

ETHNICITY AND VIOLENCE DURING DEMOCRATIC TRANSITIONS: EVIDENCE FROM SOUTH AFRICA

Francesco Amodio
McGill University

Giorgio Chiovelli
London Business School

Abstract

This paper shows that ethnic diversity within the disenfranchised majority is a strong correlate of the incidence of violence during democratization. We study the relationship between ethnicity and conflict in South Africa during the fall of apartheid. Migration flows following the implementation and repeal of apartheid segregation laws induce cross-sectional and time variation in the ethnic composition of districts. Using Census data from the years before and after democratization, we compare the evolution of conflict across districts experiencing differential changes in ethnic composition. We find that ethnic polarization and inequality within the black majority correlate strongly and positively with the incidence of armed confrontations between black-dominated groups. Results suggest that during democratic transitions ethnic markers can become a salient technology to separate individuals into well-identified groups and mobilize them for political violence. (JEL: D74, J15, N47, N97, O15, P48, R23)

1. Introduction

Transitions to democracy are often violent. Conflict plagued almost half of all democratic transitions that occurred between 1973 and 2000 (Karatnycky 2005). Existing studies of democratization conceptualize violence as manifestation of the

The editors in charge of this paper were Nicola Gennaioli and Paola Giuliano.

Acknowledgments: We thank the Editors Nicola Gennaioli and Paola Giuliano, and five anonymous referees for their insightful comments and suggestions. We are indebted to our advisors Matteo Cervellati and Albrecht Glitz for their guidance and support. We received valuable comments from Jean-Marie Baland, Laia Balcells, Marcelo Bérigolo, Bruno Caprettini, Francesco Caselli, Rodrigo Deiana, Giovanni Di Mauro, Oeindrila Dube, Ruben Enikolopov, Joan Esteban, Margherita Fort, Oded Galor, Nicola Gennaioli, Andrea Ichino, Laura Mayoral, Stelios Michalopoulos, Massimo Morelli, Hannes Mueller, Elias Papaioannou, Giacomo Ponzetto, Giovanni Prarolo, Debraj Ray, Marta Reynal-Querol, Simon Roberts, Dominic Rohner, Martín Rossi, Uwe Sunde, Alessandro Tarozzi, Mathias Thoenig, John Tsoukalis, Fabrizio Zilibotti, and seminar participants at Università di Bologna, Universitat Pompeu Fabra, Toulouse School of Economics, University of Lausanne, University of Zurich, Brown University, the Barcelona GSE “Understanding Civil Conflict” Summer Forum, 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.

E-mail: francesco.amodio@mcgill.ca (Amodio); gchiovelli@london.edu (Chiovelli)

struggle for power between a small ruling elite and a large disenfranchised majority, regarding the latter as a unitary, homogeneous group (Acemoglu and Robinson 2006). This paper investigates the role of divisions within the disenfranchised majority. Once the transition begins, disenfranchised groups struggle among themselves to achieve dominance. The ability of political parties to mobilize individuals for violence determines the extent to which political confrontations escalate into violent conflict.

Ethnic markers have the potential to separate individuals into well-identified groups. In their struggle for power, political parties can use ethnicity as a tool to mobilize individuals for violence. We therefore hypothesize that, as democratization unfolds, societies that are more ethnically diverse are more prone to experience conflict.

We test this hypothesis in the context of South Africa through the end of the *apartheid* regime. There are several reasons why South Africa represents an ideal setting to explore the relationship between ethnicity and conflict during democratization. First, it is a newly established democracy, with the first democratic elections held in 1994. Second, democratization was violent. The vast majority of armed confrontations involved black-dominated parties and unions that struggled among themselves to benefit from the transition and dominate the new institutional scenario. Moreover, violence was not evenly distributed within the country. Third, above and beyond the separation between the white elite and the black majority, the newly enfranchised black population was highly heterogeneous along ethnic lines, which allows us to explore this specific dimension. Fourth, the implementation and repeal of apartheid segregation laws induced cross-sectional and time variation in ethnic composition of districts. We can therefore study the relationship between ethnic diversity and conflict before, during, and after democratization, analyzing changes over time. This rules out the possibility that time-invariant unobserved characteristics at the district level - such as geographical characteristics - determine both ethnic composition and conflict incidence, generating a spurious correlation between the two.

Our analysis shows that the changes in ethnic composition that occurred in the democratization period correlate significantly with the evolution of conflict incidence at the local level. As the apartheid-era restrictions on internal mobility were lifted, 2.5 million blacks moved out of segregated districts and relocated in others.¹ We show that a strong relationship exists between the resulting changes in ethnic composition of districts and the pattern of political violence in the period. In particular, ethnolinguistic polarization within the blacks at the district level is positively correlated with the number of armed confrontations between black-dominated organized groups. This result holds across several regression model specifications and identification strategies. First, it is robust to the inclusion of controls

1. Source: 1996 Population Census of South Africa (Statistics South Africa 1998). This is the estimated number of individuals who moved from one magisterial district to another between 1991 and 1996 (see Section 3). This corresponds to a district-level outflow rate of 8.66%. Reed (2013) reports very similar estimates using data from the South Africa Migration and Health Survey (SAMHS).

capturing the demographic and economic characteristics of districts. Importantly, the extent of immigration per se has no systematic relationship with conflict incidence, but rather matters only when it changes the local ethnic distribution. Second, the relationship between changes in ethnic diversity and conflict still holds within clusters of neighboring districts. This undermines the concern that unobserved characteristics at a finer geographical level may account for the result. Third, we dig deeper in the history of South Africa to implement an identification strategy based on instrumental variables. We recover information on the ethnic distribution of districts before the full implementation of segregation laws, and use it to predict the change in polarization that occurred in the democratization period. The extent of historical return migration determines the strength of our instrument and its scope for identification. Results still support the hypothesis of a positive relationship between ethnic diversity and conflict during democratization.

Several pieces of quantitative and qualitative evidence support the hypothesis that, during the democratic transition, ethnicity became a salient trait for the separation of individuals in society. First, the measure of ethnic diversity that predicts conflict is the polarization index. In their theory of ethnic conflict, Esteban and Ray (2011b) identify the polarization index as capturing the *degree of publicness* of the disputed *prize*. This is consistent with our conceptualization of the struggle within the black majority in South Africa during democratization as a dispute for political power, a prize whose nature is intrinsically public. Second, the narratives from the period reveal how ethnicity suddenly became salient for the political struggle at the onset of the democratic transition, while it was not in the period before or after. Third, ethnic diversity maps into diverse support for black political parties. Using electoral data from the first democratic elections, we show that a positive correlation exists between ethnolinguistic polarization within the black majority and political polarization across major black political parties.

Our interpretation of results rest on the assumption that incentives to achieve political power were highest during democratization, also at the local level. This is the period when political parties negotiated the new form of government, and local institutions were held responsible for the implementation of key policies that were to shape the future of South Africa. We therefore expect the salience of ethnicity as a separation and mobilization technology to be higher in those districts where achieving control over policy implementation is more valuable. To test this hypothesis, we focus on the issue of land reform and redistribution. We augment the main conflict regression specification with the interaction between within-black ethnolinguistic polarization and a district-level measure of agricultural suitability. We find that the estimated coefficient of the interaction term is positive and significant. This indicates that the salience of ethnic diversity for conflict is higher in those districts where land is more suitable for agriculture, and incentives to achieve control over its redistribution are higher.

Above and beyond ethnic composition, we also investigate the role of ethnic inequality. Conditional on the relative size of groups, income differences shape the incentives to mobilize and be mobilized for political violence. Building on Alesina, Michalopoulos, and Papaioannou (2016), we develop an index that captures heterogeneity in the fraction of top earners across ethnic groups. Conditionally on the level of within-black ethnolinguistic polarization, we find that ethnic inequality at the district level correlates positively with the local incidence of conflict during democratization.² Moreover, the estimated coefficient of the interaction between ethnic polarization and inequality is negative and significant. We interpret this as suggestive evidence that income inequality across groups matters positively for conflict incidence, and more so when groups are unequal in both income and relative size. For instance, conflict incidence is high when a large poor majority faces a small rich minority.

To the best of our knowledge, ours is the first paper to investigate the salience of ethnic diversity in explaining the incidence of violence during democratic transitions. We bring together and thus contribute to several strands of the literature. First, we contribute to the empirical literature on democratization. Papaioannou and Siourounis (2008) and Acemoglu, Naidu, Restrepo, and Robinson (2016) show that democracy has a significant and robust positive effect on per capita GDP growth and levels respectively. But, violent transitions to democracy correlate with subsequent lower growth, lower degree of civil liberties, and higher incidence of civil conflict (Cervellati, Fortunato, and Sunde 2012, 2014; Sunde and Cervellati 2014; Cervellati and Sunde 2014). We contribute to this debate by showing that ethnic diversity and inequality within the disenfranchised are two strong correlates of violence during democratization. Our analysis suggests that, apart from the struggle between the ruling elite and the majority, the study of violence within the latter and its determinants can contribute to our understanding of why some democratic transitions are violent and others are not.

We also contribute to the literature on ethnic conflict. Esteban and Ray (1994, 1999) argue that polarized societies - comprising internally homogeneous but externally very distant groups - are more prone to experience conflict. Ethnic lines become salient as they allow for separation of individuals into groups that are homogenous under one trait (ethnicity), but significantly different along other dimensions (Esteban and Ray 2008, 2011a). Caselli and Coleman (2013) conceptualize ethnicity as a technological device that prevents indiscriminate access to the expected gains of the winning group. Montalvo and Reynal-Querol (2005) provide the first empirical evidence showing that ethnic polarization correlates with the probability of conflict outbreak and incidence. Finally, Esteban and Ray (2011b) develop an extensive theory of conflict that highlights the role of different ethnic distribution measures. The

2. Using a rainfall-based measure of inequality between ethnic groups at the country level, Guariso and Rogall (2017) show that higher ethnic inequality is positively correlated with the incidence of ethnic civil conflict.

empirical evidence in Esteban, Mayoral, and Ray (2012) supports their theoretical arguments. More generally, this paper builds upon a growing body of the literature that studies conflict and violence at a disaggregated level.³

Our contribution to the literature on ethnic conflict is twofold. First, the evidence in this paper suggests that the prospect of regime change can be a mediating factor in the relationship between ethnic diversity and conflict: the salience of ethnicity for conflict can change with the incentives to achieve political power and mobilize individuals for violence. This is consistent with a strand of literature in political science that shows how ethnic traits are redefined and mobilized accordingly depending on the incentives of the relevant actors in society, and the institutional framework in which they operate (Posner 2005; Eifert, Miguel, and Posner 2010).

Second, to the best of our knowledge, ours is the first paper that exploits variation over time in the ethnic composition of districts and investigate its relationship with the evolution of conflict through time.⁴ Our analysis in first differences exploits the variation in ethnic distribution originated by the internal migration of South Africans, allowing us to control for all time-invariant characteristics at the district level that could confound the relationship between ethnic diversity and conflict. Our results show that this relationship holds true in differences and thus is not attributable to such confounders.

The paper is organized as follows. In Section 2, we provide a short overview of the history of contemporary South Africa. Section 3 introduces the data and measures we use, together with preliminary evidence. The main empirical results are presented in Section 4. Section 5 concludes.

2. South Africa and the Democratic Transition

The history of South Africa's democratic transition traces back to the *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. The apartheid legislation identified four different races living in the South Africa: black, white, indian, and

3. See for example Rohner, Thoenig, and Zilibotti (2013), Dube and Vargas (2013), Besley and Reynal-Querol (2014), Harari and La Ferrara (2016), Berman et al. (2017), Michalopoulos and Papaioannou (2016), Bazzi and Gudgeon (2016).

4. Existing studies of ethnic conflict make use of time-invariant ethnic distribution indices, drawn from historical and encyclopedic sources (Montalvo and Reynal-Querol 2005, 2010; Desmet, Ortuño Ortín, and Wacziarg 2012). One exception is Novta (2016), who uses Census data from Bosnia to derive ethnic distribution measures at the municipality level. She studies the relationship between the latter and the spread of civil conflict, but considers the ethnic distribution of municipalities in 1991 only and does not investigate changes over time.

colored. The black population was further classified into native ethnicities. The long-term goal of apartheid was the complete separation between a white Republic of South Africa and a number of independent black states, one for each black ethnolinguistic group. In order to achieve this goal, the Bantu Authorities Act (1951) and Bantu Resettlement Act (1954) established ten black ethnicity-based homeland reserve areas known as Bantustans.⁵ In less than thirty years, 3.5 million blacks were forced to relocate to the homelands, and the government enforced strict control over the mobility of blacks within the country (Clark and Worger 2011). By mid 1980s, the apartheid system began to collapse. This was mainly due to the internal contradictions and conflicting effects of its policies (Acemoglu and Robinson 2006). With the UN embargoes in the late 1980s, black opposition parties started to foresee the possibility of a regime change.

At the eve of the democratic transition, several black parties vied for power. The African National Congress (ANC), founded in 1912, was the historical black party. Until the state banned it in 1950s, the ANC represented almost all the black population. Later on, the exile of ANC leaders and the party alignment with the socialist cause created the space for new political representation on the ground. The Pan Africanist Congress (PAC) formed in 1953, and the United Democratic Front (UDF) arose in 1983. Mangosuthu Buthelezi, a former member of ANC and member of the Zulu royal family, founded the Inkatha Freedom Party (IFP) in 1975. Until the 1979 London Congress, opposition to apartheid was fairly cohesive. With the beginning of the end of the apartheid regime and the rising prospects of democratization, the alliance broke up and confrontations started among black-dominated groups. The struggles between the IFP and both the ANC and UDF were the most violent. The IFP drew support from the traditional Zulu chiefly structures, and ran on a strongly conservative political platform closely aligned with large business interests in the KwaZulu region (Carver 1996). Most urban youth and working class supported instead the ANC or UDF, drawn by their vision of a non-racial unitary state in contrast with traditional chief power.

In the late 1980s, the white government decided to repeal several apartheid laws. The turning point for the democratization process was the unbanning of the ANC and the liberation of its leaders in 1990. Between 1990 and 1991, the government restored free mobility of blacks within South Africa. In the meanwhile, representatives of all major political parties took part in several rounds of negotiations over the new constitutional framework. While the ANC strongly supported the creation of a centralized state,

5. Figure A.1 in Appendix A shows the map of Bantustans in South Africa as defined by the apartheid legislation as of 1986. These were: Transkei and Ciskei (Xhosa ethnicity), Bophuthatswana (Tswana), Venda (Venda), Gazankulu (Tsonga), Lebowa (Sotho), Qwaqwa (Sotho), KwaZulu (Zulu), KaNgwane (Swazi) and KwaNdebele (Sotho). 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.

both the ruling white National Party and the IFP favored a decentralized form of government. From 1990 to 1993, several meetings trying to reach an agreement failed. An interim constitution setting electoral rules was approved in November 1993. In December 1993, the IFP announced that it would not take part in the elections because of omissions from the interim constitution including the absence of a clear acknowledgement of the role of traditional leaders and the Zulu king. Conflict between the ANC and the IFP reached its climax. Deadly violence ensued in Creighton, KwaMashu, and Ndwedwe as the IFP substituted violence for voting against the ANC. The IFP decided to enter the election just seven days before the vote in April 1994.

The 1993 constitutional arrangements held South African provinces responsible for such crucial matters as agriculture, roads and transportation, and welfare. Decentralization made the Magisterial District a key seat of power.⁶ Transitional Rural Councils were established at the Magisterial District level and held responsible for the allocation of land.⁷ This deepened the confrontations between the ANC, which favored land reform, and the IFP, which favored preserving the power of traditional chief leaders to allocate land.

The process of democratic transition and decentralization raised the incentives for black political parties to achieve control of local level institutions. Ethnicity was a readily available technology to separate individuals into well-identified groups (Esteban and Ray 2008, 2011a; Caselli and Coleman 2013). Ethnic lines suddenly became salient for the expression of political tensions. Anecdotal evidence shows how the perception of ethnicities changed in the transition scenario:

“Who are you?” “We are Zulus,” they reply. “What are you doing here? Where are all the other people?” “They ran away. They were Xhosa and Shangaans [Tsonga]. We live here now, so there is no trouble.” “What trouble?” “This thing about Inkatha and ANC. For years we lived here. We were Zulus and Shangaans and Xhosa all mixed up but now we are fighting and I don’t know why”. (Sunday Tribune, April 1992)

6. The Magisterial District is a sub-provincial territorial unit defined by the judicial system under the administration of the Department of Justice and Constitutional Development. The 1994 elections structured the voting districts around these Magisterial Districts.

7. The Transitional Rural Council (TRP) and the Transitional Representative Council (TRPC) were the provisional local municipalities from 1994 to 2000. The implementation of the 2000 municipal system substituted both TRPC and TRP with the actual local municipalities.

3. Data

3.1. Sources and Measurement

Our database combines several different data sources. The geographical unit of analysis is the Magisterial District (MD). Figure A.2 in Appendix A shows a map of South Africa with the district boundaries. Pre-1994 demarcations still inform these boundaries. This makes the MDs particularly valuable as units of comparative analysis on a small-scale geographical basis, as all other administrative boundaries have been re-demarcated numerous times in the 1990s (Case and Deaton 1999; Hoffman and Todd 2000). We exclude those MDs that the 1991 Census of the Republic of South Africa did not cover as they belonged to independent Bantustans. Our final sample consists of the 294 districts for which information is available in both the 1991 and 1996 Census.

Part of our empirical analysis also covers the democratic period, focusing on the years 2001 through 2011. The 2011 Census data do not report the MD of residence of interviewed households. Therefore, we use the Local Municipality (LM) as alternative unit of analysis. LMs became the official most local tier of government after 2000, substituting the Transitional Councils that were previously established at the MD level. Their boundaries stayed reasonably constant thereafter.

Ethnicity. We derive information on the ethnic composition of districts and its changes using South Africa Census data (Statistics South Africa 2011, 1991, 1998, 2004, 2014). We identify individuals belonging to different ethnolinguistic groups according to the first language they speak. The African native groups in the database are Swazi, Xhosa, Zulu, Sotho, Tswana, Tsonga, and Venda.⁸ For each district and Census year, we count the total number of individuals in each group and compute the relative share of each group within the black majority. We then use these within-black shares to compute the relevant measures of ethnic distribution. We first follow Esteban and Ray (2011b) and Esteban, Mayoral, and Ray (2012), and focus on three different distributional indexes of ethnic diversity: the *polarization* index P , the

8. Desmet, Ortuño Ortín, and Wacziarg (2012) show how to use the genealogical relationships between world languages in the construction of distribution indexes. In line with their approach, we use the information contained in the *Ethnologue* linguistic database (Lewis 2009), assuming that all languages at the same level are equally distant from the proto-languages of their respective families. We build our measure considering black ethnolinguistic groups in South Africa that correspond to level 11 in the world language tree and reported in Figure C.1 in the Online Appendix. The *Ethnologue* database has been used as source of ethnolinguistic information also by Alesina, Devleeschauwer, Easterly, Kurlat, and Wacziarg (2003), Desmet, Ortuño Ortín, and Weber (2009), and Esteban, Mayoral, and Ray (2012).

fractionalization index F , and *Greenberg-Gini* index G . They are defined as

$$P = \sum_{i=1}^m \sum_{j=1}^m n_i^2 n_j d_{ij} \quad F = \sum_{i=1}^m n_i (1 - n_i) \quad G = \sum_{i=1}^m \sum_{j=1}^m n_i n_j d_{ij}$$

where n_i is the population share of ethnolinguistic group i and m is the overall number of groups, while d_{ij} is a measure of distances between groups.⁹ Esteban and Ray (2011b) show the relationship between the equilibrium intensity of conflict and these three distribution measures. In particular, their theory posits that the degree of publicness of the disputed prize determines the relative salience of P as opposed to F in explaining conflict incidence.

We also employ the *binary* version of the polarization index as implemented in Reynal-Querol (2002) and Montalvo and Reynal-Querol (2005, 2010), which falls within the class of polarization measures proposed by Esteban and Ray (1994). This is computed as

$$RQ = 1 - \sum_{i=1}^m \left(\frac{1/2 - n_i}{1/2} \right)^2 n_i = 4 \times \sum_{i=1}^m n_i^2 (1 - n_i)$$

The difference between the RQ and the P index is that the former considers all inter-group distances d_{ij} to be binary. It thus returns a distribution measure linked to the generation of social tension between equally distant groups. It is worth noticing that, when groups have different sizes, changes in the population size of a given group do not map systematically into changes in the polarization index. Esteban and Ray (1994) provide a series of examples illustrating this point.

Finally, we refer to Alesina, Michalopoulos, and Papaioannou (2016) in building a measure of *ethnic inequality*. This is a Gini index that captures the extent to which ethnic groups differ in their relative income. We derive it by focusing on the fraction of top earners in each ethnic group, as derived from Census data. Let y_h be the fraction of individuals in ethnolinguistic group h that belongs to the top 10% of the overall black income distribution. Let the group identifier $h = 1$ to m be indexed in non-decreasing order ($y_h \leq y_{h+1}$). The ethnic inequality index EI is computed as

$$EI = \frac{1}{m} \left(m + 1 - 2 \frac{\sum_{h=1}^m (m + 1 - h) y_h}{\sum_{h=1}^m y_h} \right)$$

Unlike the polarization index, the value of the EI index is independent of the relative size of each group.

9. For a detailed discussion of these indexes please refer to Esteban and Ray (2011b). In the computation of P , we follow Esteban, Mayoral, and Ray (2012) and use $d_{ij} = 0.05$. Results are robust to the choice of d_{ij} .

Conflict. We measure conflict incidence using the Geo-referenced Event Dataset of the Uppsala Conflict Data Program (UCDP-GED v1.5).¹⁰ 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.¹¹ For each conflict event, the data provide information on the date, place of the event (with coordinates), actors participating, and estimates of fatalities.¹² We measure the yearly incidence of armed confrontations between black-dominated organized groups in South Africa at the district level by counting the number of related geo-referenced conflict events in each district per year. These clashes all belong to the category of non-state conflict.¹³

The UCDP-GED dataset records very few non-state conflict events in South Africa after 2000, and none of them in 2001 and 2011. We thus measure conflict incidence in these years using the Armed Conflict Location and Event Data Project (ACLED) database (Raleigh, Linke, Hegre, and Karlsen 2010). The dataset covers the period from 1997 onwards. It reports detailed geo-referenced information on all reported political violence events in over 60 developing countries in Africa and Asia. It records also episodes of political violence that take place outside the context of a civil war, such as violence against civilians, militia interactions, communal conflict, and rioting. It draws information on each event from a variety of sources, including reports from developing countries and local media, humanitarian agencies, and research publications. Although the nature of events can differ from UCDP-GED, ACLED data bring about substantial variation in conflict incidence between the years 2001 and 2011.¹⁴

10. 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 latest data presentation.

11. *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 organized armed groups, neither of which is the government of a state, that 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 organized group against civilians that results in at least 25 deaths in a year in the period (Sundberg, Lindgren, and Padskocimaite 2010).

12. In the study of the relationship between ethnic polarization and conflict, the literature has used conflict data from the Peace Research Institute Oslo (PRIO) (Montalvo and Reynal-Querol 2005, 2010; Esteban, Mayoral, and Ray 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 used to test the robustness of results.

13. Table A.1 in Appendix A shows the total number of One-sided and Non-state conflict events per year that can be mapped into magisterial districts, together with the estimated total number of deaths. One-sided conflict events refer to violence perpetrated by organized groups (including the Government of South Africa or the Afrikaner Resistance Movement (AWB) white supremacist movement) against civilians.

14. Please refer to Raleigh, Linke, and Dowd (2014) for further information. See also Eck (2012) for a complete discussion of the UCDP-GED database and its comparison with the ACLED database.

Political Parties. In order to provide direct evidence of the link between ethnolinguistic diversity and political mobilization, we use data on election results from the 1994 South Africa general elections (Alvarez-Rivera 2004). The data are disaggregated at the district level. Out of the 298 MDs surveyed in the 1991 Census, we are able to map 270 of them into 1994 electoral districts. Given our interest in political mobilization within the black majority, we focus on black-dominated parties. We consider those parties that conquered at least one seat in the National Assembly. These are the ANC, IFP, PAC, and the African Christian Democratic Party (ACDP). Knowing the number of votes for each party, we can derive a measure of within-black political polarization at the district level using the binary version of the polarization index explained before.

Socio-economic and Geographical Controls. Using Census data, we derive a number of additional socio-economic variables by aggregating information at the district level. 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 Census, conflict, and election data, we rely on several other sources to derive other variables such as night-time lights as proxy for local economic conditions, terrain ruggedness and slope, accessibility as proxied by travel time to the nearest big city, and local suitability for agriculture. We discuss the sources of these data in Appendix B.

3.2. Descriptive Evidence

Table 1 shows the total population sizes of each population group in our sample according to the 1991 and 1996 Census, and their difference over time. It does so for the seven black ethnolinguistic groups, and for the main population groups as defined in the Census: whites, blacks, asians, and colored.¹⁶ The table shows evidence of a substantial inflow of individuals belonging to those black groups that were previously segregated in independent Bantustans. 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. According to 1996 Census data, 2.5 million blacks moved from one district to another between 1991 and 1996. Figure A.3 in Appendix A shows the change in the population share of the three biggest ethnolinguistic groups at the district level between 1991 and 1996 plotted against the share of black population in 1991. It is worth noticing that the most

15. In our regression specification we use the logarithm of these variables, augmenting all values by 0.01. Results go through other variable specifications and level shifts.

16. Our study sample excludes those districts that were part of independent Bantustans in 1991. Table A.3 in Appendix A shows the same totals, but calculated over all districts in the 1991 and 1996 Census. Not surprisingly, the inclusion of former independent Bantustan territories in the 1996 Census brings about a dramatic increase in the population stocks of the corresponding ethnolinguistic groups.

TABLE 1. Group population sizes.

	TOTAL POPULATION		DIFFERENCE 1991–1996
	1991 Census	1996 Census	
Xhosa	2,493,382	2,570,752	77,370
Zulu	8,416,125	8,579,680	163,555
Sotho	6,414,684	6,648,697	234,038
Swazi	943,989	989,354	45,365
Tswana	1,437,660	1,900,704	463,044
Tsonga	1,450,874	1,533,665	82,791
Venda	116,533	118,829	2,296
Black	21,642,682	22,697,442	1,054,760
White	5,065,193	4,398,995	-666,198
Asian	985,317	1,038,281	52,964
Colored	3,259,250	3,313,530	54,280

Notes: The table reports the group population sizes in South Africa in 1991 and 1996, together with their difference. Totals are computed over all those districts for which information is available in both the 1991 and 1996 Census, and are thus part of our study sample (Sources: Statistics South Africa 1991, 1998).

relevant changes do not seem to be concentrated in those districts where the overall share of blacks in 1991 was either negligible or close to one.¹⁷

Heterogeneity in black group population sizes and their changes map into variation in within-black ethnic distribution measures. Figure A.5 in Appendix A shows the changes in the distribution of within-black polarization P and fractionalization F over time across districts. According to the results of the Kolmogorov-Smirnov test of equality of distributions, and for both measures, we cannot reject the hypothesis that the 1991 and 1996 distributions are equal. Indeed, Table A.2 in Appendix A shows that both the averages and standard deviations of both P and F are very similar across the two years. Nonetheless, evidence shows meaningful variation in the two measures within districts. Figure 1 plots the 1996 values of within-black polarization and fractionalization measures in each district against their corresponding values in 1991. Those points located away from the 45-degree line are districts where the ethnic composition changed between 1991 and 1996. The time difference is as large

17. See also Figure A.4 in Appendix A for further analysis of migration patterns in the period.

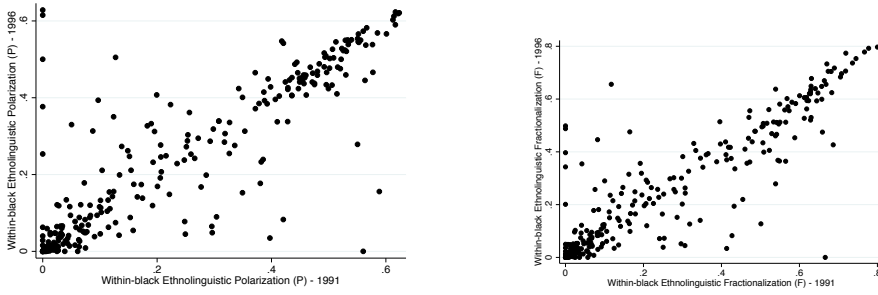


FIGURE 1. Ethnic diversity over time. The figures plot the levels of within-black ethnolinguistic polarization P and fractionalization F in 1996 against their value in the same district in 1991. The farther points are from the 45 degree line, the larger the change in ethnic composition over time (Sources: Statistics South Africa 1991, 1998).

as 0.6 for some districts. The standard deviation of changes over time is as high as half of the 1991 cross sectional standard deviation of both measures. Moreover, the values of polarization and fractionalization in 1991 explain no more than 10% of the variation in their corresponding changes between 1991 and 1996, so that there is enough independent variation between levels and changes in the two. In the same period, average ethnic inequality decreases by 30%, but also in this case the change is highly heterogeneous, with the index dropping by more than 0.7 in some districts while rising by 0.59 in others.

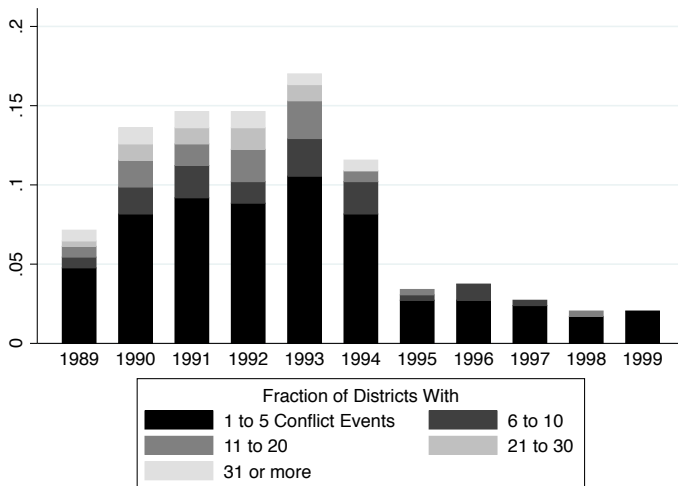


FIGURE 2. Distribution of conflict over time. The figure shows the evolution of the distribution of the total number of Non-state conflict events across districts in South Africa as recorded in the geo-referenced Event Dataset of the Uppsala Conflict Data Program - UCDP-GED v1.5.

In the period 1989-1996, we count more than 2,000 armed confrontations between black-dominated organized groups. Approximately 85% involved the ANC in conflict with the IFP. The remaining 15% involved the UDF opposing the IFP or the ANC Greens faction conflicting with the ANC Reds faction. Table A.2 in Appendix A shows that the average number of conflict events per district is 1.2 in 1991, with large cross-district variation. The same average is 0.1 in 1996, but with a cross-district variation of 0.7. The results of the Kolmogorov-Smirnov test of equality of distributions lead us to reject the hypothesis that the 1991 and 1996 distributions are equal. Figure 2 shows the evolution of the distribution of the number of these conflict events across districts in South Africa from 1989 to 1999. The fraction of districts in which at least one event occurred increases from 1989 onwards, reaching a height in 1993. In the year prior to the first democratic elections, 50 different districts experienced at least one conflict event. Conflict prevalence decreases thereafter, but some districts still experience as many as 10 conflict events in 1996. Overall, from 1989 to 1999, the data record at least one conflict in 72 different districts. Conflict prevalence is highest in Kwa-Zulu Natal and Gauteng. These are the provinces with the highest within-black share of Zulus, and thus support to the IFP.

3.2.1. Basic Correlations. Table A.4 in Appendix A reports the correlation structure between the main variables of interest in 1991 and 1996 in levels (Panel A and B) and changes (Panel C). The table shows the estimated linear correlation coefficients with p-values in parenthesis. Within-black polarization P and fractionalization F are highly positively correlated in both levels and changes. However, there is residual variation to exploit. For example, in 1991, the magisterial districts of Reddersburg, Barkly-East, Westonaria, and Oberholzer all have values of P between 0.35 and 0.4. However, the values of F in the same districts ranges from 0.28 to 0.8. The same holds for the ethnic inequality measure EI . The correlation with P and F is negative and significant in levels. Nonetheless, while for example the districts of Virginia, Jacobsdal, and Clanwilliam all have values of P between 0.5 and 0.55, their EI values range from 0.07 to 0.77. Notice also that the correlation between changes in EI and changes in P and F is close to zero and insignificant.¹⁸

The correlation between conflict and ethnic distribution and inequality measures is close to zero and insignificant in both 1991 and 1996. On the contrary, we estimate a positive and significant correlation in 1991 between conflict and observable characteristics at the district level such as the size of the black population, overall population size, and economic development as measured by night-time lights. Consistent with Michalopoulos (2012), Table A.4 in Appendix A shows that agricultural suitability is negatively and significantly correlated with ethnic diversity, but positively and significantly correlated with conflict incidence. This is an example

18. Table C.1 in the Online Appendix shows the correlation between P , F , EI , and conflict across periods.

of how time-invariant characteristics at the district level may confound the estimated relationship in levels between ethnic diversity and conflict. When looking at changes over time, we find a very different pattern. The correlation between conflict and ethnic distribution and inequality measures becomes positive and highly significant. For example, the districts of Richmond and Port Shepstone (both in Kwa-Zulu Natal) are among the 10 districts that experienced the biggest decrease in within-black polarization, and are also among the 10 districts that experienced the biggest drop in conflict incidence. Similarly, the district of Mooi River, one of those with the highest decrease in conflict incidence, also experienced one of the biggest declines in ethnic inequality. In what follows, we adopt a regression framework to analyze these patterns of correlation in a more systematic way.

4. Regression Analysis

4.1. Cross-sectional Estimates

We start by examining cross-sectional patterns and investigate the link between ethnic distribution and conflict separately for the years 1991 and 1996. We follow Esteban, Mayoral, and Ray (2012) and adopt the following regression specification

$$\text{conf}_{ipt} = \theta_p + \beta P_{ipt} + \gamma F_{ipt} + \delta G/N_{ipt} + \mathbf{Z}'_{ip}\omega + \mathbf{X}'_{ipt}\varphi + u_{ipt} \quad (1)$$

where conf_{ipt} is the number of non-state conflict events recorded in district i in province p in year t . P_{ipt} , F_{ipt} and G/N_{ipt} are the within-black distribution indexes in the same district and year, with N being the size of the black population in thousands. \mathbf{Z}_{ip} is a set of time-invariant geographical characteristics (ruggedness, slope index, accessibility, agricultural suitability). \mathbf{X}_{ipt} is the vector of time-varying demographic and economic controls. θ_p captures province fixed effects, netting out average differences across provinces. The residual u_{ipt} captures those remaining unobserved factors that affect conflict incidence.

Panel A and B of Table 2 show the corresponding set of cross-sectional estimates for the three distribution indexes. Columns (1) and (2) report ordinary least squares estimates. Columns (3) and (4) report negative binomial estimates, which suit the count nature of the dependent variable better. To analyze also the extensive margin of conflict, in columns (5) and (6) we replace as dependent variable a dummy equal to one if the database records at least one conflict event in the district in the year, and estimate a linear probability model. In columns (7) and (8), we take into account the skewness of the outcome distribution and replace as dependent variable the logarithm of the conflict count (augmented by 0.01). Finally, columns (9) and (10) report estimates using as dependent variable the cumulative number of conflict events in

the three years prior to the Census. For each model estimation, we add a first set of geographical and socio-economic control variables in the second column.¹⁹

The estimated coefficients in Table 2 are mostly insignificant at the standard levels. Cross-sectional estimates do not support the hypothesis that a systematic relationship exists between within-black ethnic diversity and conflict at the district level. This is consistent with the basic correlation structure discussed in the previous Section.

4.2. First-difference Estimates

As shown in Section 3, the restoration of free internal mobility of blacks induced meaningful variation over time in the ethnolinguistic composition of districts. We can thus implement a first-difference strategy, where we look at the evolution of both within-black ethnolinguistic diversity and conflict over time across districts. We implement the following regression specification

$$\begin{aligned} \Delta conf_{i96-91} = \\ \alpha + \beta \Delta P_{i96-91} + \gamma \Delta F_{i96-91} + \delta \Delta G/N_{i96-91} + \Delta \mathbf{X}'_{i96-91} \varphi + \Delta u_{i96-91} \end{aligned} \quad (2)$$

where $\Delta conf_{i96-91}$ is the change in the number of recorded non-state conflict events in district i between year 1991 and 1996, while ΔP_{i96-91} , ΔF_{i96-91} and $\Delta G/N_{i96-91}$ are the corresponding changes in the within-black ethnic distribution indexes. The constant term α captures the effect of nationwide events (such as democratization in 1994) and general time trends. 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_{i96-91} captures those unobserved additional factors which affect the change in conflict incidence. The proposed first-difference specification allows to cancel out both observable and unobservable time-invariant characteristics at the district level.

Table 3 shows the corresponding set of estimates. We order model specifications as in Table 2. Given that we only have two periods, the first-difference estimates from a linear model are equivalent to those obtained from a level specification with district fixed effects. This is not the case for non-linear models, so that in columns (3) and (4) we report negative binomial fixed effects estimates. Also, the district fixed effects now absorb the province fixed effects and geographical controls included in the cross-sectional model. We thus include only time-varying socio-economic controls as additional regressors in the second column of each model specification.

All estimates consistently support the hypothesis that changes in the within-black ethnic composition of districts correlate with the evolution of conflict. This is again

19. Geographical controls are: ruggedness, slope index, accessibility, agricultural suitability. Socio-economic controls are: levels of black population, total population, night-time satellite light value in logs.

TABLE 2. Ethnic distribution and conflict: cross-sectional estimation.

	Total Number of Non-state conflict Events in the District									
	OLS		NB		LPM		Log		Cumulative	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
PANEL A. YEAR 1991										
<i>P</i>	3.186 (3.683)	5.545 (4.173)	5.497 (5.167)	4.352 (4.826)	0.151 (0.404)	0.272 (0.424)	0.579 (2.253)	1.573 (2.385)	6.872 (8.565)	17.710* (10.090)
<i>F</i>	8.105 (4.930)	6.601 (6.415)	9.164 (7.097)	8.034 (10.029)	0.462 (0.508)	0.407 (0.528)	3.585 (2.744)	2.763 (2.880)	18.835* (10.620)	5.849 (12.448)
<i>G/N</i>	-2.930 (1.913)	-3.298 (2.508)	-3.494 (2.490)	-2.961 (4.409)	-0.157* (0.091)	-0.190** (0.091)	-1.102* (0.580)	-1.220** (0.574)	-6.861* (3.897)	-6.141 (4.256)
<i>R</i> ²	0.185	0.204	n.a.	n.a.	0.412	0.453	0.411	0.457	0.194	0.243
Partial <i>R</i> ²	0.023	0.039	n.a.	n.a.	0.019	0.038	0.022	0.051	0.017	0.070
PANEL B. YEAR 1996										
<i>P</i>	-1.905 (1.583)	-1.948 (1.651)	-21.474 (13.937)	-13.039 (12.822)	-0.466 (0.321)	-0.456 (0.332)	-2.652 (1.876)	-2.621 (1.946)	-1.176 (2.334)	-1.019 (2.830)
<i>F</i>	1.541 (1.190)	1.939 (1.419)	13.917 (12.943)	13.743 (13.049)	0.555 (0.357)	0.592 (0.394)	2.886 (1.844)	3.162 (2.083)	5.604** (2.548)	-0.032 (2.975)
<i>G/N</i>	0.113 (0.201)	0.020 (0.203)	2.229 (4.350)	0.040 (4.792)	-0.006 (0.063)	-0.020 (0.065)	0.018 (0.324)	-0.067 (0.334)	-1.357 (1.093)	0.541 (1.104)
<i>R</i> ²	0.030	0.052	n.a.	n.a.	0.053	0.066	0.049	0.063	0.201	0.244
Partial <i>R</i> ²	0.016	0.042	n.a.	n.a.	0.022	0.035	0.020	0.035	0.005	0.053
Controls	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Observations	294	291	294	291	294	291	294	291	294	291

Notes: Standard errors in parenthesis. The unit of observation is a Magisterial District in South Africa for which information is available in both 1991 and 1996. The dependent variable is the intensity or presence of Non-state conflict events in the district as recorded in the UCDDP-GED v1.5 dataset. *P*, *F* and *G/N* are the ethnolinguistic distribution measures. Panel A and B report cross-sectional coefficient estimates from 1991 and 1996 respectively, with the inclusion of province fixed effects. Columns (1) and (2) report Ordinary Least Squares (OLS) estimates having as dependent variable the total number of conflict events in the district. Columns (3) and (4) report Negative Binomial (NB) estimates. Columns (5) and (6) report OLS estimates from a Linear Probability Model (LPM) where the dependent variable is a dummy equal to one if at least one conflict event was recorded in the district. Columns (7) and (8) report OLS estimates from a specification where the dependent variable is the logarithm of the number of Non-state conflict events (augmented by 0.01). Columns (9) and (10) report OLS estimates having as dependent variable the cumulative number of conflict in 1989-1991 (Panel A) and 1994-1996 (Panel B). Heteroskedastic residuals are assumed in all specifications except NB. The set of controls includes the geographical controls (ruggedness, slope index, accessibility, agricultural suitability) and the socio-economic controls (black population size, total population size, night-time satellite light value in logs) (Sources: UCDDP-GED v1.5, Statistics South Africa 1991, 1998; NOAA 2012).

TABLE 3. Ethnic distribution and conflict: first-difference/fixed-effects estimation.

	Total Number of Non-state conflict Events in the District										
	OLS		NB				LPM		Log		Cumulative
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	
ΔP	26.647** (12.789)	26.993** (12.765)	26.907* (15.159)	30.445* (18.132)	1.495*** (0.512)	1.523*** (0.472)	10.150*** (3.549)	10.362*** (3.348)	41.404** (16.820)	41.492** (17.082)	
ΔF	4.299 (13.311)	3.099 (12.861)	-9.523 (13.220)	-16.761 (19.577)	0.158 (0.482)	-0.095 (0.489)	0.813 (3.381)	-0.464 (3.353)	5.110 (17.403)	5.254 (16.325)	
$\Delta G/N$	-6.323 (4.160)	-6.268 (4.026)	-3.976 (7.275)	-1.751 (9.006)	-0.318** (0.129)	-0.291*** (0.109)	-2.130** (0.943)	-2.021** (0.842)	-9.483* (5.640)	-9.516* (5.435)	
R^2	0.073	0.087	n.a.	n.a.	0.045	0.083	0.052	0.080	0.030	0.041	
Controls	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	
Observations	294	294	94	94	294	294	294	294	294	294	

Notes: Standard errors in parenthesis. The unit of observation is a Magisterial District in South Africa for which information is available in both 1991 and 1996. The dependent variable is the change in the intensity or presence of Non-state conflict events in the district as recorded in the UCDDP-GED v1.5 dataset. *P*, *F* and *G/N* are the ethnolinguistic distribution measures. The table reports first-difference or fixed-effects coefficient estimates from 1991-1996. Columns (1) and (2) report Ordinary Least Squares (OLS) estimates having as dependent variable the change in the total number of conflict events in the district. Columns (3) and (4) report Negative Binomial (NB) estimates having as dependent variable the total number of conflict events in the district, with the inclusion of district fixed effects (those districts with zero conflict events in both 1991 and 1996 are in this case automatically dropped from the estimation sample, reducing the number of observations to 94). Columns (5) and (6) report OLS estimates from a Linear Probability Model (LPM) where the dependent variable is the change in a dummy equal to one if at least one conflict event was recorded in the district. Columns (7) and (8) report OLS estimates from a specification where the dependent variable is the change in the logarithm of the number of Non-state conflict events (augmented by 0.01). Columns (9) and (10) report OLS estimates having as dependent variable the change in the cumulative number of conflict between 1989-1991 and 1994-1996. Heteroskedastic residuals are assumed in all specifications except NB. The set of controls includes changes in socio-economic controls (black population size, total population size, night-time satellite light value in logs) or their levels (Columns (3) and (4)) (Sources: UCDDP-GED v1.5, Statistics South Africa 1991, 1998, NOAA 2012).

consistent with basic correlation patterns. According to the results in columns (1) and (2), an increase in one cross-district standard deviation in the within-black ethnolinguistic polarization measure is associated with more than five additional conflict events per district: more than four times the 1991 cross-district national average. The estimated coefficient is significant at the 5% level.²⁰ Negative binomial estimates in columns (3) and (4) are also significant, although only at the 10% level.²¹ Columns (5) and (6) indicate that one cross-district standard deviation increase in the within-black ethnolinguistic polarization is associated with an increase of 30 percentage points in the probability of experiencing at least one conflict per year. Estimates are still significant at the 1% level when the logarithm of conflict is used as outcome in columns (7) and (8), and are significant at the 5% level when we replace cumulative conflict in a three-year period as dependent variable in columns (9) and (10).

4.3. Discussion and Robustness

Evidence shows that ethnic distribution measures do not correlate or correlate weakly with conflict in both 1991 and 1996. Nevertheless, the correlation is strongly significant when looking at changes over the same period. We interpret these results as indicating the presence of unobservable characteristics at the district level that correlate with ethnic distribution and conflict incidence in opposite ways, confounding the cross-sectional estimates. Table A.4 in Appendix A shows how agricultural suitability can act as such confounder. Another example is ancestral exposure to contagious diseases, which may be higher in some districts than in others. Cervellati, Chiovelli, and Esposito (2017) show that ancestral exposure to malaria increases the benefits of interacting in small and geographically separated groups, thus increasing ethnic diversity. At the same time, disease exposure can increase the cost of mobilizing individuals for political violence due to poor health of potential fighters. In this scenario, variation in the disease environment generates a spurious negative correlation between diversity and conflict, biasing downwards the estimated correlation in levels between the two. The first-difference specification allows to net out all time-invariant potential confounders, agricultural suitability and ancestral exposure to contagious diseases being two of them.

The results in Table 3 also show that the polarization index P is the relevant ethnic diversity measure in explaining conflict incidence during the democratic transition. This is in line with our argument, as we conceptualize the conflict among black-dominated groups as a fight to achieve political power. Political power is an

20. Table A.5 in Appendix A shows that the significance of this estimate is not sensitive to the choice of the time interval over which we track the evolution of conflict.

21. Those districts in which no conflict events are recorded in both 1991 and 1996 are automatically dropped from the estimation sample. This reduces the number of observations to 94.

TABLE 4. Polarization and conflict.

	Change in Total No. of Non-state conflict Events 1991-1996							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ΔRQ	5.924*** (1.593)	5.562*** (1.628)	5.178*** (1.659)	5.178* (3.044)	4.690*** (1.669)	5.174*** (1.663)	4.758*** (1.66)	4.107*** (1.409)
Δ Between- RQ					7.023** (3.515)			
$\frac{\text{Immigrants}}{\text{Population 1991}}$						-0.052 (0.506)		
$\overline{\Delta RQ}_j$							7.692** (3.332)	
Non-state Conf 89-90								-0.273*** (0.026)
R^2	0.045	0.058	0.096	0.096	0.108	0.096	0.112	0.354
Main Controls	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Additional Controls	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Observations	294	294	294	294	294	294	294	294

Notes: Standard errors in parenthesis. The table reports first-difference coefficient estimates. The unit of observation is a Magisterial District in South Africa for which information is available in both 1991 and 1996. The dependent variable is the change in total number of Non-state conflict events recorded in the district in the UCDP-GED v1.5 dataset between 1991 and 1996. ΔRQ is the change in within-black polarization in the same period. Δ Between- RQ is the change in the polarization between the four races (black, white, asian and colored) in the district. Immigrants/Population 1991 is the number of blacks immigrating in the district between 1991 and 1996 as a share of the black population in 1991. $\overline{\Delta RQ}_j$ is the average change in within-black ethnolinguistic polarization in neighboring districts. *Non-state Conf 89-90* is the total number of conflict events coded in the MD in 1989-1990. (1) to (3) and (5) to (8) are first-difference estimates assuming homoskedastic difference residuals; in (4) 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). The set of Main Controls includes the change in: black population, total population, night-time satellite light value in logs. The set of Additional Controls includes the change in: rural population, number of individuals with no education, number of individuals unemployed, number of individuals out of the labor force, number of citizens of South Africa in logs (Sources: UCDP-GED v1.5, Statistics South Africa 1991, 1998; NOAA 2012).

intrinsically public prize. Results are thus consistent with the theoretical framework in Esteban and Ray (2011b) showing that the degree of publicness of the disputed prize determines the salience of polarization P relative to fractionalization F in explaining conflict incidence.

For this reason, in what follows, we restrict our attention to the polarization index. We implement the same first-difference regression specification specified above, but focusing on the binary polarization measure RQ . Table 4 reports coefficient estimates from this specification. The results in columns (1) and (2) mirror the same columns of Table 3, showing the salience of the polarization index when computed using

its binary version RQ . This is not surprising since the variability in RQ overlaps largely to the variability in P .²² An increase in one cross-district standard deviation in the within-black ethnolinguistic polarization measure is now associated with almost two more conflicts per district, still more than the 1991 nationwide average. The corresponding coefficient is significant at the 1% level. In column (3), we add a vector of additional time-varying regressors as controls.²³ The coefficient of interests is somewhat smaller in magnitude, but still significant at the 1% level. In column (4), we follow Conley (1999) and allow for a non-zero spatial correlation of first-difference residuals 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.

One possible concern with these estimates is that a high correlation may exist between the within-black ethnolinguistic polarization and the overall polarization between main population groups in South Africa, meaning - as defined in the 1991 Census - whites, blacks, asians, and colored. We compute the corresponding value of between-polarization, and add it as additional regressor in column (5). The coefficient of within-black polarization decreases in magnitude by around 10%, but remains significant at the 1% level. Together with the positive and significant coefficient of the between-polarization measure, this indicates that the two measures are indeed positively correlated, but the within-black polarization carries enough independent variation to predict the evolution of conflict between black organized parties.²⁴

Another concern has to do with the main source of variation over time in the within-black polarization measure: internal migration. Indeed, the increase (or decrease) in polarization may just be capturing the inflow of migrants in the district. If that was the case, our estimates would be informative of the salience of divisions between new immigrants and former residents rather than separation along ethnic lines: a hypothesis in line with the *sons of the soil* argument (Fearon 2004; Fearon and Laitin 2011). To address this concern, we add in column (6) an additional regressor equal to the size of the immigrant population that moved into the district between 1991 and 1996 (as retrievable from the 1996 Census) relative to the district population

22. As Section 3 explained, we identify the black ethnolinguistic groups in South Africa using the information contained in the *Ethnologue* linguistic database (Lewis 2009). The groups we identify share up to 9/11 of their path along the ethnolinguistic tree, so that there is little heterogeneity in terms of pairwise distances among them.

23. These are the size of the rural population, the number of individuals with no education, the number of individuals unemployed, the number of individuals not in the labor force, and the number of citizens of South Africa in logs.

24. In Table C.2 in the Online Appendix, we include a measure of between-polarization that only captures the polarization between whites and blacks in the population. Results are quantitatively and qualitatively similar. Table A.6 in Appendix A also shows how the estimated coefficient of the within-black ethnolinguistic polarization measure remains highly significant when controlling separately for the changes of each within-black group's share. This indicates that none of the latter is individually responsible for the relationship we find in the baseline specification.

size in 1991. When compared to columns (1) to (4), the coefficient of within-black polarization experiences no change in magnitude, and is still significant at the 1% level.²⁵

In column (7), we directly take into account and control for spatial spillovers by including as regressor the average change in within-black ethnolinguistic polarization in neighboring districts, $\overline{\Delta RQ}_{j96-91}$. The estimated coefficient of within-black polarization remains significant at the 1% level.²⁶ Finally, in the last column of Table 4, we augment the first-difference specification by including as regressor the number of non-state conflict events in 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. The coefficient of the latter is negative and highly significant. This means that conflict prevalence decreased more between 1991 and 1996 in those districts with a higher number of conflicts in 1989-1990. Nonetheless, the estimated coefficient of within-black polarization is still significant at the 1% level.²⁷

Taken altogether, the evidence in Table 4 shows that changes in within-black ethnolinguistic polarization correlate strongly with the evolution of conflict across districts through the democratization period. In order to further qualify these results, it is important to understand the extent to which the results may be driven by omitted variable bias. We follow Oster (2017) and calculate the degree of selection on unobservables relative to observables which would be necessary to explain away the result, meaning to produce a coefficient of our variable of interest equal to zero. These calculations suggest that the unobservables would need to be as important as the observables to explain away the result.²⁸

25. In Tables C.5 and C.6 in the Online Appendix, we consider two additional variables that capture the salience of divisions between the new immigrant and the resident population in the district. First, we divide the size of the immigrant population by the district population size in 1996 rather than 1991. Second, we compute the polarization index between the immigrant population and former residents in 1996. Including these variables as additional regressors does not change the estimated coefficient of within-black polarization.

26. Table C.7 in the Online Appendix shows that the coefficient of the within-black ethnolinguistic polarization is 14% lower and significant at the 1% level when we include the average change in conflict (the dependent variable) in neighboring districts as regressor, and 21% lower and significant at the 5% level when we instrument the latter with the average change in within-black ethnolinguistic polarization in neighboring districts.

27. The Online Appendix provides additional tables showing that results do not change when we use the estimated number of deaths in non-state conflicts per district as the measure of conflict incidence and when we control for a measure of intensity of governmental repression.

28. In Oster (2017), this is captured by the value of the parameter δ , which in our case equals 1.04. The approach in Oster (2017) differs from the one in Altonji, Elder, and Taber (2005) in that it also factors movements in the R^2 in the calculation. It thus requires to assume a value for the R^2 attainable in a hypothetical regression that includes both observed and unobserved explanatory variables. We refer to Esteban, Mayoral, and Ray (2012) and set this value equal to 0.4. This means that we regard 0.4 as the maximum value of R^2 attainable by a hypothetical model where controls include both observable and

4.3.1. Agricultural Suitability and Ethnic Inequality. Section 2 explained how, during democratization, magisterial district institutions were assigned competences over land titling and redistribution. Black organized parties took widely divergent positions on this issue. If land issues matter for confrontations among black parties, and these mobilize individuals for violence along ethnic lines, the relationship between ethnic diversity and conflict will be stronger in districts with higher suitability for agriculture.

TABLE 5. Polarization and agricultural suitability.

	Total No. of Non-state conflict Events 1991-1996			
	(1)	(2)	(3)	(4)
RQ	-6.129** (2.847)	-5.838** (2.863)	-4.919* (2.824)	-4.919 (3.607)
$RQ \times \text{Suitability}$	29.944*** (5.953)	28.853*** (6.178)	27.558*** (6.107)	27.558* (16.181)
R^2	0.554	0.563	0.586	0.586
Main Controls	No	Yes	Yes	Yes
Additional Controls	No	No	Yes	Yes
Observations	586	586	586	586

Notes: Standard errors in parenthesis. The table reports fixed effects coefficient estimates. The unit of observation is a Magisterial District in South Africa for which information is available in both 1991 and 1996. The dependent variable is the total number of Non-state conflict events recorded in the district in the UC DP-GED v1.5 dataset in 1991 and 1996. RQ is the within-black polarization in the same year. *Suitability* is the district-level measure of agricultural suitability derived from Ramankutty, Foley, Norman, and McSweeney (2002). The set of Main Controls includes: black population, total population, night-time satellite light value in logs. The set of Additional Controls includes: rural population, number of individuals with no education, number of individuals unemployed, number of individuals out of the labor force, number of citizens of South Africa in logs. (1) to (3) are first-difference estimates assuming homoskedastic residuals; those in (4) are first-difference estimates assuming heteroskedastic residuals (Sources: UC DP-GED v1.5, Statistics South Africa 1991, 1998; NOAA 2012).

We test this hypothesis by implementing a regression specification with district fixed effects that matches the first-difference one in equation 2, but including as additional regressor the interaction between within-black ethnolinguistic polarization and agricultural suitability (Ramankutty, Foley, Norman, and McSweeney 2002).²⁹ Table 5 shows the corresponding estimation results. The coefficient of the interaction variable is positive and significant at least at the 10% level in all specifications. This indicates that those districts where land is more suitable for agriculture drive the positive relationship between within-black ethnolinguistic polarization and conflict. Evidence supports the hypothesis that incentives to mobilize individuals for political

unobservable variables. This choice is very conservative, as the R^2 for our model specification in column (3) of Table 4 is 0.096.

29. Agricultural suitability per se is time-invariant and therefore its variability absorbed by district fixed effects.

violence were higher in districts with higher agricultural suitability, and that such mobilization occurred along ethnic lines.

TABLE 6. Ethnic inequality and polarization.

	Change in Total No. of Non-state conflict Events 1991-1996					
	(1)	(2)	(3)	(4)	(5)	(6)
ΔEI	3.358** (1.363)	3.061** (1.376)	3.679*** (1.382)		3.710*** (1.345)	3.483*** (1.333)
ΔRQ				5.178*** (1.659)	8.645*** (2.141)	5.223** (2.469)
$\Delta EI \times \Delta RQ$						-23.612*** (8.772)
R^2	0.021	0.042	0.095	0.096	0.146	0.168
Main Controls	No	Yes	Yes	Yes	Yes	Yes
Additional Controls	No	No	Yes	Yes	Yes	Yes
Observations	283	283	283	294	283	283

Notes: Standard errors in parenthesis. The table reports first-difference coefficient estimates. The unit of observation is a Magisterial District in South Africa for which information is available in both 1991 and 1996. The dependent variable is the change in total number of Non-state conflict events recorded in the district in the UCDP-GED v1.5 dataset between 1991 and 1996. ΔRQ and ΔEI are changes in within-black polarization and within-black ethnic inequality respectively in the same period. The set of Main Controls includes the change in: black population, total population, night-time satellite light value in logs. The set of Additional Controls includes the change in: rural population, number of individuals with no education, number of individuals unemployed, number of individuals out of the labor force, number of citizens of South Africa in logs (Sources: UCDP-GED v1.5, Statistics South Africa 1991, 1998; NOAA 2012).

The relative income of groups may also shape the incentives to mobilize and be mobilized for political violence. The relationship between ethnic inequality and conflict incidence is a priori ambiguous.³⁰ On the one hand, differences in income across groups increase the scope for redistribution, raising the incentives to achieve political power. On the other hand, violence is a less viable option for poor groups as they have limited access to resources to fund conflict. Also, ethnic inequality and polarization can interact positively or negatively in the generation of conflict. While income differences can make ethnic traits even more salient in polarized societies, a framing of conflict as rich vs. poor can be more powerful when groups are more unequal in size, and a large poor majority faces a small rich elite.

To shed light on these issues, we first implement the same first-difference specification presented above, but having ethnic inequality as the main regressor. Column (1) of Table 6 shows that when the change in ethnic inequality is the only included regressor, its coefficient is positive and significant at the 5% level. An increase in the ethnic

30. We thank one anonymous referee for pushing us to think in this direction.

inequality index of one cross-district standard deviation is associated with almost one more conflict event per district. The same is true when we include time-varying economic controls in columns (2) and (3), with the coefficient of ethnic inequality becoming significant at the 1% level. We then move to explore the complementarity or substitutability between within-black ethnic inequality and polarization in generating conflict. Column (4) of Table 6 replicates column (3) in Table 4, where the main regressor is the change in within-black ethnolinguistic polarization. In column (5), we include both the ethnic inequality index and the binary polarization index. The coefficients of both the within-black ethnolinguistic polarization and inequality indexes are positive and highly significant. Finally, in column (6), we include the interaction between the two. The corresponding coefficient is negative and highly significant. These results suggest that income differences between ethnic groups matter positively for conflict incidence, and more so when group size differences are larger. For instance, this is the case when a large poor majority faces a small rich minority, and the scope for redistribution is higher.

4.4. Additional Robustness Checks

First-difference estimates could still reflect the presence of unobserved factors which affect the evolution of both ethnic distribution and conflict, generating a spurious correlation between the two. We address this concern 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, García-Jimeno, and Robinson (2012), we assume 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, it is therefore possible to net out those common unobserved sources of heterogeneity and time-varying omitted variables that may correlate with the evolution of both polarization and conflict incidence. We report the details of the implemented procedure and results in Table A.7 in Appendix A. We still find a significant relationship between the change in within-black ethnolinguistic polarization and the evolution of conflict. The corresponding point estimate is comparable to the ones in columns (3) and (4) of Table 4, and significant at the 1% level.

Exploiting differential changes in ethnic composition across neighboring districts can still be problematic. The relocation choices of internal migrants could possibly be endogenous to the evolution of conflict.³¹ We address this issue using instrumental variables. To establish physical separation of blacks and whites, the apartheid regime forcedly relocated blacks to designated homelands. Forced removals started in the

31. Migration across contiguous districts represents a high fraction of total district-level moves in the 1991-1996. Table C.10 in the Online Appendix shows that, on average, 35% of Xhosa and 28% of Zulu movers are estimated to relocate in neighboring districts.

1960s, but continued well into the 1980s. Figure A.6 in Appendix A plots the distribution of within-black ethnolinguistic polarization across districts in the years 1970, 1991, and 1996. Comparing the 1970 and the 1991 distribution, we can see how the average level of within-black ethnolinguistic polarization diminishes over the period. In particular, the mode of the distribution shifts from around 0.7 to less than 0.2. We interpret this as evidence that, by implementing a model of physical separation and segregation, the apartheid government successfully enforced an ethnic distribution characterized by a lower number of polarized districts and a higher number of homogenous ones. Between 1991 and 1996, the distribution appears to move in the opposite direction, back to its shape in 1970. This indicates that the internal migration flows that followed the repeal of apartheid legislation partially overturned the spatial ethnic-based architecture of the regime.

We can exploit this finding to construct an instrumental variable. We use the difference between within-black ethnolinguistic polarization in 1970 and its value in 1991 to predict the difference between 1996 and 1991. That is, we use $\Delta \widetilde{RQ}_i = RQ_{i,1970} - RQ_{i,1991}$ as instrument for $\Delta RQ_i = RQ_{i,1996} - RQ_{i,1991}$. Identification requires that the instrument induces meaningful variation in the observed change in polarization, and is independent from the change in conflict incidence in 1991-1996. The first condition is satisfied as long as historical return migration at least partially informs the patterns of internal migration in the period. The extent to which the second condition is satisfied depends instead on the extent to which the ethnic distribution in 1970 is independent from the struggle between black organized parties in the 1990s.

Table A.8 in Appendix A shows OLS estimates in Panel A, and both first and second stage IV estimates in Panel B. The need of retrieving ethnolinguistic information consistently from 1970 to 1996 forces us to reduce the sample size to 270 magisterial districts. The OLS estimates in Panel A are comparable with those reported in Table 4, and significant at the 1% level across all specifications. First stage estimates in Panel B show that our instrument correlates with the actual change in within-black ethnolinguistic polarization. However, the *F-statistics* for the significance test of the instrument in the first stage regression is only close to the threshold value of 10. The strength of our instrument is therefore borderline, so we interpret these estimates with some caution. The specifications in columns (1) to (3) mirror the ones in the same columns of Table 4, where we progressively include our sets of main and additional time-varying controls. The second stage estimate of our coefficient of interest is significant at the 10% level, and almost three times bigger than its corresponding OLS estimate. This suggests that endogenous migration decisions may be responsible for a downward bias in previous estimates. However, results from a Hausman test do not allow us to reject the hypothesis of both the OLS and IV estimators being

consistent (Hausman 1978).³² One possible concern with our instrumental variable strategy is that historical return migration itself can be among the determinants of the struggle among black organized parties at the local level, violating the exclusion restriction. To address this issue - and mirroring column (6) of Table 4 - we include as additional regressor the size of the immigrant population which moved in the district between 1991 and 1996 relative to the district population size in 1991. Column (4) of Table A.8 in Appendix A shows that the coefficient of within-black polarization experiences little change in magnitude and remains significant at the 10% level.³³ Overall, the results in this Section confirm the existence of a positive relationship between the change in within-black ethnolinguistic polarization and the evolution of conflict during democratization.

4.5. *Additional Supporting Evidence*

Our main argument is that during democratization ethnicity stands out as a salient technology to separate individuals into well-identified groups and mobilize them for political violence. Embedded in this reasoning is the claim that the relationship between ethnolinguistic polarization and conflict incidence occurs through political mobilization. If data on the political affiliation of each individual were available for both 1991 and 1996, we could study how within-black ethnolinguistic diversity maps into political diversity and thus conflict incidence in both periods in a systematic way. Unfortunately, no data on political affiliation are available at the magisterial district level for those years. We can nonetheless provide at least suggestive evidence of this link by looking at the results from the first democratic elections in 1994. We can use disaggregated information on election results at the district level to compute an index of within-black political polarization, and investigate whether a meaningful

32. One possible source of difference between OLS and IV estimates is measurement error. Focusing on the results in column (1) of Table A.8, a simple calculation based on the formula for the attenuation bias indicates that, if random measurement error was solely responsible for the difference we observe in coefficient estimates, it would account for 65% of the variability in the observed change of within-black polarization. A different explanation is in the possible endogeneity of internal migration moves to the evolution of conflict. If internal migrants avoided relocating in districts where they expected conflict to increase, the OLS estimate would be downward bias to the extent to which these relocation choices made the increase in ethnolinguistic polarization lower than what it would have been otherwise. Finally, the difference between the OLS and IV coefficient estimates can also reflect the local interpretation of the instrumental variable estimate of the parameter we wish to identify (Angrist, Imbens, and Rubin 1996).

33. Table C.11 in the Online Appendix shows that results do not change when we replace as controls the other two proxies for the salience of internal migration that we use in Tables C.5 and C.6. Another concern with the proposed strategy is that the value of within-black ethnolinguistic polarization in 1991 appears negatively both in our instrument and the endogenous regressor, possibly accounting for all the predictive power of our instrument. To address this concern, we implement a falsification test where we use polarization in 1991 as instrument for the change between 1991 and 1996. The corresponding estimates in Table A.9 in Appendix A are very different from the ones in Table A.8, ruling out the possibility that the variability in 1991 polarization is by itself driving our results.

relationship exists between the latter and within-black ethnolinguistic polarization in 1991.³⁴

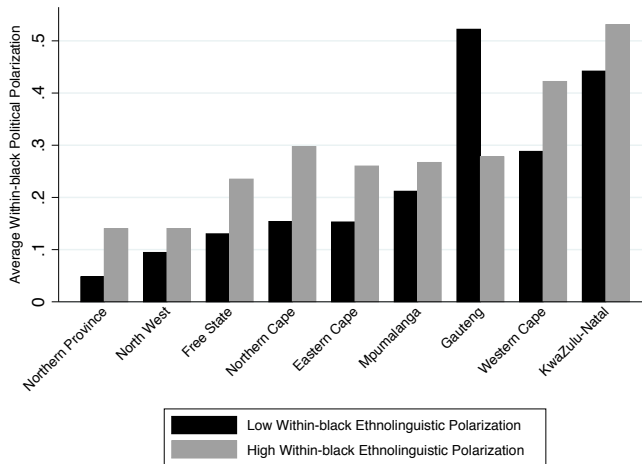


FIGURE 3. Ethnolinguistic and political polarization. The figure plots the average level of political polarization among black-dominated political parties in the 1994 general elections. For each province, the average level of political polarization is shown separately for those districts in the top and bottom quartile of the within-black ethnolinguistic polarization distribution. Both polarization measures are computed using the binary version of the polarization measure as explained in Section 3. The political parties considered in the computation of the political polarization index are: ANC, IFP, PAC, ACDP (Sources: Statistics South Africa 1991; Alvarez-Rivera 2004).

Figure 3 plots the average level of within-black political polarization across districts separately for each province, and further split according to the level of within-black ethnolinguistic polarization. In particular, the figure shows the average level of political polarization in the top and the bottom quartile of the ethnolinguistic polarization distribution within each province. There is meaningful variation in the level of political polarization across provinces. Not surprisingly, the province with the highest level of within-black political polarization is the KwaZulu-Natal: the IFP stronghold. Despite this, a clear pattern emerges within provinces in the relationship between political polarization and ethnolinguistic diversity. With the exception of the Gauteng province, those districts with the highest within-black ethnolinguistic polarization in 1991 show higher levels of within-black political polarization in 1994. This relationship is significant. The estimated coefficient of the RQ index in 1991 from a regression of within-black political polarization in 1994 on the former and province fixed effects is equal to 0.10 and significant at the 6% level. The result for the Gauteng province reflects a rural/urban divide: Johannesburg and Pretoria,

34. As Section 3 explains, in the computation of the within-black political polarization index at the electoral district level we consider the ANC, IFP, PAC and ACDP parties, and use the binary version of the polarization index RQ .

two of the biggest urban clusters in the country, are both located in Gauteng, and ethnic lines are less salient in the political struggle within the young urban black population.³⁵ We interpret this evidence as suggestive of the existence of a mapping between ethnolinguistic and political polarization.

TABLE 7. Ethnic distribution and conflict in democratic South Africa.

	Change in Total No. of Conflict Events 2001-2011			
	(1)	(2)	(3)	(4)
ΔRQ	-0.359 (0.341)	-0.300 (0.395)	-0.324 (0.406)	-0.324 (0.271)
R^2	0.005	0.020	0.043	0.043
Main Controls	No	Yes	Yes	Yes
Additional Controls	No	No	Yes	Yes
Observations	231	231	231	231

Notes: Standard errors in parenthesis. The table reports first-difference coefficient estimates. The unit of observation is a Local Municipality in South Africa for which information is available in both 2001 and 2011. The dependent variable is the change in total number of conflict events coded in the ACLED dataset in the same period. ΔRQ is the corresponding change in the within-black ethnolinguistic polarization measure. The set of Main Controls includes the change in: black population, total population, night-time satellite light value in logs. The set of Additional Controls includes the change in: rural population, number of individuals with no education, number of individuals unemployed, number of individuals out of the labor force, number of citizens of South Africa in logs. (1), (2), (3) are first-difference estimates assuming homoskedastic difference residuals; those in (4) are first-difference estimates assuming heteroskedastic difference residuals (Sources: Raleigh, Linke, Hegre, and Karlsen 2010; Statistics South Africa 2004, 2014; NOAA 2012).

The narratives presented in Section 2 also make clear that incentives to achieve political power were highest during democratization, also at the local level. The period of democratization was crucial to establish the rules of the game as politicians designed and implemented policies that were expected to shape the future of the country. We therefore expect the incentives to mobilize individuals for political violence to peak during this period, and to decrease thereafter. We test this hypothesis using Census and conflict data from the years 2001 and 2011.³⁶ Table 7 reports the corresponding coefficient estimates. Consistent with Table 4, we focus on the binary version of the polarization index.³⁷ Evidence shows no systematic relationship between within-black ethnolinguistic polarization and conflict incidence in 2001 through 2011. This indicates that the salience of ethnicity in the violent political

35. The qualitative evidence in Section 2 also highlights the salience of the confrontations between the multi-ethnic ANC and the Zulu-based IFP. When we regress the same measure of within-black political polarization in 1994 over an RQ index that only considers Zulus and non-Zulus, the coefficient of the latter is equal to -0.04 and insignificant. This suggests that the measure that is relevant to our results is the one that consider all seven black groups.

36. As Section 3 explained, in this part of the analysis we adopt the local municipality as unit of analysis and use ACLED data to derive our measure of local conflict incidence.

37. Table C.12 in the Online Appendix shows the results from a specification where the three within-black distribution indexes are included, in accordance with Tables 2 and 3.

struggle diminished after democratization, and that current political tensions in South Africa do not manifest themselves through ethnic markers. This corroborates our view that the incentives to mobilize individuals for political violence were stronger during the process of democratization than in the period thereafter, when the institutions and main policies of democratic South Africa are already established.

5. Conclusions

This paper exploits the features of history of contemporary South Africa to study the relationship between ethnic diversity and political violence during democratization. The implementation and repeal of apartheid segregation laws induced meaningful changes in the ethnic composition of districts. At the same time, black-dominated political groups struggled violently among themselves to achieve power and dominate the new institutional scenario. Our analysis shows that the evolution of conflict correlates significantly with the changes in ethnic composition that occurred in the democratization period. Results are robust to the inclusion of an array of controls, and are not confounded by unobserved factors at a fine geographical level. An instrumental variable strategy that exploits and predicts historical return migration also confirms the presence of a strong positive relationship between within-black ethnolinguistic polarization and the incidence of armed confrontations between black-dominated groups.

We interpret the evidence in this paper as showing that, during democratization, ethnic markers stand out as a readily available technology to separate individuals into well-identified groups, and mobilize them for political violence. Does this affect the quality of democracy and its consolidation? In their study of Kenya, Burgess, Jedwab, Miguel, Morjaria, and Padró i Miquel (2015) show that ethnic favoritism disappears during periods of democracy. Is this still the case when the transition is violent and the battle for power is fought along ethnic lines? We plan to investigate these questions in our future research.

Appendix: Additional Tables and Figures

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 can be mapped into Magisterial Districts, together with the estimated total number of deaths.

TABLE A.2. Summary statistics.

Variable	Mean	St. Dev.	Min	Max	N
PANEL A: YEAR 1991					
Non-state conflict Events	1.271	5.192	0	41	303
<i>P</i>	0.226	0.209	0	0.623	296
<i>F</i>	0.240	0.244	0	0.800	296
<i>G/N</i>	0.588	0.617	0	3.856	296
<i>RQ</i>	0.368	0.336	0	0.989	296
<i>EI</i>	0.501	0.194	0	0.819	285
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	2.249	1.448	0.237	6.538	354
Slope Index	73.009	22.623	13	99	351
Agricultural Suitability	0.451	0.256	0.008	0.927	353
PANEL B: YEAR 1996					
Non-state conflict Events	0.105	0.692	0	6	354
<i>P</i>	0.210	0.207	0	0.628	350
<i>F</i>	0.218	0.236	0	0.797	350
<i>G/N</i>	0.527	0.571	0	2.337	350
<i>RQ</i>	0.34	0.332	0	0.997	350
<i>EI</i>	0.358	0.203	0	0.787	345
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					
Non-state conflict Events	-1.044	4.9	-41	6	298
<i>P</i>	0.003	0.107	-0.560	0.628	294
<i>F</i>	-0.004	0.111	-0.666	0.539	294
<i>G/N</i>	-0.010	0.411	-3.856	2.337	294
<i>RQ</i>	0.002	0.177	-0.889	0.997	294
<i>EI</i>	-0.155	0.217	-0.775	0.589	283
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 and Appendix B (Sources: UCDDP-GED v1.5, Statistics South Africa 1991, 1998; Nunn and Puga 2012; IIASA/FAO 2012; NOAA 2012; Ramankutty, Foley, Norman, and McSweeney 2002).

TABLE A.2. Summary statistics (continued).

Variable	Mean	St. Dev.	Min	Max	N
PANEL D: YEAR 1970					
<i>RQ</i>	0.430	0.318	0	0.989	270
PANEL E: YEAR 2001					
ACLEDD Conflict Events	0.263	1.502	0	16	232
<i>P</i>	0.210	0.191	0	0.617	231
<i>F</i>	0.214	0.217	0	0.785	231
<i>G/N</i>	0.490	0.509	0	2.600	231
<i>RQ</i>	0.343	0.307	0	0.980	231
Blacks	151.378	277.870	0.145	2362.160	232
Population	191.242	408.135	6.152	3221.259	232
Rural Population	83.446	99.124	0.364	552.296	232
No Education	46.241	70.787	1.625	566.975	232
Unemployed	35.750	85.270	0.862	725.955	232
Night-time Lights	3.068	6.367	0.009	52.782	232
PANEL F: YEAR 2011					
ACLEDD Conflict Events	0.368	1.763	0	21	234
<i>P</i>	0.248	0.180	0.021	0.612322	234
<i>F</i>	0.260	0.222	0.019	0.7872139	234
<i>G/N</i>	0.628	0.624	0.041	3.405	234
<i>RQ</i>	0.403	0.290	0.037	0.979	234
Blacks	175.218	370.333	0.198 3388.92	234	234
Population	221.242	521.223	7.018	4434.922	234
Rural Population	82.141	104.292	0.123	528.938	234
No Education	7.031	13.049	0.292	99.981	234
Unemployed	37.604	92.734	1.117	776.606	234
Night-time Lights	3.626	6.790	0.011	55.668	234
PANEL G: DIFFERENCE 2001-2011					
ACLEDD Conflict Events	0.108	0.774	-4	5	232
<i>P</i>	0.036	0.093	-0.349	0.527	231
<i>F</i>	0.043	0.116	-0.280	0.714	231
<i>G/N</i>	0.125	0.435	-0.548	3.405	231
<i>RQ</i>	0.057	0.150	-0.551	0.842	231
Blacks	25.010	105.964	-183.029	1026.761	232
Population	31.322	128.022	-182.0745	1213.663	232
Rural Population	-0.795	28.910	-187.639	176.196	232
No Education	-39.168	58.155	-477.167	5.976	232
Unemployed	2.094	15.028	-26.916	154.716	232
Night-time Lights	0.001	0.001	-0.002	0.005	232

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 and Appendix B (Sources: Raleigh, Linke, Hegre, and Karlsen 2010; Statistics South Africa 2011, 2004, 2014; NOAA 2012; Ramankutty, Foley, Norman, and McSweeney 2002).

TABLE A.3. 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
Black	21,642,766	31,141,365	9,498,599	1,054,760
White	5,069,377	4,426,117	-643,260	-666,198
Asian	985,363	1,045,168	59,805	52,964
Colored	3,277,020	3,593,966	316,946	54,280

Notes: The table reports the ethnic group population sizes in South Africa in 1991 and 1996, together with overall difference and the difference computed over those districts for which information is available in both the 1991 and 1996 Census (Sources: Statistics South Africa 1991, 1998).

TABLE A.4. Correlations of main variables and controls.

PANEL A: YEAR 1991							
	<i>P</i>	<i>F</i>	<i>EI</i>	Conflict	Blacks	Population	Lights
<i>F</i>	0.9529 (0.0000)						
<i>EI</i>	-0.2810 (0.0000)	-0.3700 (0.0000)					
Conflict	-0.0015 (0.9799)	0.0336 (0.5655)	0.0296 (0.6191)				
Blacks	0.1460 (0.0122)	0.1983 (0.0006)	-0.1501 (0.0113)	0.1608 (0.0057)			
Population	0.0960 (0.1004)	0.1796 (0.0020)	-0.1662 (0.0050)	0.1723 (0.0030)	0.7929 (0.0000)		
Night-time Lights	0.2104 (0.0003)	0.3155 (0.0000)	-0.1945 (0.0010)	0.1471 (0.0115)	0.2854 (0.0000)	0.4739 (0.0000)	
Suitability	-0.2344 (0.0001)	-0.1846 (0.0015)	0.0833 (0.1623)	0.2438 (0.0000)	0.3624 (0.0000)	0.3116 (0.0000)	0.1199 (0.0402)
PANEL B: YEAR 1996							
	<i>P</i>	<i>F</i>	<i>EI</i>	Conflict	Blacks	Population	Lights
<i>F</i>	0.9396 (0.0000)						
<i>EI</i>	-0.2719 (0.0000)	-0.3612 (0.0000)					
Conflict	-0.0465 (0.4273)	0.0052 (0.9287)	-0.0468 (0.4279)				
Blacks	0.0478 (0.4147)	0.1517 (0.0092)	0.1233 (0.0361)	0.0601 (0.3040)			
Population	0.0449 (0.4432)	0.1732 (0.0029)	-0.0353 (0.5497)	0.0797 (0.1730)	0.7793 (0.0000)		
Night-time Lights	0.1643 (0.0047)	0.2819 (0.0000)	-0.1953 (0.0008)	0.0084 (0.8856)	0.3977 (0.0000)	0.5448 (0.0000)	
Suitability	-0.2857 (0.0000)	-0.2192 (0.0002)	0.1463 (0.0129)	0.0739 (0.2070)	0.3989 (0.0000)	0.3653 (0.0000)	0.1646 (0.0047)
PANEL C: DIFFERENCE 1991-1996							
	ΔP	ΔF	ΔEI	Δ Conflict	Δ Blacks	Δ Population	
ΔF	0.9441 (0.0000)						
ΔEI	-0.0376 (0.5290)	-0.0238 (0.6897)					
Δ Conflict	0.1860 (0.0014)	0.1512 (0.0094)	0.1454 (0.0144)				
Δ Blacks	-0.1723 (0.0030)	-0.1969 (0.0007)	-0.0729 (0.2212)	-0.0895 (0.1259)			
Δ Population	-0.1761 (0.0024)	-0.1940 (0.0008)	0.0055 (0.9270)	-0.0754 (0.1974)	0.6883 (0.0000)		
Δ Night-time Lights	-0.0355 (0.5447)	-0.0379 (0.5178)	-0.0199 (0.7383)	-0.1091 (0.0618)	0.0278 (0.6353)	0.0327 (0.5761)	

Notes: The Table shows the estimated correlation between the main variables of interest and main controls in levels in 1991 and 1996 and in differences 1991-1996, with p-values in parenthesis (Sources: UCDP-GED v1.5; Statistics South Africa 1991, 1998).

TABLE A.5. First-difference estimation over alternative time intervals.

	Change in Total No. of Non-state conflict Events							
	1990-1995	1992-1997	1993-1998	1990-1997	1991-1998	1992-1999	1992-1995	1993-1994
ΔP	18.316** (7.714)	26.309** (10.644)	22.109** (9.435)	18.545** (7.845)	26.781** (12.131)	26.836** (10.704)	26.080** (10.608)	11.938** (4.401)
ΔF	-3.112 (7.433)	-17.424 (13.022)	-11.872 (10.914)	-4.937 (7.665)	-1.893 (11.641)	-16.858 (13.068)	-15.599 (13.078)	-5.176 (5.077)
$\Delta G/N$	-0.003 (0.002)	-0.002 (0.003)	-0.003 (0.003)	-0.003 (0.003)	-0.005 (0.003)	-0.003 (0.003)	-0.003 (0.003)	-0.002 (0.001)
R^2	0.083	0.111	0.129	0.085	0.130	0.114	0.102	0.089
Main Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Additional Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	294	294	294	294	294	294	294	294

Notes: Standard errors in parenthesis. Heteroskedastic difference residuals are assumed in all columns. The table reports first-difference coefficient estimates. The unit of observation is a Magisterial District in South Africa for which information is available in both 1991 and 1996. The dependent variable is the change in total number of Non-state conflict events recorded in the district in the UCDDP-GED v1.5 dataset between the two years indicated in each column. ΔP , ΔF and $\Delta G/N$ are the changes in the corresponding ethnolinguistic distribution measures between 1991 and 1996. The set of Main Controls includes the change in: black population, total population, night-time satellite light value in logs. The set of Additional Controls includes the change in: rural population, number of individuals with no education, number of individuals unemployed, number of individuals out of the labor force, number of citizens of South Africa in logs (Sources: UCDDP-GED v1.5, Statistics South Africa 1991, 1998; NOAA 2012).

TABLE A.6. Within-black group shares as controls.

	Change in Total No. of Non-state conflict Events 1991-1996							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ΔRQ	6.499*** (1.685)	4.961*** (1.635)	5.403*** (1.778)	5.362*** (1.679)	5.174*** (1.662)	5.215*** (1.680)	5.234*** (1.663)	6.282*** (1.748)
Δ Sh. Xhosa _{WB}	10.865*** (3.402)							
Δ Sh. Zulu _{WB}		-13.569*** (4.314)						-43.828*** (7.074)
Δ Sh. Sotho _{WB}			-1.503 (4.223)					-10.817** (4.354)
Δ Sh. Swazi _{WB}				-7.346 (9.799)				-52.918*** (12.440)
Δ Sh. Tswana _{WB}					-1.648 (4.642)			-2.448 (4.477)
Δ Sh. Tsonga _{WB}						1.114 (7.385)		-35.492*** (10.264)
Δ Sh. Venda _{WB}							32.563 (46.678)	20.663 (48.925)
R^2	0.127	0.126	0.096	0.098	0.096	0.096	0.097	0.209
Main Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Additional Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	294	294	294	294	294	294	294	294

Notes: Standard errors in parenthesis. Heteroskedastic difference residuals are assumed in all columns. The table reports first-difference coefficient estimates. The unit of observation is a Magisterial District in South Africa for which information is available in both 1991 and 1996. The dependent variable is the change in total number of Non-state conflict events recorded in the district in the UCDP-GED v1.5 dataset between 1991 and 1996. ΔRQ is the change in within-black polarization in the same period. For each black ethnolinguistic group, Δ Sh. *Group*_{WB} is the change in the within-black share of that group. The set of Main Controls includes the change in: black population, total population, night-time satellite light value in logs. The set of Additional Controls includes the change in: rural population, number of individuals with no education, number of individuals unemployed, number of individuals out of the labor force, number of citizens of South Africa in logs (Sources: UCDP-GED v1.5, Statistics South Africa 1991, 1998; NOAA 2012).

TABLE A.7. First-difference with cluster-specific trends.

	Change in Total No. of Non-state conflict Events		
	(1)	(2)	(3)
ΔRQ	6.08*** (1.831)	6.36*** (1.982)	5.26*** (1.514)
Main Controls	No	Yes	Yes
Additional Controls	No	Yes	Yes
Polynomial of Controls	No	No	Yes
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. We obtain coefficient estimates and standard errors as follows. 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 non-treated neighboring districts. We obtain a sub-sample of 227 out of the initial 294 districts, with the final subsample containing 105 treated and 122 non-treated districts. For each treated district $g \in M$ and its non-treated neighbors $f \in N(g)$, we implement the following regression specification:

$$\begin{aligned} \Delta conf_{g96-91} &= \delta + \gamma_g + \beta \Delta RQ_{g96-91} + \Delta X'_{g96-91} \varphi + \Delta \varepsilon_{g96-91} & g \in M \\ \Delta conf_{f96-91} &= \delta + \gamma_g + \beta \Delta RQ_{f96-91} + \Delta X'_{f96-91} \varphi + \Delta \varepsilon_{f96-91} & f \in N(g) \end{aligned} \quad (A.1)$$

where $\Delta conf_{i96-91}$ is the change in the number of recorded non-state conflict events in district i between year 1991 and 1996, ΔRQ_{i96-91} is the change in the within-blacks polarization index (with $i = g, f$). Time trends are controlled for by δ , and 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 include a dummy for each cluster of districts, taking value one for all treated and non-treated observation in the cluster. X_{it} is the vector of time-variant district characteristics in year t (population, blacks, night-time light, rural, etc.). The difference residual $\Delta \varepsilon_{i96-91}$ captures those idiosyncratic unobserved changes and factors which affect the change in conflict incidence, net of cluster-specific trends. 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 set of Main Controls includes the change in: black population, total population, night-time satellite light value in logs. The set of Additional Controls includes the change in: rural population, number of individuals with no education, number of individuals unemployed, number of individuals out of the labor force, number of citizens of South Africa in logs. Specification (3) is augmented with a 3rd order Polynomial of all Controls (Sources: UCDP-GED v1.5; Statistics South Africa 1991, 1998; NOAA 2012).

TABLE A.8. Instrumental variable estimation.

PANEL A. ORDINARY LEAST SQUARES ESTIMATES				
	Change in Total No. of Non-state conflict Events			
	(1)	(2)	(3)	(4)
ΔRQ	6.830*** (1.805)	6.544*** (1.867)	6.226*** (1.891)	6.258*** (1.900)
$\frac{\text{Immigrants}}{\text{Population 1991}}$				-0.177 (0.771)
R^2	0.051	0.065	0.104	0.104
PANEL B. INSTRUMENTAL VARIABLE ESTIMATES				
	First Stage: Change in Polarization Measure $\Delta \widetilde{RQ}$			
	(1)	(2)	(3)	(4)
$\Delta \widetilde{RQ}$	0.168*** (0.054)	0.162*** (0.054)	0.165*** (0.054)	0.158*** (0.055)
<i>F-stat</i>	9.72	9.14	9.44	8.21
	Second Stage: Change in Total No. of Non-state conflict Events			
	(1)	(2)	(3)	(4)
$\Delta \widehat{RQ}$	19.882* (10.510)	20.320* (11.123)	19.166* (10.768)	21.043* (11.788)
$\frac{\text{Immigrants}}{\text{Population 1991}}$				-0.623 (0.909)
Main Controls	No	Yes	Yes	Yes
Additional Controls	No	No	Yes	Yes
Observations	270	270	270	270

Notes: Standard errors in parenthesis. The table reports Ordinary Least Squares (Panel A) and Instrumental Variable (Panel B) first-difference coefficient estimates. The unit of observation is a Magisterial District in South Africa for which information is available in 1970, 1991 and 1996. The dependent variable is the change in total number of Non-state conflict events recorded in the district in the UC DP-GED v1.5 dataset between 1991 and 1996. ΔRQ is the change in within-black polarization in the same period. $\Delta \widetilde{RQ}$ is the instrumental variable: the difference between the 1970 and the 1991 polarization, $RQ_{1970} - RQ_{1991}$. Immigrants/Population 1991 is the number of blacks immigrating in the district between 1991 and 1996 as a share of the black population in 1991. The set of Main Controls includes the change in: black population, total population, night-time satellite light value in logs. The set of Additional Controls includes the change in: rural population, number of individuals with no education, number of individuals unemployed, number of individuals out of the labor force, number of citizens of South Africa in logs (Sources: UC DP-GED v1.5, Statistics South Africa 1991, 1998; NOAA 2012).

TABLE A.9. Instrumental variable falsification instrument: within-black polarization in 1991.

	First Stage: Change in Polarization Measure ΔRQ			
	(1)	(2)	(3)	(4)
RQ_{1991}	-0.139*** (0.030)	-0.141*** (0.029)	-0.140*** (0.030)	-0.138*** (0.030)
<i>F-stat</i>	21.95	23.95	22.04	21.45
	Second Stage: Change in Total No. of Non-state conflict Events			
$\Delta \widehat{RQ}$	3.734 (6.573)	4.046 (6.445)	5.263 (6.642)	5.337 (6.731)
$\frac{\text{Immigrants}}{\text{Population 1991}}$				-0.149 (0.780)
Main Controls	No	Yes	Yes	Yes
Additional Controls	No	No	Yes	Yes
Observations	270	270	270	270

Notes: Standard errors in parenthesis. The table reports Instrumental Variable first-difference coefficient estimates. The unit of observation is a Magisterial District in South Africa for which information is available in 1970, 1991 and 1996. The dependent variable is the change in total number of Non-state conflict events recorded in the district in the UCDP-GED v1.5 dataset between 1991 and 1996. ΔRQ is the change in within-black polarization in the same period. RQ_{1991} is the 1991 polarization, which we use as instrument. Immigrants/Population 1991 is the number of blacks immigrating in the district between 1991 and 1996 as a share of the black population in 1991. The set of Main Controls includes the change in: black population, total population, night-time satellite light value in logs. The set of Additional Controls includes the change in: rural population, number of individuals with no education, number of individuals unemployed, number of individuals out of the labor force, number of citizens of South Africa in logs (Sources: UCDP-GED v1.5, Statistics South Africa 1991, 1998; NOAA 2012).

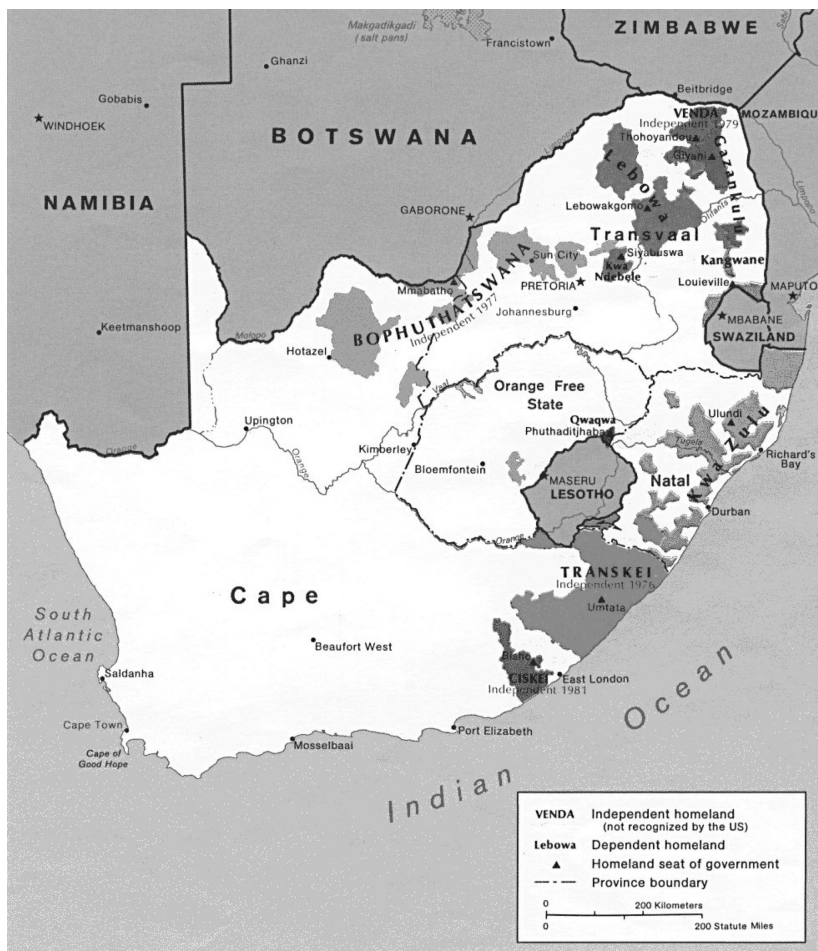


FIGURE A.1. . Map of Bantustans in South Africa as of 1986, produced by the U.S. CIA (Source: University of Texas at Austin 1986).

Map of Magisterial Districts in South Africa

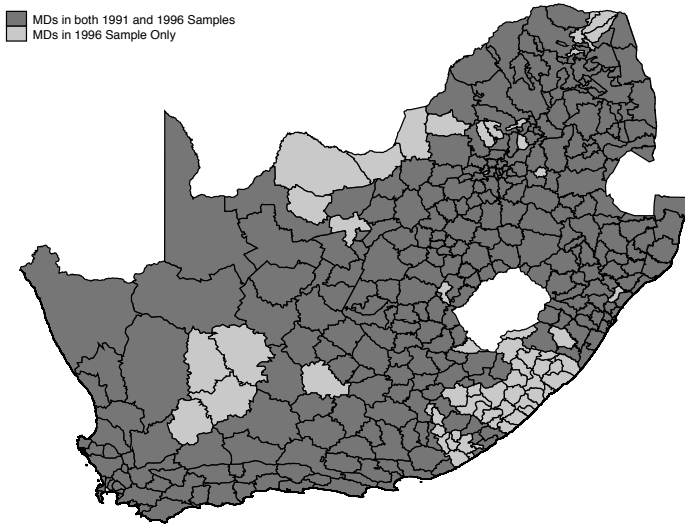


FIGURE A.2. Map of magisterial districts in South Africa. Districts in dark grey are those for which we can recover information from both the 1991 and 1996 Census, and thus belong to the study sample. Those Bantustans which were already granted independence are not covered by the 1991 Census of the Republic of South Africa (Sources: Statistics South Africa 1991, 1998).

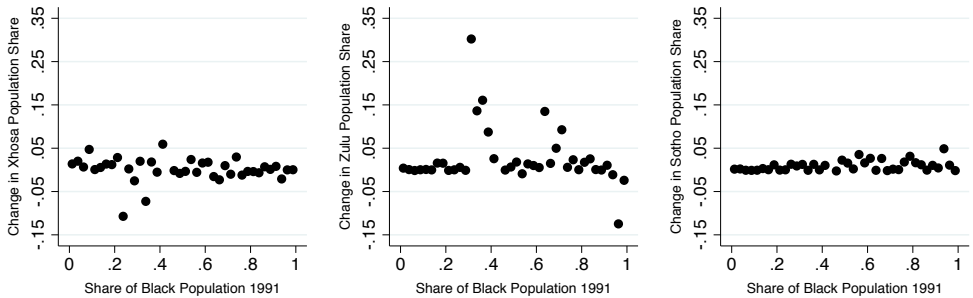


FIGURE A.3. Ethnic group populations per district: changes 1991–1996. The figure shows the changes in the population share of each ethnolinguistic group in each Magisterial District plotted over the initial share of blacks. Observations are those districts for which information is available in both 1991 and 1996. 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 2.5% (Sources: Statistics South Africa 1991, 1998).

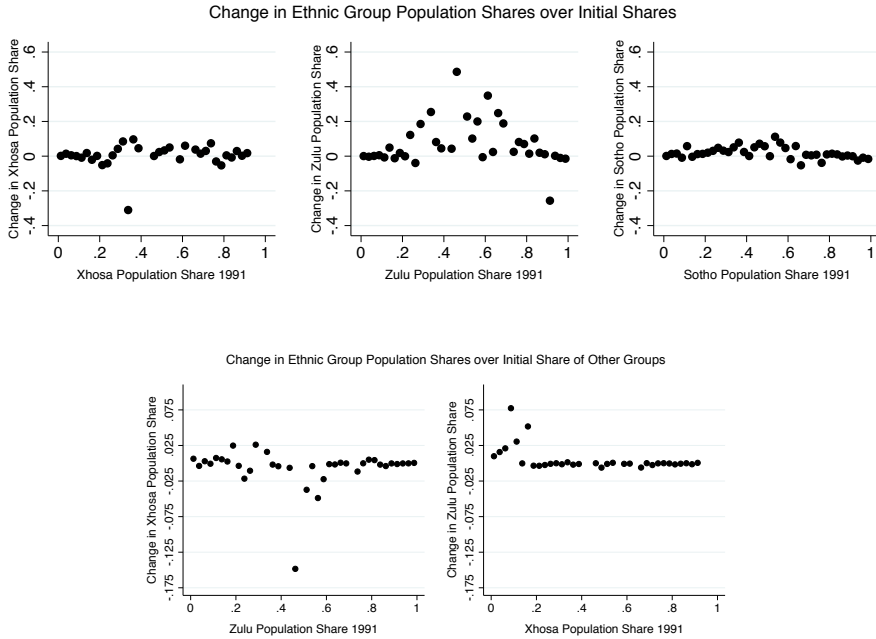


FIGURE A.4. Ethnic group populations per district: changes 1991-1996. . The top graphs in the figure show the change in population share of each of the three biggest black ethnolinguistic groups in each Magisterial District plotted over the initial share of each group. The bottom graphs plot instead the changes for the two biggest groups, Xhosa and Zulu, over the population share of the other in 1991. Observations are those districts for which information is available in both 1991 and 1996. 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 2.5%. Taking these figures and Figure A.3 together, we can notice the following patterns. Districts where the population share of Zulu grew the most between 1991 and 1996 are those where both the Zulu population share and the share of blacks in 1991 were large but not close to one. According to the bottom graphs, the Xhosa population share was very low in 1991 in these same districts. The Zulu group population seems thus to have increased more than proportionally in those districts with a relevant presence of non-black population, and where the Zulu themselves were already a large share of the population. Similarly, looking at changes in the Xhosa population shares, these seem to be disproportionately more positive in those districts where the Zulu group population share was low in 1991, and more negative otherwise. Again, the most important changes in absolute terms seem to be observed where non-black population shares were positive in 1991 (Sources: Statistics South Africa 1991, 1998).

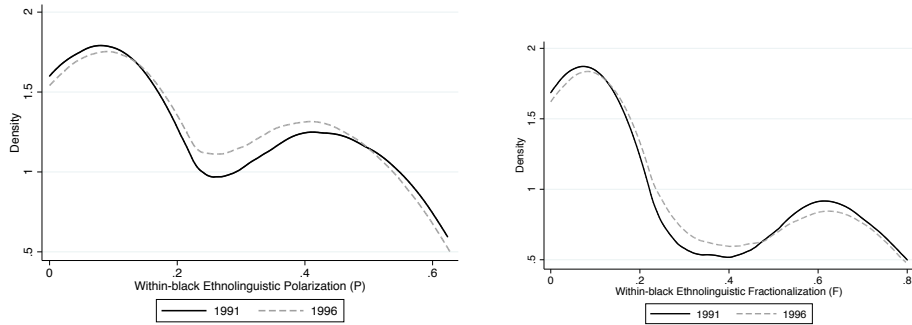


FIGURE A.5. Ethnic distribution across districts. The figures plot the distributions of the within-black ethnolinguistic polarization P and fractionalization F across districts in South Africa in the years 1991 and 1996. Districts are weighted according to their population of blacks. A Kolmogorov-Smirnov test for equality of distribution does not reject the null hypothesis of equality *between* the two distributions. Nonetheless, we find evidence of substantial *within*-variation, as shown in Figure 1 (Sources: Statistics South Africa 1991, 1998).

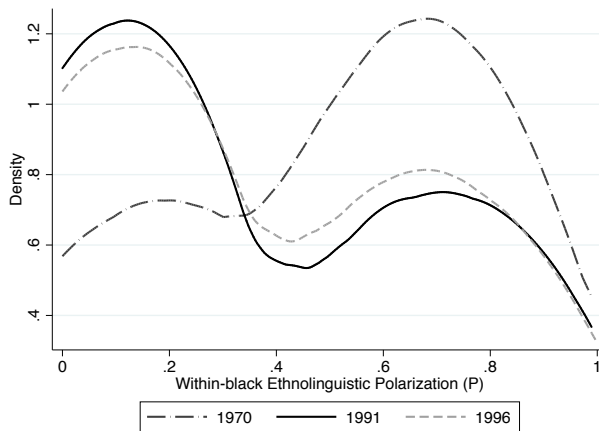


FIGURE A.6. Ethnic polarization across districts: 1970–1996. The figure plots the distribution of the within-black ethnolinguistic polarization index RQ across districts in South Africa in the years 1970, 1991, and 1996. Districts are weighted according to their population of blacks. The figure shows how the apartheid legislation and initiatives lead to a shift in the distribution, with an increase in the density of less polarized districts. The distribution moves back towards its 1970 shape in 1996 (Sources: Statistics South Africa 2011, 1991, 1998).

Appendix B: Data Appendix

This Section provides additional information on the data sources and variables that we use in the empirical analysis.

Census Data - We use data from the South African Census of 1970, 1991, 1996, 2001, and 2011 to derive information on the population size of African native groups and races in each magisterial district (Statistics South Africa 2011, 1991, 1998, 2004, 2014). We then use this information to compute the value of ethnic distribution measures (see Section 3 for further details). We use the same data to also identify respondents belonging to the top 10% of the overall black income distribution, and compute the value of the ethnic inequality index. We also derive a number of additional socio-economic variables: total 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. In the derivation of all these variables we collapse individual-level data at the district level using population weights.

Night-time Lights - We follow Michalopoulos and Papaioannou (2013, 2014), and use National Oceanic and Atmospheric Administration (NOAA 2012) night-time light satellite images data for 1992 (the first available year), 1996, 2001 and 2011 to derive a proxy for economic conditions in South Africa at the local level. Michalopoulos and Papaioannou (2014, 2013) provide an extensive discussion of these data and the validity of using night-time light as proxy for economic conditions in the African territories. See also Doll, Muller, and Morley (2006) and Sutton, Elvidge, and Ghosh (2007). We average night-time light density within each district across 30 arc-second grid areas (approximately 1 square kilometer at the equator). In our regression specifications, 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.

Terrain Ruggedness - We take this measure from Nunn and Puga (2012). Data are available at <http://diegopuga.org/data/rugged/>. As for night-time lights, we average

the ruggedness measure across 30 arc-second grid areas (approximately 1 square kilometer at the equator) within each district.

Terrain Slope Index - We use the measure from the Global Aero-ecological Zones (GAEZ) project (IIASA/FAO 2012). This is available at the 5 arc-minute by 5 arc-minute (approximately 9 km by 9 km at the equator) grid-cell level. We average its values across grid cells within each district.

Accessibility Index - We use the measure from the Global Aero-ecological Zones (GAEZ) project (IIASA/FAO 2012). This is computed as the estimated travel time to the nearest city with 50,000 or more inhabitants in year 2000. The measure is available at the 5 arc-minute by 5 arc-minute (approximately 9 km by 9 km at the equator) grid-cell level. We average its values across grid cells within each district.

Agricultural Suitability - We take this measure from Ramankutty, Foley, Norman, and McSweeney (2002). The measure is available at the 0.5 by 0.5 decimal degrees (approximately 55 km by 55 km at the equator) grid-cell level. It takes values between 0 and 1, and indicates for each the probability that a particular grid cell may be cultivated. We average its value across cells within magisterial districts to derive a district-level measure of agricultural suitability. See also Michalopoulos (2012) for a detailed explanation of this measure.

References

- Acemoglu, Daron, Camilo García-Jimeno, and James A. Robinson (2012). "Finding Eldorado: Slavery and Long-run Development in Colombia." *Journal of Comparative Economics*, 40(4), 534–564.
- Acemoglu, Daron, Suresh Naidu, Pascual Restrepo, and James A. Robinson (2016). "Democracy Does Cause Growth." *Journal of Political Economy*. Forthcoming.
- Acemoglu, Daron and James A. Robinson (2006). *Economic Origins of Dictatorship and Democracy*. Cambridge University Press, Cambridge, UK, URL <http://www.cambridge.org/us/catalogue/catalogue.asp?isbn=0521855268>.
- Alesina, Alberto, Arnaud Devleeschauwer, William Easterly, Sergio Kurlat, and Romain Wacziarg (2003). "Fractionalization." *Journal of Economic Growth*, 8(2), 155–94.
- Alesina, Alberto, Stelios Michalopoulos, and Elias Papaioannou (2016). "Ethnic Inequality." *Journal of Political Economy*, 124(2), 428–488.
- Altonji, Joseph G., Todd E. Elder, and Christopher R. Taber (2005). "Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools." *Journal of Political Economy*, 113(1), 151–184.

- Alvarez-Rivera, Manuel (2004). *Election Resources on the Internet: Republic of South Africa 1994 General Election*. available at <http://www.electionresources.org/> (accessed August 28, 2015).
- Angrist, Joshua D., Guido W. Imbens, and Donald B. Rubin (1996). "Identification of Causal Effects Using Instrumental Variables." *Journal of the American Statistical Association*, 91(434), pp. 444–455.
- Bazzi, Samuel and Matthew Gudgeon (2016). "Local Government Proliferation, Diversity, and Conflict." Boston University.
- Berman, Nicolas, Mathieu Couttenier, Dominic Rohner, and Mathias Thoenig (2017). "This Mine Is Mine! How Minerals Fuel Conflicts in Africa." *American Economic Review*, 107(6), 1564–1610.
- Besley, Tim and Marta Reynal-Querol (2014). "The Legacy of Historical Conflict: Evidence from Africa." *American Political Science Review*, 108, 319–336.
- Burgess, Robin, Remi Jedwab, Edward Miguel, Ameet Morjaria, and Gerard Padró i Miquel (2015). "The Value of Democracy: Evidence from Road Building in Kenya." *American Economic Review*, 105(6), 1817–51.
- Carver, Richard (1996). *Kwazulu-Natal - Continued Violence and Displacement*. WRITENET. available at <http://www.refworld.org/docid/3ae6a6bc4.html> (accessed August 10, 2014).
- Case, Anne and Angus Deaton (1999). "School Inputs and Educational Outcomes in South Africa." *The Quarterly Journal of Economics*, 114(3), pp. 1047–1084.
- Caselli, Francesco and Wilbur John Coleman (2013). "On the Theory of Ethnic Conflict." *Journal of the European Economic Association*, 11, 161–192.
- Cervellati, Matteo, Giorgio Chiovelli, and Elena Esposito (2017). "Bite and Divide: Ancestral Exposure to Malaria and the Emergence and Persistence of Ethnic Diversity in Africa." University of Bologna.
- Cervellati, Matteo, Piergiuseppe Fortunato, and Uwe Sunde (2012). "Consensual and Conflictual Democratization." *The B.E. Journal of Theoretical Economics*, 12(1), 1–51.
- Cervellati, Matteo, Piergiuseppe Fortunato, and Uwe Sunde (2014). "Violence During Democratization and the Quality of Democratic Institutions." *European Economic Review*, 66, 226 – 247.
- Cervellati, Matteo and Uwe Sunde (2014). "Civil Conflict, Democratization, and Growth: Violent Democratization as Critical Juncture." *The Scandinavian Journal of Economics*, 116(2), 482–505.
- Clark, Nancy L. and William H. Worger (2011). *South Africa. The Rise and Fall of Apartheid*. 2nd ed., Pearson.
- Conley, Timothy G. (1999). "GMM Estimation with Cross Sectional Dependence." *Journal of Econometrics*, 92(1), 1–45.
- Desmet, Klaus, Ignacio Ortuño Ortín, and Romain Wacziarg (2012). "The Political Economy of Linguistic Cleavages." *Journal of Development Economics*, 97(2), 322–338.
- Desmet, Klaus, Ignacio Ortuño Ortín, and Shlomo Weber (2009). "Linguistic Diversity and Redistribution." *Journal of the European Economic Association*, 7(6), 1291–1318.
- Doll, Christopher N.H., Jan-Peter Muller, and Jeremy G. Morley (2006). "Mapping Regional Economic Activity from Night-time Light Satellite Imagery." *Ecological Economics*, 57(1), 75 – 92.
- Dube, Oeindrila and Juan F. Vargas (2013). "Commodity Price Shocks and Civil Conflict: Evidence from Colombia." *Review of Economic Studies*, 80(4), 1384–1421.
- Eck, Kristine (2012). "In Data We Trust? A Comparison of UCDP GED and ACLED Conflict Events Datasets." *Cooperation and Conflict*, 47(1), 124–141.
- Eifert, Benn, Edward Miguel, and Daniel N. Posner (2010). "Political Competition and Ethnic Identification in Africa." *American Journal of Political Science*, 54(2), 494–510.
- Esteban, Joan, Laura Mayoral, and Debraj Ray (2012). "Ethnicity and Conflict: An Empirical Study." *American Economic Review*, 102(4), 1310–42.
- Esteban, Joan and Debraj Ray (1994). "On the Measurement of Polarization." *Econometrica*, 62(4), 819–51.

- Esteban, Joan and Debraj Ray (1999). "Conflict and Distribution." *Journal of Economic Theory*, 87(2), 379–415.
- Esteban, Joan and Debraj Ray (2008). "On the Saliency of Ethnic Conflict." *American Economic Review*, 98(5), 2185–2202.
- Esteban, Joan and Debraj Ray (2011a). "A Model Of Ethnic Conflict." *Journal of the European Economic Association*, 9(3), 496–521.
- Esteban, Joan and Debraj Ray (2011b). "Linking Conflict to Inequality and Polarization." *American Economic Review*, 101(4), 1345–74.
- Fearon, James D. (2004). "Why Do Some Civil Wars Last So Much Longer than Others?" *Journal of Peace Research*, 41(3), 275–301.
- Fearon, James D. and David D. Laitin (2011). "Sons of the Soil, Migrants, and Civil War." *World Development*, 39(2), 199–211.
- Guariso, Andrea and Thorsten Rogall (2017). "Rainfall Inequality, Political Power, and Ethnic Conflict in Africa." Trinity College Dublin.
- Harari, Mariaflavia and Eliana La Ferrara (2016). "Conflict, Climate and Cells: A Disaggregated Analysis." University of Pennsylvania.
- Hausman, Jerry A (1978). "Specification Tests in Econometrics." *Econometrica*, 46(6), 1251–71.
- Hoffman, Timm 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.
- IIASA/FAO (2012). *Global Agro-ecological Zones (GAEZ v3.0)*. available at <http://www.gaez.iiasa.ac.at/> (accessed April 20, 2012), Laxenburg, Austria and FAO, Rome, Italy.
- Karatnycky, Adrian (2005). *How Freedom is Won: From Civic Resistance to Durable Democracy*. Freedom House, New York.
- Lewis, M. Paul (2009). *Ethnologue: Languages of the World, Sixteenth edition*. SIL international Dallas, TX, available at <http://www.ethnologue.com/> (accessed March 11, 2012).
- Melander, Erik and Ralph 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, Stelios (2012). "The Origins of Ethnolinguistic Diversity." *American Economic Review*, 102(4), 1508–39.
- Michalopoulos, Stelios and Elias Papaioannou (2013). "Pre-colonial Ethnic Institutions and Contemporary African Development." *Econometrica*, 81(1), 113–152.
- Michalopoulos, Stelios and Elias Papaioannou (2014). "National Institutions and Subnational Development in Africa." *The Quarterly Journal of Economics*, 129(1), 151–213.
- Michalopoulos, Stelios and Elias Papaioannou (2016). "The Long-Run Effects of the Scramble for Africa." *American Economic Review*, 106(7), 1802–48.
- Montalvo, Jose and Marta Reynal-Querol (2005). "Ethnic Polarization, Potential Conflict and Civil War." *American Economic Review*, 95(3), 796–816.
- Montalvo, Jose and Marta Reynal-Querol (2010). "Ethnic Polarization and the Duration of Civil Wars." *Economics of Governance*, 11(2), 123–143.
- NOAA (2012). *Version 4 DMSP-OLS Nighttime Lights Time Series*. available at <http://www.ngdc.noaa.gov/dmsp/downloadV4composites.html> (accessed April 20, 2012).
- Novta, Natalija (2016). "Ethnic Diversity and the Spread of Civil War." *Journal of the European Economic Association*, 14(5), 1074–1100.
- Nunn, Nathan and Diego Puga (2012). "Ruggedness: The Blessing of Bad Geography in Africa." *The Review of Economics and Statistics*, 94(1), 20–36.
- Oster, Emily (2017). "Unobservable Selection and Coefficient Stability: Theory and Evidence." *Journal of Business Economics and Statistics*. Forthcoming.
- Papaioannou, Elias and Gregorios Siourounis (2008). "Democratisation and Growth." *The Economic Journal*, 118(532), 1520–1551.

- Posner, Daniel N. (2005). *Institutions and Ethnic Politics in Africa*. Cambridge University Press, New York.
- Raleigh, Clionadh, Andrew Linke, and Caitriona Dowd (2014). *Armed Conflict Location and Event Data Project (ACLED) Codebook 3*. Mimeo.
- Raleigh, Clionadh, Andrew Linke, Håvard Hegre, and Joakim Karlsen (2010). "Introducing ACLED: An Armed Conflict Location and Event Dataset: Special Data Feature." *Journal of Peace Research*, 47(5), 651–660.
- Ramankutty, Navin, Jonathan A. Foley, John Norman, and Kevin McSweeney (2002). "The Global Distribution of Cultivable Lands: Current Patterns and Sensitivity to Possible Climate Change." *Global Ecology and Biogeography*, 11(5), 377–392.
- Reed, Holly E. (2013). "Moving Across Boundaries: Migration in South Africa, 1950-2000." *Demography*, 50(1), 71–95.
- Reynal-Querol, Marta (2002). "Ethnicity, Political Systems, and Civil Wars." *Journal of Conflict Resolution*, 46(1), 29–54.
- Rohner, Dominic, Mathias Thoenig, and Fabrizio Zilibotti (2013). "Seeds of Distrust: Conflict in Uganda." *Journal of Economic Growth*, 18(3), 217–252.
- 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 (2004). *South African Census 2001 [dataset]. Version 1.1*. Pretoria: Statistics South Africa [producer]. Cape Town: DataFirst [distributor]. <http://www.datafirst.uct.ac.za/> (accessed August 31, 2015).
- Statistics South Africa (2011). *South African Census 1970 [dataset]. Version 1.2*. Pretoria: Statistics South Africa [producer]. Cape Town: DataFirst [distributor]. <http://www.datafirst.uct.ac.za/> (accessed August 31, 2015).
- Statistics South Africa (2014). *South African Census 2011 [dataset]. Version 1.2*. Pretoria: Statistics South Africa [producer]. Cape Town: DataFirst [distributor]. <http://www.datafirst.uct.ac.za/> (accessed August 31, 2015).
- Sundberg, Ralph, Mathilda Lindgren, and Ausra Padskocimaite (2010). *UCDP GED Codebook Version 1.0-2011*. Department of Peace and Conflict Research, Uppsala University.
- Sunde, Uwe and Matteo Cervellati (2014). "Democratizing for Peace? The Effect of Democratization on Civil Conflicts." *Oxford Economic Papers*, 66(3), 774.
- Sutton, Paul C., Christopher D. Elvidge, and Tilottama 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).