640	2.640
	(0.83)	(0.76)	(5.54)	(6.15)
Δ Night-time Lights (log)	-0.252	-0.252	-0.243	-0.243
	(0.16)	(0.24)	(0.16)	(0.24)
Constant	-0.969	-0.969	-2.822	-2.822
	(0.32)	(0.29)	(2.12)	(1.75)
Economic Controls	N	N	Y	Y
Observations	294	294	294	294

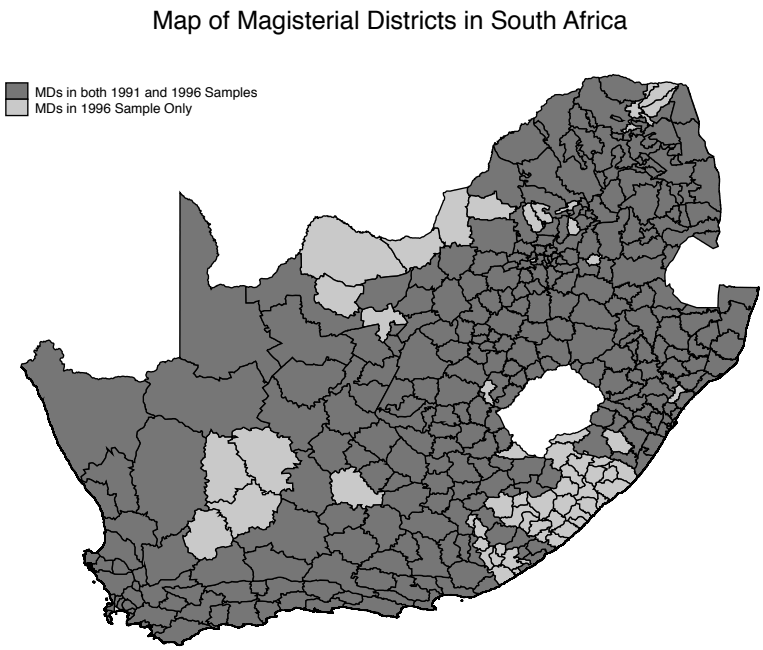
Notes. Standard errors in parenthesis. The table reports first-stage and second-stage instrumental variable estimates of the first-difference baseline model. The unit of observation is a MD in South Africa for which information is available in both periods. The dependent variable is the change in the total number of non-state conflict events coded in the MD in between 1991 and 1996 in the UCDP-GED dataset. ELP_{WB} is the district-level within-black polarization measure. The total number of blacks and total population, and night-time satellite light value in logs are used as controls in both the first and second stage in (1) and (2), while all other economic controls are included in (3) and (4). First-difference residuals are assumed to be homoskedastic in columns (1) and (3), and heteroskedastic in (2) and (4), where Eicker-Huber-White robust standard errors (White 1980) are estimated (Sources: UCDP-GED v1.5; Statistics South Africa 1991, 1998; NOOA 2012).

Figure 1: Map of Bantustans



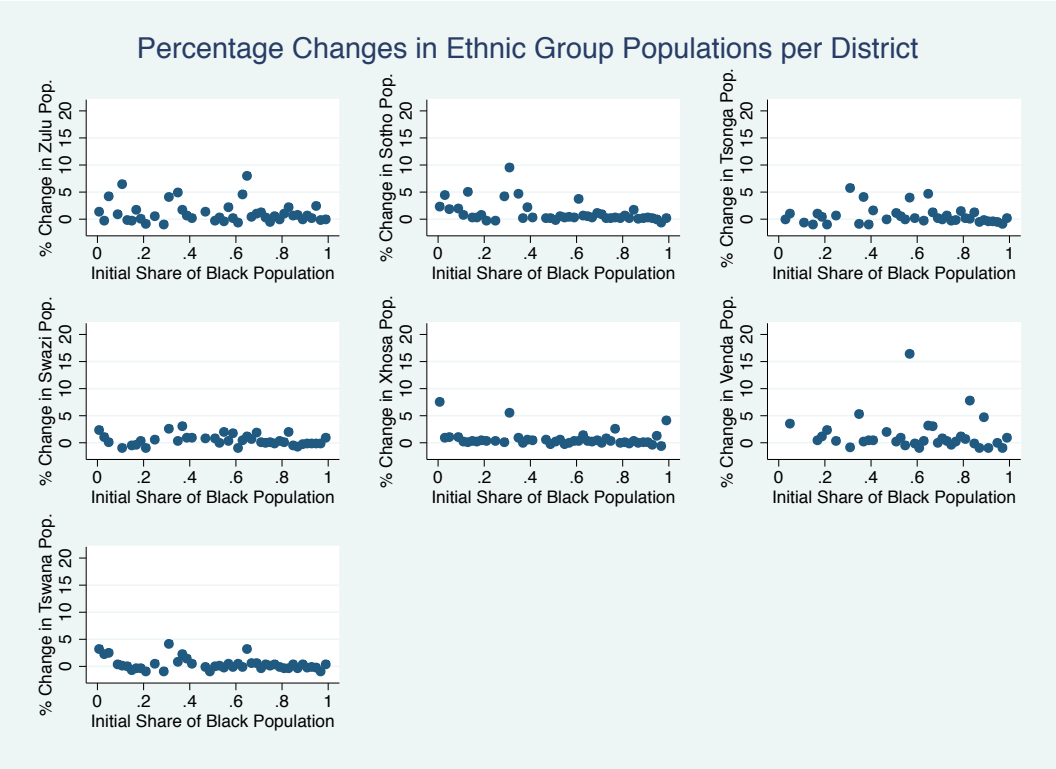
Notes. Map of Bantustans in South Africa as of 1986 (produced by the U.S. CIA. Source: University of Texas at Austin 1986).

Figure 2: Map of Magisterial Districts in South Africa



Notes. Map of Magisterial Districts in South Africa, indicating in dark grey those we for which information can be retreived from both the 1991 and 1996 Census (Statistics South Africa 1991, 1998). Those Bantustans which were already granted independence are not covered by the 1991 Census of the Republic of South Africa (Source: authors' elaboration using Stata).

FIGURE 3: ETHNIC GROUPS POPULATION PER DISTRICT: CHANGES 1991-1996



Notes. The figure shows the percentage change in population size of each ethnolinguistic group at the district level plotted over the initial share of blacks. Observations are districts for which we are able to retrieve information consistently from both 1991 and 1996 Census (Source: Statistics South Africa 1991, 1998). This excludes districts in the apartheid homelands which were granted independence and were no longer part of the Republic of South Africa. Observations are averaged per bins of share of blacks of size 2%.

A Appendix

Table A.1: One-sided and Non-state Conflict events: 1989-1998

Year	Gov. Repression		Non-state Conflicts	
	Conflict Events	Est. Deaths	Conflict Events	Est. Deaths
1989	34	54	185	226
1990	147	195	399	1243
1991	48	49	385	657
1992	61	61	479	665
1993	73	67	415	643
1994	16	14	230	444
1995	0	0	39	143
1996	0	0	37	156
1997	0	0	20	30
1998	0	0	23	44
1999	0	0	11	19
2000	0	0	1	0
2002	3	2	0	0
2004	0	0	4	0
2010	1	1	0	0

Notes. The Table shows the total number of One-sided and Non-state conflict events per year in South Africa recorded in the geo-referenced Event Dataset of the Uppsala Conflict Data Program (UCDP-GED v1.5) which we are able to map into MDs, together with the estimated total number of deaths.

Table A.2: Summary Statistics

Panel A: Year 1991

Variable	Mean	St. Dev.	Min	Max	N
Non-state Conflict Events	1.271	5.192	0	41	303
ELP_{WB}	0.368	0.336	0	0.989	296
Blacks	72.418	110.795	0	972.838	299
Population	103.67	156.61	3.04	1546.067	299
Rural Population	44.876	76.325	0	419.321	299
No Education	29.879	37.919	1.136	253.145	299
Unemployed	7.194	15.502	0.034	189.362	299
Night-time Lights	4.133	12.79	0	63	354
Accessibility	2.04	0.762	1	4	351
Ruggedness	-39.602	12759.379	-21431	21467	354
Slope Index	73.009	22.623	13	99	351

Panel B: Year 1996

Variable	Mean	St. Dev.	Min	Max	N
Non-state Conflict Events	0.105	0.692	0	6	354
ELP_{WB}	0.34	0.332	0	0.997	350
Blacks	87.97	109.609	0	896.042	354
Population	114.63	139.528	3.557	902.861	354
Rural Population	53.123	73.148	0	404.352	354
No Education	21.674	23.123	0.929	137.231	354
Unemployed	13.403	20.166	0.214	189.748	354
Night-time Lights	4.816	13.26	0	63	354

Panel C: Difference 1991-1996

Variable	Mean	St. Dev.	Min	Max	N
Non-state Conflict Events	-1.044	4.9	-41	6	298
ELP_{WB}	0.002	0.177	-0.889	0.997	294
Blacks	3.541	69.902	-665.71	338.236	298
Population	2.752	78.429	-788.692	347.736	298
Rural Population	-2.697	40.793	-224.642	175.135	298
No Education	-10.822	23.11	-195.972	52.845	298
Unemployed	4.695	12.893	-118.249	78.861	298
Night-time Lights	0.785	3.15	-15	29	298

Notes. Data for Blacks, Population, Rural Population, No Education, Unemployed, Not Economically Active and Citizens of South Africa are in thousands. All variables are discussed in Section 3 (Sources: UCDP-GED v1.5, Statistics South Africa 1991, 1998; Nunn and Puga 2012; IIASA/FAO 2012; NOOA 2012).

B Appendix for Online Publication

B.1 Poisson Model Estimation Results

Table B.1: Poisson Model, 1991 Cross-sectional Estimation

	Total Number of Non-state Conflict Events 1991				
	(1)	(2)	(3)	(4)	(5)
ELP_{WB}	2.542	3.747	3.749	3.829	4.195
	(0.20)	(0.28)	(0.28)	(0.33)	(0.38)
Blacks (log)		0.512	0.517	0.650	0.377
		(0.13)	(0.13)	(0.18)	(0.34)
Population (log)		-0.083	-0.095	-0.196	-3.161
		(0.12)	(0.13)	(0.17)	(1.22)
Night-time Lights (log)			0.005	-0.062	-0.104
			(0.02)	(0.03)	(0.03)
Rural Population (log)					0.046
					(0.02)
No Education (log)					0.416
					(0.39)
Unemployed (log)					1.276
					(0.29)
Constant	-3.105	-6.736	-6.631	-20.738	-21.336
	(0.51)	(0.83)	(0.93)	(1226.57)	(1409.08)
Province Fixed Effects	Y	Y	Y	Y	Y
Geographic Controls	N	N	N	Y	Y
Other Economic Controls	N	N	N	N	Y
Observations	294	294	294	291	291

Notes. Standard errors in parenthesis. The table reports Poisson model coefficients estimates from the 1991 cross-sectional specification. The unit of observation is a MD in South Africa for which information is available in 1991. The dependent variable is the total number of non-state conflict events coded in the MD in 1991 in the UCDP-GED dataset. ELP_{WB} is the district-level within-black polarization measure. Other controls are defined as in the data section (Sources: Statistics South Africa 1991, 1998; Nunn and Puga 2012; IIASA/FAO 2012; NOOA 2012).

Table B.2: Poisson Model, Fixed-Effect Estimation

	Total Number of Non-state Conflict Events 1991		
	(1)	(2)	(3)
ELP_{WB}	5.814	17.129	24.092
	(2.35)	(5.38)	(8.14)
Blacks (log)		-4.149	-5.645
		(1.19)	(1.60)
Population (log)		6.182	4.175
		(1.45)	(4.44)
Night-time Lights (log)		-0.458	-0.456
		(0.27)	(0.29)
Rural Population (log)			-0.073
			(0.08)
No Education (log)			-2.426
			(1.84)
Unemployed (log)			0.097
			(1.21)
Constant	-2.076	-2.108	-4.958
	(0.19)	(0.24)	(2.48)
Other Economic Controls	N	N	Y
Observations	94	94	94

Notes. Standard errors in parenthesis. The table reports coefficients estimates from a fixed-effects Poisson model. With two time periods only (1991 and 1996), estimates from a first-difference linear regression model are equivalent to the ones of a fixed-effects linear regression model. Estimation is performed over those MDs for which a positive number of non-state conflict events is observed in at least one of the two considered year. The dependent variable is the total number of non-state conflict events coded in the MD in the year in the UCDP-GED dataset. ELP_{WB} is the district-level within-black polarization measure. Other controls are defined as in the data section (Sources: Statistics South Africa 1991, 1998; NOOA 2012).

B.2 Conditional Logit Estimation Results

Table B.3: Conditional Logit Estimation

	Migrant from i is observed in j
Distance between i and j (km)	-0.009
	(5.31×10^{-6})
Observations	894'758'631
Pseudo R^2	0.21

Notes. Empirical standard errors in parenthesis. The table reports the estimated coefficient of the distance variable for the Conditional Logit Model in equation (5) in the text. The number of observations is given by multiplying the number of migrants as recorded in the 1996 Census (Statistics South Africa 1998) for the total number of MDs they had the option to locate in (353).

B.3 Governmental Repression as Control

Table B.4: 1991 Cross-sectional Estimation

	Total Number of Non-state Conflict Events 1991				
	(1)	(2)	(3)	(4)	(5)
ELP_{WB}	2.120	2.145	2.155	2.131	2.090
	(0.99)	(1.01)	(1.01)	(1.02)	(1.08)
Gov. Repression	3.335	3.305	3.311	3.285	3.368
	(0.46)	(0.46)	(0.47)	(0.47)	(0.50)
Blacks (log)		-0.032	-0.041	-0.306	-0.111
		(0.39)	(0.39)	(0.44)	(0.48)
Population (log)		0.171	0.199	0.467	-3.658
		(0.48)	(0.51)	(0.55)	(3.65)
Night-time Lights (log)			-0.020	-0.036	-0.051
			(0.11)	(0.12)	(0.12)
Rural Population (log)					-0.030
					(0.12)
No Education (log)					-1.031
					(1.53)
Unemployed (log)					0.373
					(0.69)
Constant	-0.446	-2.015	-2.303	-0.876	-0.287
	(0.68)	(2.96)	(3.38)	(3.72)	(4.56)
Province Fixed Effects	Y	Y	Y	Y	Y
Geographic Controls	N	N	N	Y	Y
Other Economic Controls	N	N	N	N	Y
Observations	294	294	294	291	291
R^2	0.309	0.309	0.309	0.319	0.329

Notes. Standard errors in parenthesis. The table reports Ordinary Least Squares coefficients estimates from the 1991 cross-sectional specification. The unit of observation is a MD in South Africa for which information is available in 1991. The dependent variable is the total number of non-state conflict events coded in the MD in 1991 in the UCDP-GED dataset. ELP_{WB} is the district-level within-black polarization measure. Governmental repression is measured counting in each MD the total number of one-sided conflict events recorded in the UGDP-GED database where the Government is involved against civilians. Other controls are defined as in the data section (Sources: Statistics South Africa 1991, 1998; Nunn and Puga 2012; IIASA/FAO 2012; NOOA 2012).

Table B.5: First-difference Estimation

	Change in Total No. of Non-state Conflict Events 1991-1996				
	(1)	(2)	(3)	(4)	(5)
ΔELP_{WB}	5.175	4.787	4.309	4.309	3.642
	(1.47)	(1.50)	(1.51)	(2.79)	(1.33)
Δ Gov. Repression	3.285	3.296	3.575	3.575	2.560
	(0.45)	(0.45)	(0.46)	(1.58)	(0.42)
Non-state Conf 89-90					-0.233
					(0.03)
Δ Blacks (log)		-0.171	-0.286	-0.286	-0.213
		(0.40)	(0.41)	(0.26)	(0.36)
Δ Population (log)		-0.184	-1.386	-1.386	-4.725
		(0.69)	(4.45)	(7.63)	(3.92)
Δ Night-time Lights (log)		-0.289	-0.254	-0.254	-0.156
		(0.15)	(0.14)	(0.22)	(0.13)
Δ Rural Population (log)			-0.718	-0.718	-0.448
			(0.20)	(0.24)	(0.18)
Δ No Education (log)			-2.143	-2.143	-0.498
			(1.60)	(1.99)	(1.41)
Δ Unemployed (log)			-0.646	-0.646	-0.493
			(0.76)	(0.52)	(0.67)
Constant	-0.586	-0.391	-2.647	-2.647	-0.879
	(0.27)	(0.28)	(1.74)	(1.75)	(1.54)
Other Economic Controls	N	N	Y	Y	Y
Observations	294	294	294	294	294
R^2	0.191	0.204	0.255	0.255	0.430

Notes. Standard errors in parenthesis. The table reports first-difference coefficients estimates. The unit of observation is a MD in South Africa for which information is available in both periods. The dependent variable is the change in total number of non-state conflict events coded in the MD in between 1991 and 1996 in the UCDP-GED dataset. ELP_{WB} is the district-level within-black polarization measure. *Non-state Conf 89-90* is the total number of conflict events coded in the MD in 1989-1990. Governmental repression is measured counting in each MD the total number of one-sided conflict events recorded in the UGDP-GED database where the Government is involved against civilians. Other controls are defined as in the data section (Sources: Statistics South Africa 1991, 1998; NOOA 2012). (1), (2), (3) and (5) are first-difference estimates assuming homoskedastic difference residuals; those in (4) are first-difference estimates assuming heteroskedastic difference residuals.

B.4 Conflict-related Deaths as Alternative Measure of Conflict Incidence

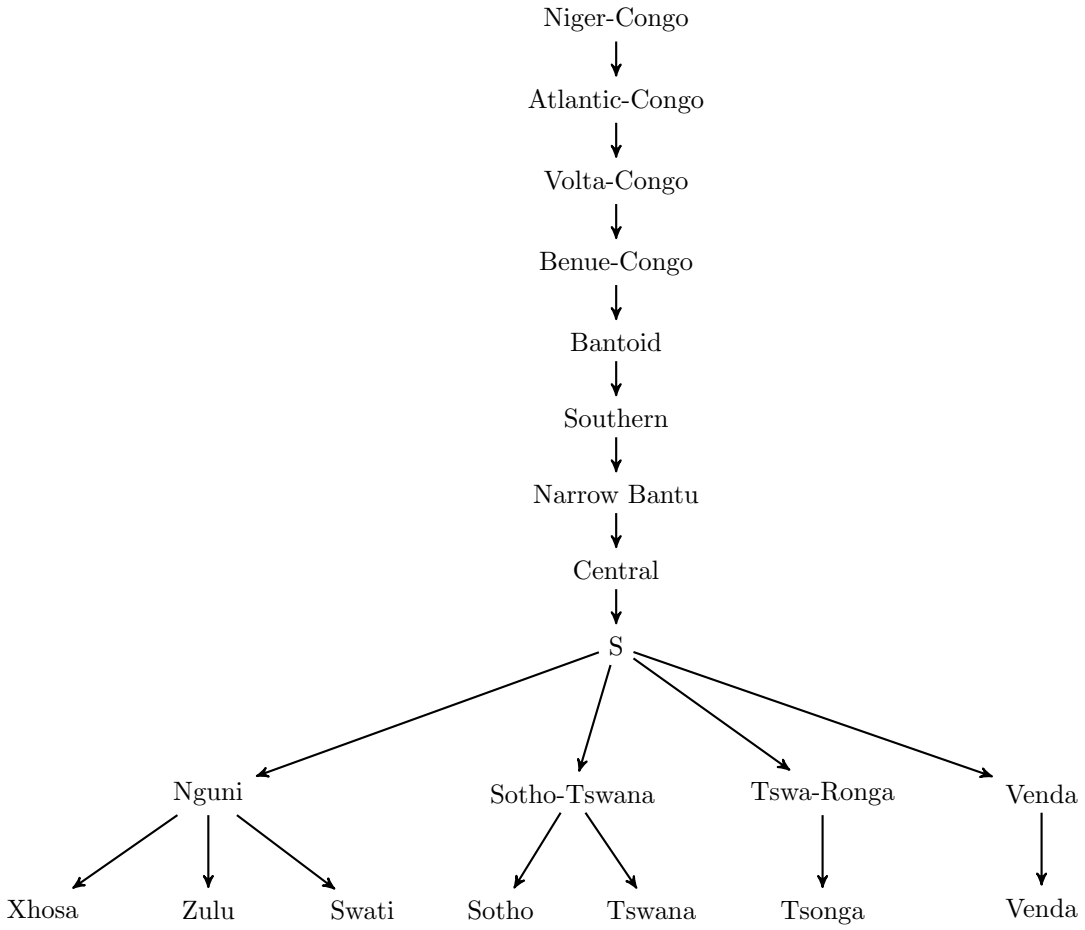
Table B.6: First-difference Estimation

	Change in Total No. of Non-state Conflict Deaths 1991-1996				
	(1)	(2)	(3)	(4)	(5)
ΔELP_{WB}	8.660	8.660	8.259	8.259	5.900
	(3.71)	(4.80)	(3.83)	(4.77)	(3.03)
Est. Deaths 89-90					-0.246
					(0.02)
Δ Blacks (log)	-0.632	-0.632	-0.494	-0.494	-0.490
	(0.98)	(0.45)	(1.03)	(0.50)	(0.81)
Δ Population (log)	0.332	0.332	7.448	7.448	-4.767
	(1.71)	(1.42)	(11.07)	(9.63)	(8.79)
Δ Night-time Lights (log)	-0.834	-0.834	-0.800	-0.800	-0.413
	(0.36)	(0.63)	(0.37)	(0.61)	(0.29)
Δ Rural Population (log)			-0.571	-0.571	0.295
			(0.51)	(0.27)	(0.41)
Δ No Education (log)			-0.238	-0.238	1.314
			(4.04)	(4.29)	(3.19)
Δ Unemployed (log)			1.138	1.138	0.047
			(1.92)	(1.18)	(1.52)
Constant	-1.012	-1.012	-8.046	-8.046	-2.699
	(0.68)	(0.57)	(4.40)	(3.88)	(3.50)
Other Economic Controls	N	N	Y	Y	Y
Observations	294	294	294	294	294
R^2	0.042	0.042	0.056	0.056	0.414

Notes. Standard errors in parenthesis. The table reports first-difference coefficients estimates. The unit of observation is a MD in South Africa for which information is available in both periods. The dependent variable is the change in total number of deaths in non-state conflict events coded in the MD in 1991 and 1996 in the UCDP-GED dataset. ELP_{WB} is the district-level within-black polarization measure. *Est. Deaths 89-90* is the total number of deaths in non-state conflict events coded in the MD in 1989-1990. Other controls are defined as in the data section (Sources: Statistics South Africa 1991, 1998; NOOA 2012). (1), (2), (3) and (5) are first-difference estimates assuming homoskedastic difference residuals; those in (4) are first-difference estimates assuming heteroskedastic difference residuals.

B.5 Ethnolinguistic Information

Figure B.1: The *Ethnologue* Tree of Native Ethnolinguistic Groups in South Africa



Source: Lewis (2009).