

# Does Local Ethnic Segregation Lead to Violence?: Evidence from Kenya

Kimuli Kasara

Columbia University\*

June 2016

## Abstract

When ethnic conflict is likely people seek safety in homogenous neighborhoods, but does this ethnic segregation decrease communal violence? This paper argues that local segregation causes violence by eroding interethnic trust. Although segregation reduces violent disputes between individuals, the possibility of positive interethnic contact is also lower in segregated areas. Where levels of interethnic trust are low, it is easier for political leaders and other extremists to build support for communal violence. I demonstrate that segregation increases the incidence of violence using a new dataset measuring ethnic composition and violence across approximately 700 small localities in Kenya's Rift Valley Province during Kenya's 2007/08 post-election crisis. Because segregation is likely endogenous to violence, I draw on Kenya's history of land settlement to instrument for segregation. I also demonstrate that it is unlikely that local segregation increases violence by increasing groups' organizational capacity for violence.

---

\*Address: Department of Political Science, Columbia University, International Affairs Building, 420 W. 118th Street, New York, NY 10027, [kk2432@columbia.edu](mailto:kk2432@columbia.edu). I thank Kate Baldwin, Shigeo Hirano, Hino Hiroyuki, Macartan Humphreys, Saumitra Jha, Bethany Lacina, Isabela Mares, Neelanjan Sircar, and seminar participants at Columbia University, Georgetown University, and the JICA Research Institute Conference on Ethnic Diversity and Economic Instability in Africa for helpful comments. Ayla Bonfiglio, Benjamin Clark, and Kate Redburn provided excellent research assistance and Eric Nyadimo, Jasper Mwenda, and G.O. Wayumbe advised on land registration and mapping. The Institute for Social and Economic Research and Policy at Columbia University and the Hoover Institution at Stanford University provided research support.

# 1 Introduction

Residential segregation along ethnic lines occurs when relations between members of different ethnic groups are poor. Ethnic conflict has increased ethnic segregation a wide range of settings – religious riots in India and Nigeria, sectarian violence in Ireland and Iraq, and civil war in the former Yugoslavia. Ethnic segregation is also costly because people are unwilling or unable to access opportunities and public services lying just outside their ethnic enclave. The costs of segregation on African Americans in the United States are well-documented (Massey and Denton 1993, Cutler and Glaeser 1997, Ananat 2011). However, segregation may also have a negative impact on members of nonmarginalized groups because it impedes potentially profitable trade and necessitates the creation of separate public services.<sup>1</sup> Despite the individual and economic costs that segregation imposes, policymakers are more likely to advocate or tacitly support ethnic segregation where conflict has occurred to keep the peace (Calame and Charlesworth 2009). However, people may willingly bear the costs segregation imposes if it reduces ethnic conflict. To individuals living where ethnic violence is likely to occur, a reasonable “working hypothesis” is that living in segregated areas reduces the probability of violent interactions with members of other ethnic groups.<sup>2</sup> However, if segregation makes communal violence more likely, residential sorting along ethnic lines may have negative externalities.

This paper argues that segregation increases communal violence by decreasing interethnic trust – that is, people’s willingness to believe the worst about the intentions of members of other ethnic groups. Existing research on communal violence suggests that interethnic mistrust exacerbates communal violence. Rumors and real or imagined hostile encounters between members of other ethnic groups play a prominent trigger and exacerbate communal violence (Brass 1997, Horowitz

---

<sup>1</sup>Research on Kosovo and Northern Ireland, for example, demonstrates that segregation leads to the costly duplication of public services (Shirlow and Murtagh 2006, Bloom, Hoxha, Sambunjak and Sondorp 2007, Deloitte 2007). Robinson (2013b) demonstrates that market segmentation is higher in segregated areas of Malawi.

<sup>2</sup>Glennerster, Miguel and Rothenberg (2010) and Steele (2009) show that safety is an important consideration for actual (and potential) victims of violence in civil wars.

2001). Ethnic tensions may place participants in a security dilemma, whereby conflict arises between members of two peacefully-intentioned groups when actions one side takes to enhance their security are perceived as threatening by the other side (Posen 1993, Scacco 2012). Several factors affect the likelihood that people will view members of other ethnic groups as hostile. de Figueiredo and Weingast (1997) and Wilkinson (2004) have argued that politicians who are more likely to lose office have more to gain from fomenting mistrust and violence. Although these political incentives to encourage violence are important, this paper focuses on local residential segregation as a factor contributing to interethnic violence.

When violence is likely to occur in a neighborhood, people must determine whether members of other ethnic groups pose a danger to them. Under this security dilemma, disputes between individuals from separate ethnic groups can escalate to episodes of communal violence because of false assumptions regarding out-group hostility and solidarity. One woman living in a clash-prone part of Rift Valley Province in Kenya described this dynamic well, “[t]hat is how it always starts: petty crimes which are blamed on entire communities, and before you know it, people are killing each other and houses are going up in flames. ... If a thief steals Baba Karanja’s cow, his people [Kikuyus] say the Kalenjin are stealing our livestock. If a Kalenjin is killed, the Kalenjin say the Kikuyu are finishing us, then the fighting begins.”<sup>3</sup>

Although ethnic segregation reduces interaction across ethnic lines, it also alters the character of and social meaning given to the interethnic contact that does occur. Several studies find that people living near members of other ethnic groups are more tolerant than those living in homogeneous places. Where interethnic trust and tolerance are low, elites with an interest in fomenting ethnic violence may find it easier to build support for ethnic violence. Moreover, where interethnic trust is low people may be more willing to stigmatize coethnic perpetrators of violence (Horowitz 2001).

The “working hypothesis” that segregated areas are safer for individuals also introduces an empirical challenge to the study of the relationship between local segregation and violence. Because

---

<sup>3</sup>“Why Police Did Not Fire Even a Single Shot in Trouble Spots.” *Daily Nation* (Nairobi), August 21, 2010.

self-segregation is a rational response to conflict for ordinary people seeking to reduce their exposure to violence and intolerance, we are more likely to observe segregation where relations between members of different ethnic groups are fraught. Even where ethnic violence has not occurred, local ethnic demography is likely endogenous to the quality of intergroup relations. As Schelling (1971) argues, neighborhoods may become segregated if even a small minority of people have a strong preference to live in a homogenous area. This paper, in contrast to existing literature on segregation and violence, addresses this endogeneity problem using instrumental variables regression.

This paper examines violence and segregation across 700 localities in Kenya's Rift Valley Province during an episode of political violence that took place between late December 2007 to March 2008. In this period roughly 1,100 people were killed, and 350,000 people displaced from several parts of the country (Kenya National Commission on Human Rights 2008). Although a disputed presidential election triggered these violent events, there is considerable local variation in the incidence of violence and many people were targeted for violence based on their ethnic identity and partisan affiliation, which often coincided.

To instrument for ethnic segregation, I draw on Kenya's colonial and post-colonial history of migration and settlement. By the end of the colonial period in 1964, about half of the agricultural land in Kenya had been transferred to Europeans and most of this land was in the Rift Valley (Okoth-Ogendo 1991, Sorrensen 1968). Africans not employed on European farms were moved to ethnically exclusive native reserves. The likely end of British colonial rule altered ethnic demography in Kenya, particularly in Rift Valley Province. In the 1950s, restrictions on internal migration by Africans were removed, and most land held by European farmers was transferred to African ownership through private purchases and government-sponsored schemes. Between 1963 and 1979 the population of some districts more than doubled due to migration (Oucho 2002). Africans formed land-buying companies and cooperatives to buy large farms, and these organizations were often ethnically homogenous. Because of this form of joint migration, farms are occupied by members of different ethnic groups in some areas.

Drawing on an original dataset of colonial-era cadastral (property) boundary maps, I use the extent to which a locality was divided by the boundaries of farms during the colonial period to instrument for present-day segregation. The potential contemporary segregation of locations that were more subdivided by property boundaries at the end of the colonial period is higher for two reasons. First, in the areas that were most divided by the boundaries of European farms, more separate transactions transferred land from Europeans to Africans. Therefore, adjacent farms are more likely to be settled by members of different ethnic groups. Second, farm boundaries, like railroad tracks in U.S. cities, facilitate self-segregation by reducing the coordination costs associated with segregation (Schelling 1971, Ananat 2011).

Because it is difficult to measure the incidence of conflict at a highly disaggregated level, I use two proxies for violence – the number of internally displaced persons from a location and the incidence of post-election fires in a location. Using these two measures, I show that segregation is positively correlated with violence in OLS regressions and if I instrument for segregation using colonial-era farm boundaries. However, because the IV estimates of the effect of segregation are much larger than the OLS estimates, I also show that the farm boundaries index is itself positively correlated with violence. Because spatial separation of groups in an area is related to its ethnic diversity, I demonstrate these findings are robust to alternative ways of conceptualizing ethnic segregation and diversity.

Besides arguing that segregation causes violence by making people more suspicious of members of other ethnic groups, I present evidence it is unlikely that segregation increases violence by enhancing organizational capacity. To do so, I take advantage of the fact that geography makes coordinating violence more difficult. If segregation increased violence by making it easier to carry out attacks, we would expect segregation to have a greater effect on violence where organizing violence is difficult; however, the evidence suggests this is not the case.

This paper advances research on the causes of communal violence in three ways. First, although local segregation is common, its relationship to communal violence remains understudied

because it is difficult to collect fine-grained data on local demography where conflict occurs. Few empirical papers demonstrate that violence is more likely to occur in segregated areas. Olzak, Shanahan and McEneaney (1996) demonstrate that the incidence of race riots is higher in more segregated American cities, and Barron, Kaiser and Pradhan (2004) show that villages in segregated districts have higher levels of conflict in Indonesia. Weidmann and Saleyhan (2011) and Field, Levinson, Pande and Visaria (2008) find that ethnic segregation decreased conflict over time in Ahmedabad and Baghdad respectively. Rather than tracking segregation in one place over time, this paper explores variation in the level of segregation and the incidence of violence across multiple localities. This cross-sectional research design is useful because changes in segregation and the incidence of violence in any one place over time are overdetermined.

Second, the findings presented here confirm the intuition that within-group heterogeneity may reduce ethnic violence. Scholars have argued that overlapping social cleavages based on features such as language, income, and kinship increase ethnic conflict. (Horowitz 1985, Ostby 2008, Dunning and Harrison 2010, Sambanis and Shayo 2013). Residential segregation, which is the unequal distribution of ethnic groups across space, can be viewed as an additional overlapping characteristic hindering interethnic cooperation.

Finally, ethnic demography has been used to explain a wide variety of outcomes in the social sciences, and this paper suggests we cannot always treat it as exogenous. Economists and political scientists studying the implications of racial diversity in the United States have begun to address this empirical challenge, but scholars focused on other regions have not. Because researchers can rarely manipulate where people live experimentally, political scientists with an understanding of local context are well placed to use history to identify how ethnic demography affects social outcomes.

This paper proceeds as follows. The following section makes the argument that segregation increases communal violence by eroding interethnic trust. The third section provides context to the empirical analysis by describing the history of ethnicity and violence in Rift Valley Province.

The fourth section describes the dataset and my empirical strategy. The fifth and sixth sections present the paper’s main findings and robustness checks respectively. The final section explores an alternative explanation for the relationship between segregation and violence.

## **2 The Relationship Between Segregation and Violence**

Although segregation occurs where interethnic relations are fraught, I argue that local segregation increases violence by contributing to mistrust along ethnic lines. A reasonable “working hypothesis” is that segregation reduces violence by limiting the number of potentially violent interactions between members of different ethnic groups. However, segregation also decreases positive contact across ethnic lines. I argue below that the absence of interethnic contact is relevant where communal violence occurs because political leaders and other extremists can better build support for communal violence where interethnic trust is low.

Recent theoretical research uses agent-based modeling to determine the relationship between segregation and violence. Weidmann and Saleyhan (2011) present an agent-based model in which civilians migrate to safer areas to avoid being attacked by insurgents from another ethnic group. Their model predicts that segregation makes violence less likely by reducing the number of members of minority groups who are vulnerable to attack in any area. Bhavnani, Donnay, Miodownik, Mor and Helbing (2014) develop a richer agent-based model to account for murders and assaults in Jerusalem. “Social distance”, which they define as the level of tension between members of different ethnic groups, is an important parameter in their model. They argue that segregation is more likely to produce peace where the social distance between two groups is high.

These agent-based models of segregation and violence treat underlying tensions between ethnic groups as exogenous to the level of segregation between them. However, evidence suggests that local segregation affects people’s attitudes towards members of other ethnic groups. Research on the psychology of prejudice shows that integration may improve intergroup relations. Allport

(1954) and subsequent authors have argued that, under certain conditions, interethnic contact reduces prejudice by allowing people to correct false beliefs about members of other ethnic groups (Pettigrew 1998). Living near members of other ethnic groups increases the potential for interethnic contact and, therefore, interethnic trust (Blau 1994). However, Putnam (2007) argues that a negative correlation between interethnic trust and local ethnic diversity arises because residents in ethnically diverse areas do not interact in ways that promote generalized trust rather than group-specific trust. Empirical research on the effect of local ethnic demography on attitudes estimates the effect of the local population share of one ethnic group on the attitudes of members of another group. Kinder and Mendelberg (1995) and Welch, Sigelman, Bledsoe and Combs (2001), find that local racial diversity increases whites' tolerance of African Americans. However, there is little consensus in the extensive *empirical* literature on whether living near members of other ethnic groups produces tolerance. Several studies focus on factors mediating the relationship between local ethnic demography such as relative group status, between-group inequality, and other group-specific attributes (Bobo and Hutchings 1996, Dixon 2006, Oliver and Wong 2003, Bowyer 2009).

With few exceptions, studies of the effect of local context on attitudes do not use data from places with a history of violence. This gap in the literature is important because contact probably would not produce tolerance where violent mobilization has already occurred. However, two quantitative studies of the effect of local context in conflict-prone places confirm that living in segregated areas is correlated with low levels of tolerance. Massey, Hodson and Sekulic (1999) find that ethnic tolerance was higher in ethnically mixed areas just before the civil war in Yugoslavia. Also, Kasara (2013) finds that Kenyans living in segregated localities were less trusting of members of other ethnic groups two years before the events examined in this paper. Examining segregation at a higher geographic scale, Robinson (2013a) finds that people living in ethnically diverse districts trust co-ethnics relatively more than others, but that segregation across districts within a country decreases interethnic trust.

Why is interethnic trust relevant for communal violence? Existing research on communal



violence confirms that only some of the numerous disputes that take place between members of different ethnic groups spark large-scale instances of communal violence (Brass 1997, Horowitz 2001). When large-scale communal violence occurs, there is often a great deal of spatial variation in its severity. Political elites incentives explain much of the spatial and temporary variation in the incidence of communal violence (Wilkinson 2004).

Although relatively few people organize and perpetrate ethnic violence, Horowitz (2001) argues that public approval of violent events is an often neglected reason riots occur where they do. Perpetrators of communal violence often remain in their home communities, and their coethnics are likely to know what they have done. Therefore, a high degree of interethnic distrust makes it easier for those who do not participate in violence and who would ordinarily condemn theft, murder, and rape to justify these acts when committed against ethnic outsiders.

Empirical research on the incidence of Hindu-Muslim riots in India also demonstrates the link between interethnic trust and violence. Varshney (2002) argues that interethnic engagement through civic organizations promotes peace because participants in these organizations can prevent small interethnic clashes from escalating into large-scale riots. Varshney argues that everyday contact and residential integration can promote peace only in rural areas because cities are large and complex. Therefore, in cities interethnic contact in civic associations must supplement everyday contact to promote peace. Jha (2013) argues that institutions supporting ethnic tolerance are a legacy of medieval interethnic complementarities in trade. He demonstrates these historical legacies can account for variation in *both* the incidence of Hindu-Muslim riots and the degree to which members of the Muslim minority trust the Hindu-dominated state.

Research on the spatial separation of groups and the incidence of civil war suggests an alternative mechanism by which segregation may increase communal violence. Several scholars argue that ethnic clustering increases the organizational capacity of groups.<sup>4</sup> Toft (2003) and Lichbach

---

<sup>4</sup>Toft (2003) and Matuszeski and Schneider (2006) argue that geographic concentration gives elites the incentive to use violence to gain political autonomy. I focus on organizational capacity rather than political autonomy as a goal because local segregation probably could not give ethnic groups the political autonomy these authors focus on at the

(1995) claim that geographic concentration makes the coordination of violence easier. Scholars proposing partition as a solution to ethnic conflict claim that ethnic clustering increases war-fighting capabilities so much that members of ethnic groups will fight to attain clear boundaries separating them from other ethnic groups (Posen 1993, Kaufmann 1996). Although organizing acts of communal violence poses less of an organizational challenge than mounting an insurgency, I explore the possibility that segregation increases violence by increasing perpetrators' organizational capacity in Section 6.

### **3 Ethnicity and Violence in Rift Valley Province**

Local ethnic demography varies considerably across Rift Valley Province, which comprised 22% of Kenya's total population (6.4 million people) in 1999. The province is shared, to varying degrees, by five of Kenya's largest ethnic groups. To provide context for the empirical analysis that follows, I describe the origins of ethnic demography and conflict in the region.

Rift Valley Province was highly ethnically diverse upon independence because of both geography and colonial policy. In 1962, the Kalenjin were the largest and most populous group in the region with 62% of the population. However, members of other ethnic groups had a significant presence, including the Kikuyu (17%), Maasai (7%), Luhya (6%), Luo (3%), and Kisii (1%).<sup>5</sup> The Kalenjin themselves are divided into seven culturally-related sub-groups that formed a politically salient identity after the 1940s (Lynch 2011). The seven Kalenjin sub-groups listed in order of their size within the broader group in 1962 are: Kipsigis (55%), Nandi (14%), Tugen (9%), Keiyo (8%), Pokot (6%), Marakwet (5%), and Sabaot (2%).

Because the province is large and centrally located, it borders regions in which five Kenyan ethnic groups are a majority. Areas along provincial boundaries are more ethnically mixed. Also, smaller geographic scale examined in this paper.

---

<sup>5</sup>Data come from the 1962 census (Republic of Kenya 1964a). These figures consider 1962 locations and wards that overlap the present-day boundaries of Rift Valley Province and not the province as it was defined at the time.

multiple ethnic groups make overlapping claims, of varying historical accuracy, to parts of Rift Valley Province. Reviewing territorial claims Africans made to a British commission that proposed boundaries prior to independence, Médard (1999) finds 11 conflicting claims in or proximate to the province involving five separate ethnic pairs. However, territorial claims have been most politically salient for members of the Kalenjin and Maasai ethnic groups.

The removal of restrictions on internal migration and the end of colonial rule changed the ethnic landscape of the province. In the study area, the proportion of Kalenjins and Maasais has decreased relative to that of other ethnic groups. Members of some Kalenjin sub-groups have migrated to new areas within the province. Although migration has also occurred in less densely populated native reserve areas, transferring European-owned farms to Africans played a large role in this demographic change. Upon independence, few Africans had the resources to purchase settlers' farms outright. However, a handful of politically connected Kenyans acquired a large share of the land. Hornsby (2012) estimates that politically-connected individuals acquired a sixth of these farms. Ordinary people acquired land on settlers' farms in one of two ways. First, many Africans who settled former European farms purchased that land together by forming land-buying companies or cooperatives (Leo 1984). Joint land-buying schemes were often ethnically exclusive because it was difficult to collect information about potential settlement sites hundreds of miles away. In the immediate post-independence period, land-buying companies and cooperatives were dominated by Kikuyus, but Luhyas, Kisii, and Kalenjins also settled former European-held lands jointly (Leys 1974). Second, some white-owned farms were redistributed to Africans through government-sponsored settlement schemes in programs funded first by the British, and later by international donors. In these schemes, the highly concessional terms upon which land was sold encouraged rent-seeking. A policy of selling land to Africans who had worked on these farms as employees made it hard for British and later Kenyan officials to keep the population on government-sponsored schemes ethnically homogenous (Carey Jones 1965, Odingo 1971).

Ethnic diversity has also increased in parts of Rift Valley Province not granted to Europeans

during the colonial period. Restrictions on migration meant that prior to independence land was scarce and land prices relatively high in areas settled by agricultural groups, particularly the Kikuyu and Kisii (Mbithi and Barnes 1975, Ocho 1988). Also, members of the two Kalenjin groups most involved in agriculture – the Kipsigis and the Nandi – have also migrated outside their colonial-era native reserves. Even where migrants purchased land from individual African owners, the descendants of some view these sales as illegitimate.

Large-scale ethnic violence in the region dates to the introduction of multi-party politics in the 1990s. Ethnic clashes, often orchestrated by politicians, have occurred in three out of five general elections since 1990 (Klopp 2001). The episode of political violence examined in this paper was triggered by a disputed election result in which Kalenjin and Luo ethnic groups supported the candidate who lost, and Kikuyus supported the candidate who won. Although the presidential election triggered these violent events, in some parts of the Rift Valley, violence was planned before the date of the election (Government of Kenya 2008). Moreover, as in earlier periods of ethnic violence, perpetrators of violence openly expressed a desire to send members of the Kikuyu and Kisii ethnic groups “home” (Kenya National Commission on Human Rights 2008). Where perpetrators coordinated violence, there was a pattern of localized centralization as prominent people coordinated the youth to attack neighborhoods and settlements. Unlike in earlier periods of violence, Kikuyus in Kikuyu-majority parts of the province attacked Kalenjins and Luos in retaliation (Human Rights Watch 2008, Kenya National Commission on Human Rights 2008).

Kenyan specialists on land and conflict note that the most violence prone areas are those in which members of different ethnic groups were settled near each other on separate farms because, ignorant of their neighbors’ intentions, people believe the worst. A chief in a part of Trans Nzoia offered the unsolicited opinion that if he were to settle the Rift Valley all over again, he would not place people in ethnically exclusive farms.<sup>6</sup>

---

<sup>6</sup>Interview, August 20, 2008.

## 4 Empirical Strategy

I explore the relationship between segregation and violence in approximately 700 localities in Kenya's Rift Valley Province. Figure 1 shows a map of the study area.<sup>7</sup> *Locations*, the administrative area used in the analysis below, have an average population of 9,000 and a median area of 22 square miles in this region. Locations are the second smallest administrative unit in Kenya. Each location falls entirely within a district – the principle administrative jurisdiction – and a constituency – an electoral jurisdiction represented by a single Member of Parliament.<sup>8</sup>

It is difficult to acquire fine-grained data on violence and local ethnic demography where ethnic conflict is common. I attempted to find other types of local-level data on violence with little success. The police aggregate reports of crime to the district level and the press focuses on areas well-known for violence. I describe the measurement of both variables and my instrument for segregation below.

### 4.1 Measuring Violence

My first outcome variable is the number of people from each location in camps for internally displaced persons (*Number of IDPs*). These data come from an IDP profiling exercise intended to collect all data on IDPs in camps carried out by the Kenya Red Cross and International Organization for Migration in March and April 2008. The dataset covers roughly 75,600 individuals but excludes IDPs who settled outside of camps and are staying with relatives or renting accommodation.

The number of IDPs is not a count of violent events because people become internally displaced because of both actual and anticipated violence. In one sense, it is an advantage this measure captures people who fled preemptively because the number of IDPs from a location reveals

---

<sup>7</sup>I exclude all locations in arid and semi-arid (ASAL) districts because it is more difficult to measure ethnic diversity reliably in those areas.

<sup>8</sup>Summary statistics are in Table A1 in the Supplementary Appendix.

information about the state of interethnic relations in an area. Although these data were produced by the actions of both perpetrators and victims, it is worth considering which people are more likely to enter IDP camps given similar conditions in their location of origin. For example, IDPs residing outside of camps are likely to be wealthier or have closer social ties outside their district of origin. Therefore, IDPs in urban areas and those in more recently settled areas are likely to be under-represented. However, the difficult conditions within camps for IDPs make it less likely that many people would abandon their homes to reside there absent serious threats to their security. I discuss the potential consequences of selection in to displacement in Section 5.

In addition to data on IDPs, I use satellite images of fires in the post-election period (the five weeks following the election) as a proxy for violent events. An advantage of this proxy is that it can be geo-coded precisely (NASA/University of Maryland 2002).<sup>9</sup> Post-election fires are especially likely to represent violent events in Rift Valley Province because perpetrators of violence sought to remove ethnic others from their homes. Besides making a political statement, fires destroy property and make it less likely displaced people will return. Therefore, fires were more likely to be associated with post-election violence in Rift Valley Province. Aggregate data on criminal arrests confirms that the incidence of arson during the post-election period is *far* higher in this part of the country than in other provinces. Data on criminal cases pending trial or under investigation for offenses committed during the post-election period indicate that cases involving arson were the most common in the province; 89% of the 736 police cases in Rift Valley Province were for arson, as compared to 13% of 86 cases in Western and Nyanza Provinces combined.<sup>10</sup> Finally, the spatial and temporal distribution of fires suggests that they are a good proxy for violent events within the

---

<sup>9</sup>The active fire observations were generated in two stages. First, the satellites observe and record specific frequencies indicative of infra-red radiation. These thermal anomalies are fires but may also be emissions of hot gas or volcanic activity (Campbell 2007). Pixels on the satellite images are classified as containing an active fire using an algorithm developed and validated by Giglio, Descloitres, Justice and Kaufman (2003), which considers the temperature of an area, the temperature of surrounding areas, and other factors. No fires will be observed in an area if it lies under cloud cover. However, as these events occur during the dry season, it is unlikely that many fires are missed, and cloud cover is unlikely to be correlated with any of the explanatory variables of interest here.

<sup>10</sup>Data acquired by the author from Kenya Police.

Province. Incidence of fires was much higher in the 2007/08 period than in any of the previous five years, and, as has been noted by others, fires occurred in unusual areas (UNOSAT 2008, Anderson and Lochery 2008).

As with the data on IDPs, the validity of the fires proxy warrants discussion. Perpetrators of violence may be more likely to light them in segregated settings where damage to the property of coethnics is less likely to occur. Therefore, if perpetrators of using violence prefer setting fires, they may be more likely to target opponents in segregated areas. Also, fires may occur that are unobserved by satellites. To understand which instances of violence are under-represented by the fires data and which election-related fires may not be observed, I spoke to chiefs and other bureaucrats in Kericho and Trans Nzoia in August 2008. My impression, though it is not a systematic one, is that fires are more likely to be observed if there are several dwellings concentrated together, and all regressions include controls for both population and area. Fires cannot be observed when the weather is poor, but cloud cover is likely exogenous to politics.

Each of these proxies for violence is imperfect. However, as the discussion above indicates, it is unlikely that the measurement errors associated with each with each of these variables lie in the same direction. Therefore, finding a relationship between both the number of IDPs and fire incidence ought to increase our confidence that segregation causes violence.

## **4.2 Measuring Local Ethnic Composition**

Because ethnic demography is politically sensitive, the Kenyan government has not released data on local ethnic composition since 1962. I use the 2006 voter register to measure local ethnic composition. The register, which was publicly available, contains the full names of registered voters and their polling station. African names in Kenya are associated with particular ethnic groups and are used socially as a gauge of ethnic identity. To match last names to groups, I calculated the probability that a person with that surname fell into an administrative location in which a group was over 90% of the population in 1962. Appendix A.1 discusses the construction

and validity of this measure. I can establish where people are registered to vote using a map I created of approximately 17,500 polling stations (Appendix A.2).

Sociologists and economists disagree on the best way to measure segregation. I use Theil's Index because it measures segregation where there are multiple ethnic groups better than other commonly used indices (Hutchens 2001, Reardon and Firebaugh 2002). Theil's Index captures that in ethnically segregated settings, knowing where a person lives reduces one's uncertainty about his ethnic affiliation. Region A is more segregated than region B if knowing a person's subregion within A reduces our uncertainty about his ethnic identity to an extent greater than the same piece of information would in B. If groups are perfectly segregated across subregions, knowing a person's subregion allows one to predict their ethnic identity perfectly; if groups are perfectly integrated, knowing a person's locality adds no additional information about their ethnicity.

The underlying measure of ethnic diversity used in constructing Theil's Index is entropy, which is the uncertainty about the ethnic identity of a person in an area chosen at random. The entropy of location  $j$  is

$$Entropy_j = \sum_{g=1}^G \pi_{gj} \ln \frac{1}{\pi_{gj}}$$

where  $\pi_{gj}$  is the proportion of registered voters in location  $j$  who come from group  $g$ . It is worth considering how entropy differs from the most commonly used measure of ethnic diversity – ethnic fractionalization – which is the probability that two randomly selected people in a country would be from two separate ethnic groups. Both measures equal zero in perfectly homogenous settings. Measures of entropy and ethnic fractionalization are closely correlated but become less so as the number of ethnic groups increases because, unlike ethnic fractionalization, entropy does not have an upper bound at one.

Theil's Index measures how segregated ethnic groups are across polling stations in a location. The value of the index for location  $j$  is:



$$Theil's\ Index_j = 1 - \frac{\sum_{k=1}^{\kappa} \left( \frac{Proportion\ of\ Location\ j}{Voters\ Registered\ in\ Polling\ Station\ k} \right) \left( \frac{Entropy\ of\ Ethnic\ Distribution\ in\ Polling\ Station\ k}{Entropy\ of\ Ethnic\ Distribution\ in\ Location\ j} \right)}{\left( \frac{Entropy\ of\ Ethnic\ Distribution\ in\ Location\ j}{Entropy\ of\ Ethnic\ Distribution\ in\ Location\ j} \right)}$$

The value of Theil's Index ranges from zero, perfect integration, to one for perfect segregation. The maps in Figure 1 show the geographic incidence of ethnic fractionalization and segregation in the locations under study. The cross-hatched areas either lie outside Rift Valley Province or are arid and semi-arid districts excluded from our analysis.

### 4.3 Identification Strategy

Determining whether segregation causes violence is complicated because people seek safety in homogenous settings once violence has occurred and, as a consequence, segregation may be endogenous to violence. Even in places with no history of ethnic violence, residential sorting may occur where ethnic tensions are high and residents anticipate future violent conflict due to factors scholars cannot observe. If so, OLS estimates of the effect of segregation on violence may suffer from omitted variable bias.

In this paper, the boundaries of European farms are a plausibly exogenous source of variation in segregation in an area. The logic of the instrument is as follows. First, a history of ethnically exclusive joint migration means that more subdivided locations are more likely to have a farm settled by one ethnic group. Second, as Schelling (1971) argues, segregation is more likely to occur when neighborhoods have fixed, well-defined borders because they shape people's expectations about the likely future ethnic composition of an area. Consistent with Schelling's argument, Ananat (2011) demonstrates that Northern cities in which nineteenth century railroad tracks created clearly defined neighborhoods are more likely to be segregated. I use the degree to which an

area was subdivided by the boundaries of European farms during the colonial period to instrument for segregation during the 2008-07 post-election crisis. Following Ananat (2011), the *Colonial Farm Division Index (CFDI)* is defined as follows:

$$CFDI = 1 - \sum_{f=1}^L \left( \frac{a_f}{A} \right)^2$$

where  $a_f$  is the area of a plot in the location, there are  $L$  farms (or parts of farms) in the location and  $A$  is the total area of the location. The *CFDI* is a modified version of Herfindahl index of market concentration and equals zero if farm boundaries do not divide a location. An alternative instrument one could construct from these data is a count of farms, but, unlike the *CFDI*, that measure is closely correlated with a location's area.

Data on the boundaries of colonial-era farms come from 87 large-scale (1:50,000) cadastral (property) maps covering Rift Valley Province created by the Survey of Kenya. Figure 2, which places the 1964 property boundary map over a satellite image of the area in the present day, demonstrates these historical boundaries still affect the landscape today. The cadastral maps are dated 1964, a year after independence and two years after the beginning of the land settlement process in the White Highlands. One would ideally map all European farms by 1959, the last year of exclusive European ownership, but no detailed maps for that year exist. However, the maps used here are largely based on earlier maps issued by the British colonial government. Though further subdivision of these parcels took place between 1959 and 1964, this does not apply often. To be conservative about potential subdivisions between 1959 and 1964, I used the most aggregate block numbers to construct farm boundaries when calculating the *CFDI*.

Why might some locations have been more subdivided than others before independence? Land may be more productive where there are many small farms. The historical record gives a mixed answer to the question of whether Europeans were most likely to sub-divide the most valuable land. Later settlers acquired smaller plots once the government limited the size of plots that Eu-

Europeans could expropriate. Also, later settlers (who made up the majority of Europeans resident in Kenya upon independence) acquired land from the first generation of Kenyan land speculators. Some early settlers families, most famously the Delamere family, retained large holdings of their (Sorrensen 1968, Hughes 2006, Kenyanchui 1992, van Zwanenberg 1975, Osolo-Nasubo 1977).

To account for the possibility that land value and agricultural potential determine the degree to which a location is subdivided, all first-stage regressions control for whether a location falls in the White Highlands as well as *Area*, *Average Rainfall*, *Altitude*, *Proportion Rangeland*, *Terrain Ruggedness*, and *Population Density in 1962*. The instrument, therefore, is exogenous conditional upon including these variables. Appendix A.3 describes the origin and construction of these variables.

## 5 Findings

In the analysis that follows, the relationship between segregation and violence is modeled as:

$$Segregation = \alpha + \beta CFI_j + X_j' \gamma + \mu_j$$

$$Y_j = \alpha + \beta Segregation_j + X_j' \gamma + \epsilon_j$$

I use *Log. Fire Brightness* instead of the count of fires because, although instrumental variable models exist for nonlinear models, linear models yield estimates that are more robust and easier to interpret.  $X_j$  is a vector of other exogenous variables. Although contemporaneous factors correlate with violence in the post-election period, controlling for post-treatment variables would introduce bias (Gelman and Hill 2007). All continuous variables are standardized to have a mean of 0 and a standard deviation of 0.5. Table 1 shows OLS, reduced form, and two-stage least-squares estimates of the effect of segregation on violence. Although fire incidence and the number of people from a location in IDP camps are imperfect measures of violence, they both support the conclusion there

is more conflict in segregated areas.<sup>11</sup> The IV estimates of the effect of diversity are higher than the OLS estimates. Weak instruments, which bias IV estimates, are not a problem here as the F-statistic for the instrument (11) exceeds the conventional rule of thumb value of 10 (Table 1). However, because of the difference in magnitude between the OLS and IV regressions, for the sake of transparency, I include the reduced form estimates of the *CFDI* on violence in Table 1.

[Table 1 about here]

Colonial-era farm boundaries may have affected the post-colonial ethnic diversity of an area *and* the spatial separation of groups. Ethnic diversity and segregation are closely related concepts. Thiel's Index is zero in ethnically homogenous places. I evaluate whether the relationship between segregation and violence is robust to considering ethnic diversity by using a measure of segregation that takes a lower value in ethnically homogenous areas and by instrumenting for ethnic diversity.

In Table 2, I replace Theil's Index with a measure of segregation (the Mutual Information Index) with a lower value where there is less uncertainty about a resident's ethnicity (Frankel and Volij 2011). *Theil's Index* is the Mutual Information Index normalized for the ethnic diversity (Entropy) in a region.

[Table 2 about here]

Table 2 shows that segregation, measured using this alternative index, has a statistically significant effect on the number of IDPs and the incidence of fires. Although the estimated effect of segregation on violence is greater when I use the MI Index to measure segregation, the substantive effects are roughly the same for both measures of segregation.<sup>12</sup>

---

<sup>11</sup>As noted in Section 4, a location's distance to the borders of Rift Valley Province may affect whether displaced persons move to IDP camps rather than to areas in which members of their ethnic group are a majority. However, the effect of segregation on the number of IDPs remains the same controlling for a location's distance to the ethnic homelands of each of four groups that settled the former White Highlands. See Table A2 in the Supplementary Appendix.

<sup>12</sup>In Table A3 in the Supplementary Appendix I instrument for present-day ethnic diversity using the average distance as the crow flies of each location to the nearest area reserved for the Kikuyu, Kalenjin, Luhya, Luo, and Maasai

Because farm boundaries were determined before independence by British settlers and officials who deprived Africans of their land, it seems implausible that the location of farm boundaries were determined (or “assigned”) referring to the ethnic divisions that later became salient in post-colonial Kenya. However, members of different ethnic groups’ propensity to migrate and exposure to political violence vary. To examine ethnic group-specific effects, I calculated an index of segregation (the *Dissimilarity Index*) that measures the degree to which members of a specific group are separated from members of all other ethnic groups. I find that segregation of Kalenjins, Kikuyus, and Luhyas increases violence, but that of Kisiis and Maasais does not.<sup>13</sup>

While Table 1 shows that the degree to which colonial-era boundaries subdivide farms is correlated with segregation, does the *CFDI* satisfy the exclusion restriction? One post-independence policy that *may* lead to a violation of the exclusion restriction is the location of government-sponsored settlement schemes. The early policy of the Settlement Fund Trustees was to settle African farmers on smaller (“mixed farms”) rather than large-scale farms (Government of Kenya. Department of Settlement 1966, Morgan 1963). Therefore, it is possible that *government-sponsored* settlement schemes are more likely to be in areas that were more subdivided by European farm boundaries. Scholars have claimed that settlement schemes are more prone to violence (Anderson and Lochery 2008). Boone (2008) argues that settlement schemes cause violence because of the politicization of land allocation. Although settlement scheme locations are more segregated and ethnically diverse, the presence of a settlement scheme is not a statistically significant predictor of displacement and is negatively associated with post-election fires.<sup>14</sup>

Another potential threat to the validity of the instrument arises if locations with higher land value are more ethnically homogenous by law, places further from these ethnically exclusive reserves are more likely to have been settled by multiple groups upon independence. When both segregation and ethnic diversity are treated as endogenous regressors, the size of the coefficient on segregation is smaller, but the substantive effect remains roughly the same as the number of IDPs. However, the estimate of the effect of segregation on fires incidence narrowly misses statistical significance at conventional levels.

<sup>13</sup>See Table A4 in the Supplementary Appendix. The negative coefficient on Maasai segregation is surprising and is probably a result of being a small minority in the most violent areas, and the *Dissimilarity Index* shows more variance for small groups.

<sup>14</sup>See Table A5 in the Supplementary Appendix.

ues were more likely to be subdivided. As discussed in Section 4.3, the historical record on land value and farm size is mixed. I claim that the *CFDI* is plausibly exogenous conditional upon the inclusion of variables controlling for land value and agricultural potential, including rainfall, altitude, rangeland, terrain ruggedness, and population density in 1962. One way to validate measures of land value is to compare them to a contemporary measure of land productivity. The Normalized Difference Vegetation Index (NDVI) is commonly used to estimate crop yields (USGS 2006). The index measures the greenness of an area from satellite imagery because healthy vegetation absorbs different parts of the electromagnetic spectrum than unhealthy or sparse vegetation.<sup>15</sup> The NDVI is closely correlated with the exogenous variables in this paper.<sup>16</sup>

I find that segregation causes violence in Kenya's Rift Valley Province but is this finding unique to this context? One possible concern regarding the external validity of these findings is that the resettlement of this region took place over a fairly short period. However, rates of economic migration are high in many developing countries, and increased segregation has often followed communal violence. Also, although segregation is often considered an urban phenomenon, there is more violence in segregated areas when urban areas are dropped from the analysis.<sup>17</sup>

## 6 Mechanisms

The evidence suggests that segregation increases violence by decreasing interethnic trust rather than by making it easier to organize violence. As Kasara (2013) shows, self-reported levels of interethnic trust are lower in ethnically segregated locations across Kenya. The substantive effect of segregation is smaller for the outcome that requires greater organizational capacity – the production of fires. This finding is surprising because, as noted above, setting fire to dwellings in integrated

---

<sup>15</sup>Harris (2012) finds that the NDVI is positively correlated with fire incidence in Rift Valley Province during this the post-election violence of 2007/08.

<sup>16</sup>If current and past land productivity are identical, the main findings are robust to including a control for the *NDVI*. See Table A6 in the Supplementary Appendix

<sup>17</sup>Table A7 in the Supplementary Appendix.

areas may put a perpetrator's coethnics at risk.

I test for whether segregation affects violence by increasing organizational capacity using variation in the degree to which violence is easy to organize. If segregation works primarily through increasing organizational capacity, the effect of segregation on violence ought to be greater in places that are more remote and in places with difficult terrain.

To explore whether segregation has a greater effect on violence where the costs of organizing violence are higher, I divided locations using two variables that affect the cost of organizing violence – distance from a major road and the ruggedness of the terrain. Both distance and ruggedness are negatively associated with fire incidence and the number of IDPs, suggesting that they make it harder for perpetrators to organize violence. Table 3 shows IV regressions in which the sample was split in half by the median values of *Distance to a Major Road* and *Terrain Ruggedness* respectively. Segregation increases the number of IDPs in places that are closer to the road and easier to access. The substantive effect of segregation on the number of IDPs is higher in this subsample. By contrast, segregation does not affect the number of IDPs in remote or rugged places. Perhaps it is unsurprising that more people flee easily accessible areas. However, segregation is positively associated with fire incidence in locations close to a road, but not in locations far from one. Segregation is also positively associated with fire incidence in less rugged places and the estimate on segregation has a  $p$ -value of 0.051.

[Table 3 about here]

## 7 Conclusion

High degrees of ethnic segregation can arise if even a small minority of people prefer to live in ethnically homogenous settings (Schelling 1971). Therefore, it is unsurprising that ethnic tension and violence increase ethnic segregation. However, individual decisions to migrate to segregated areas may have the perverse effect of making communal violence more likely to occur. Although

fear of members of other ethnic groups may motivate people to move to ethnically exclusive places, ethnic segregation reduces the possibility of *both* positive and negative contact across ethnic lines. Therefore, segregation is correlated with low levels of interethnic trust. As existing research shows, widespread mistrust along ethnic lines helps to explain the severity of communal violence for two reasons. First, elites and extremists find it easier to mobilize support for violence where underlying mistrust is high. Second, perpetrators of violence are less likely to face social sanctions where the public supports violence against members of other ethnic groups.

Determining whether ethnic segregation causes violence presents important empirical challenges. Segregation is difficult to capture because its measurement requires fine-grained data on the distribution of ethnic groups within small areas. Data on actual local ethnic composition are controversial in violence-prone places. Therefore, most of what we know about the impact of segregation comes from the developed world or studies of temporal changes in segregation in a single city. Also, segregation is likely endogenous to violence because people seek ethnically homogenous areas when ethnic conflict occurs.

This paper estimates the effect of segregation on the incidence of violence across 700 localities in Kenya's Rift Valley Province after the disputed 2007 general election. In addition to using original data on local ethnic composition, this paper establishes that segregation causes violence by drawing on Kenya's colonial history to instrument segregation.

Although interethnic trust is lower in segregated areas, research on the spatial distribution of groups in civil wars suggests that segregation at the regional level improves groups' capacity to conduct violence. However, the alternative mechanism linking segregation and violence is unlikely to account for these findings. If segregation improved organizational capacity, we would expect it to have a greater effect on violence where geography makes it more difficult to organize violence, but this is not the case.

This paper draws attention to one reason communal violence recurs. Conflict may be self-reinforcing because violence alters local ethnic demography in ways that increase future conflict.



Although there is little policymakers can do to influence individuals' residential choices, they have greater control over other institutions and organizations that could promote ethnic integration. Therefore, the findings presented have implications for debates across the social sciences regarding whether political institutions that enhance the social and political separation of members of different ethnic groups reduce ethnic conflict.

## References

- Allport, Gordon W. 1954. *The Nature of Prejudice*. Reading, MA: Addison-Wesley.
- Ananat, Elizabeth Oltmans. 2011. "The Wrong Side(s) of the Tracks: The Causal Effects of Racial Segregation on Urban Poverty and Inequality." *American Economic Journal: Applied Economics* 3(2):34–66.
- Anderson, David M. and Emma Lochery. 2008. "Violence and Exodus in Kenya's Rift Valley, 2008: Predictable and Preventable." *Journal of Eastern African Studies* 2(2):328–343.
- Barron, Patrick, Karl Kaiser and Menno Pradhan. 2004. "Local Conflict in Indonesia: Measuring Incidence and Identifying Patterns." *World Bank Policy Research Paper* 3384.
- Bhavnani, Ravi, Karsten Donnay, Dan Miodownik, Maayan Mor and Dirk Helbing. 2014. "Group Segregation and Urban Violence." *American Journal of Political Science* 58(1):226–245.
- Blau, Peter M. 1994. *Structural Contexts of Opportunities*. Chicago, IL: University of Chicago Press.
- Bloom, J.D., I. Hoxha, D. Sambunjak and E. Sondorp. 2007. "Ethnic segregation in Kosovo's post-war health care system." *The European Journal of Public Health* 17(5):430–436.
- Bobo, Lawrence and Vincent L. Hutchings. 1996. "Perceptions of Racial Group Competition: Extending Blumer's Theory of Group Position to a Multiracial Social Context." *American Sociological Review* 61(6):951–972.
- Boone, Catherine. 2008. "Politically-Allocated Land Rights and the Geography of Electoral Violence: The case of Kenya in the 1990s."
- Bowyer, Benjamin T. 2009. "The Contextual Determinants of Whites Racial Attitudes in England." *British Journal of Political Science* 39(03):559–586.
- Brass, Paul. 1997. *Theft of an Idol: Text and Context in the Representation of Collective Violence*. Princeton, N.J.: Princeton University Press.
- Calame, Jon and Esther Charlesworth. 2009. *Divided Cities: Belfast, Beirut, Jerusalem, Mostar, and Nicosia*. Philadelphia, PA: University of Pennsylvania Press.
- Campbell, James D. 2007. *Introduction to Remote Sensing*. New York: The Guildford Press.
- Carey Jones, N. S. 1965. "The Decolonization of the White Highlands." *The Geographical Journal* 131(2):186–201.
- Cutler, David M. and Edward L. Glaeser. 1997. "Are Ghettoes Good or Bad?" *Quarterly Journal of Economics* 112(3):827–872.

- de Figueiredo, Rui J. P. and Barry R. Weingast. 1997. "The Rationality of Fear: Political Opportunism and Ethnic Conflict." *Working Paper. Institute for War and Peace Studies* .
- Deloitte. 2007. *Research into the financial cost of the Northern Ireland divide*. London, England: Deloitte.
- Department of Resource Surveys and Remote Sensing (DRSRS). Ministry of Environment and Natural Resources. 2003. *DRSRS GIS Database*. Nairobi: DRSRS.
- Dixon, Jeffrey C. 2006. "The Ties That Bind and Those That Don't: Toward Reconciling Group Threat and Contact Theories of Prejudice." *Social Forces* 84(4):2179–2204.
- Dunning, Thad and Lauren Harrisson. 2010. "Cross-Cutting Cleavages and Ethnic Voting: An Experimental Study of Cousinage in Mali." *American Political Science Review* 104(1):21–39.
- Field, Erica, Matthew Levinson, Rohini Pande and Sujata Visaria. 2008. "Segregation, Rent Control, and Riots: The Economics of Religious Conflict in an Indian City." *AEA Papers and Proceedings* 98(2):50510.
- Frankel, David M. and Oscar Volij. 2011. "Measuring school segregation." *Journal of Economic Theory* 146(1):1–38.
- Gelman, Andrew and Jennifer Hill. 2007. *Data Analysis Using Regression and Multi-level/Hierarchical Models*. New York: Cambridge University Press.
- Giglio, Louis, Jacques Descloitres, Christopher O. Justice and Yoram J. Kaufman. 2003. "An Enhanced Contextual Fire Detection Algorithm for MODIS." *Remote Sensing of Environment* 87(2-3):273–282.
- Glennerster, Rachel, Edward Miguel and Alexander D. Rothenberg. 2010. "Collective Action in Diverse Sierra Leone Communities." *NBER Working Paper* 16196.
- Government of Kenya. 2008. *Commission of Inquiry into Post-Election Violence (Waki Report)*. Nairobi: Government of Kenya.
- Government of Kenya. Department of Settlement. 1966. *Annual Report, 1965-1966*. Nairobi: Government of Kenya.
- Harris, J. Andrew. 2012. "'Stain Removal': Measuring the Effect of Violence on Local Ethnic Demography in Kenya." *Working Paper, Harvard University* .
- Hornsby, Charles. 2012. *Kenya: A history since independence*. London: I. B. Tauris. DT433.58 .H67 2012g.
- Horowitz, Donald L. 1985. *Ethnic Groups in Conflict*. Berkeley: University of California Press.
- Horowitz, Donald L. 2001. *The Deadly Ethnic Riot*. Berkeley: University of California Press.

- Hughes, Lotte. 2006. *Moving the Maasai: A Colonial Misadventure*. Basingstoke, New York: .
- Human Rights Watch. 2008. *Ballots to Bullets: Organized Political Violence and Kenya's Crisis of Governance*. New York: Human Rights Watch.
- Hutchens, Robert. 2001. "Numerical Measures of Segregation: Desirable Properties and Their Implications." *Mathematical Social Sciences* 42(1):13–29.
- International Livestock Research Institute (ILRI). 2007. *Kenya Contours (Almanac Characterization Tool)*. Nairobi: Kenya.
- Jha, Saumitra. 2013. "Trade, Institutions and Religious Tolerance: Evidence from India." *American Political Science Review* 107(4).
- Kasara, Kimuli. 2013. "Separate and Suspicious: Local Social and Political Context and Ethnic Tolerance in Kenya." *The Journal of Politics* 75(04):921–936.
- Kaufmann, Chaim D. 1996. "Possible and Impossible Solutions to Ethnic Wars." *International Security* 20(4):136–175.
- Kenya National Commission on Human Rights. 2008. "On the Brink of the Precipice: A Human Rights Account of Kenyas Post-2007 Election Violence."
- Kenyanchui, Simon S.S. 1992. European Setler Agriculture. In *An Economic History of Kenya*, ed. William Robert Ochieng and Robert M. Maxon. Nairobi, Kenya: East African Educational Publishers.
- Kinder, Donald R. and Tali Mendelberg. 1995. "Cracks in American Apartheid: The Political Impact of Prejudice among Desegregated Whites." *The Journal of Politics* 57(2):402–424.
- Klopp, Jacqueline M. 2001. "'Ethnic Clashes' and Winning Elections: The Case of Kenya's Electoral Despotism." *Canadian Journal of African Studies* 35(3):473–518.
- Leo, Christopher. 1984. *Land and Class in Kenya*. Toronto: University of Toronto Press.
- Leys, Colin. 1974. *Underdevelopment in Kenya. The Political Economy of Neo-Colonialism, 1964-1971*. Berkeley: University of California Press.
- Lichbach, Mark. 1995. *The Rebel's Dilemma*. Ann Arbor: University of Michigan Press.
- Lynch, Gabrielle. 2011. *I say to you: Ethnic politics and the Kalenjin in Kenya*. Chicago: The University of Chicago Press.
- Massey, Douglas S. and Nancy Denton. 1993. *American Apartheid: Segregation and the Making of the Underclass*. Cambridge, Mass.: Harvard University Press.
- Massey, Garth, Randy Hodson and Dusko Sekulic. 1999. "Ethnic Enclaves and Intolerance: The Case of Yugoslavia." *Social Forces* 78(2):669–694.

- Matuszeski, Janina and Frank Schneider. 2006. "Patterns of Ethnic Group Segregation and Civil Conflict." *Working Paper, Harvard University*.
- Mbithi, Philip M. and Carolyn Barnes. 1975. *The Spontaneous Settlement Problem in Kenya*. Kampala: East African Literature Bureau.
- Médard, Claire. 1999. Territoires de l'éthnicité: Encadrement, revendications et conflits territoriaux au Kenya Phd Université de Paris I.
- Morgan, W. T. W. 1963. "The 'White Highlands' of Kenya." *The Geographical Journal* 129(2):140–155.
- NASA/University of Maryland. 2002. "Hotspot/Active Fire Detections. Data Set."
- Nunn, Nathan and Diego Puga. 2012. "Ruggedness: The blessing of bad geography in Africa." *Review of Economics and Statistics* 94(1):20–36.
- Odingo, Richard S. 1971. *The Kenya Highlands: Land Use and Agricultural Development*. Nairobi: East Africa Publishing House.
- Okoth-Ogendo, H. W. O. 1991. *Tenants of the Crown: Evolution of Agrarian Law and Institutions in Kenya*. Nairobi: African Centre for Technology Studies (ACTS) Press.
- Oliver, J. Eric and Janelle Wong. 2003. "Intergroup Prejudice in Multiethnic Settings." *American Journal of Political Science* 47(4):567–582.
- Olzak, Susan, Susanne Shananhan and Elizabeth McEneaney. 1996. "Poverty, Segregation, and Race Riots: 1960-1993." *American Sociological Review* 61(4):590–613.
- Osolo-Nasubo, Ngweno. 1977. *A socioeconomic study of the Kenya Highlands from 1900-1970 : a case study of the Uhuru Government*. Washington, DC: University Press of America.
- Ostby, Gudrun. 2008. "Polarization, Horizontal Inequalities and Violent Civil Conflict." *Journal of Peace Research* 45(2):143–162.
- Oucho, John O. 1988. Spatial Population Change in Kenya: A District-Level Analysis. In *Kenya's Population Growth and Development to the year 2000*, ed. Simeon H. Ominde. Nairobi: Heinemann Kenya.
- Oucho, John O. 2002. *Undercurrents of Ethnic Conflict in Kenya*. Leiden: Brill.
- Pettigrew, Thomas F. 1998. "Intergroup Contact Theory." *Annual Review of Psychology* 49:65–85.
- Posen, Barry R. 1993. "The Security Dilemma and Ethnic Conflict." *Survival* 35(1):27–47.
- Putnam, Robert D. 2007. "E Pluribus Unum: Diversity and Community in the Twenty-first Century The 2006 Johan Skytte Prize Lecture." *Scandinavian Political Studies* 30(2):137–174.

- Reardon, Sean F. and Glenn Firebaugh. 2002. "Measures of Multigroup Segregation." *Sociological Methodology* 32(33-67).
- Republic of Kenya. 1964a. *Kenya Population Census, 1962*. Vol. 3. African Population Nairobi: Economics and Statistics Division. Ministry of Finance and Planning.
- Republic of Kenya. 1964b. *Kenya Population Census, 1962. Advance Report of Volumes 1 & 2*. Nairobi: Ministry of Finance and Economic Planning.
- Robinson, Amanda. 2013a. "Ethnic Diversity, Segregation, and Ethnocentric Trust in Africa." *Working Paper*. Ohio State University. .
- Robinson, Amanda. 2013b. "Internal Borders: Ethnic Diversity and Market Segmentation in Malawi." *Working Paper*. Ohio State University. .
- Sambanis, Nicholas and Moses Shayo. 2013. "Social Identification and Ethnic Conflict." *American Political Science Review* 107(02):294–325.
- Scacco, Alexandra. 2012. Who Riots?: Explaining Individual Participation in Ethnic Violence PhD thesis.
- Schelling, Thomas C. 1971. "Dynamic Models of Segregation." *Journal of Mathematical Sociology* 1(2):143–186.
- Shirlow, Peter and Brendan Murtagh. 2006. *Belfast: Segregation, Violence, and City*. London: Pluto Press.
- Sorrensens, M. P. K. 1968. *Origins of European Settlement in Kenya*. Nairobi, Kenya: Oxford University Press.
- Steele, Abbey. 2009. "Seeking Safety: Avoiding Displacement and Choosing Destinations in Civil Wars." *Journal of Peace Research* 46(3):419–429.
- Toft, Monica. 2003. *The Geography of Ethnic Violence*. Princeton, NJ: Princeton University Press.
- UNOSAT. 2008. "Chronology of Fire Locations in Rift Valley Province Following Kenyan National Elections."
- USGS. 2006. "eMODIS TERRA Normalized Difference Vegetation Index (NDVI).".
- van Zwanenberg, Roger M. A. 1975. *An Economic History of Kenya and Uganda, 1800-1970*. London: Macmillan.
- Varshney, Ashutosh. 2002. *Ethnic Conflict and Civic Life: Hindus and Muslims in India*. New Haven: Yale University Press.
- Weidmann, Nils B. and Idean Saleyhan. 2011. "Violence and Ethnic Segregation: A Computational Model Applied to Baghdad." *Working Paper*. Yale University. .

Welch, Susan, Lee Sigelman, Timothy Bledsoe and Michael Combs. 2001. *Race and Place: Race Relations in an American City*. New York: Cambridge University Press.

Wilkinson, Steven I. 2004. *Votes and Violence: Electoral Competition and Ethnic Riots in India*. Cambridge: Cambridge University Press.

## 8 Figures and Tables

Figure 1: Ethnic Fractionalization and Segregation (Theil's Information Index)

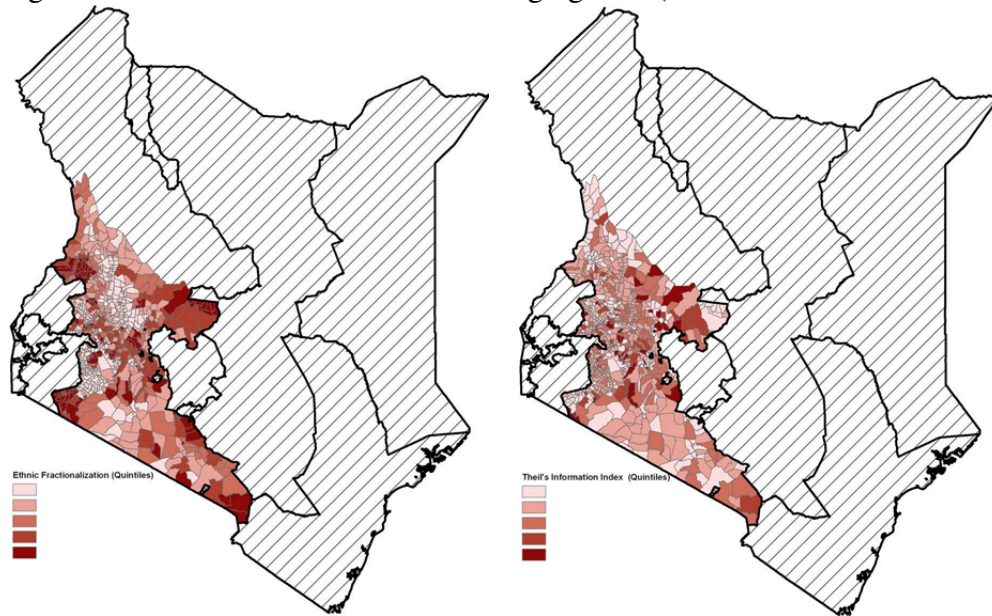




Figure 2: Colonial Cadastral Boundaries over Satellite Images



Table 1: Segregation and Violence

<i>Second Stage:</i>	<i>Log. No. of IDPs</i>			<i>Log. Fire Brightness</i>		
	[1] OLS	[2] RF	[3] IV	[4] OLS	[5] RF	[6] IV
Segregation (Theil's Index)	0.10*** ( 0.03)		1.49 *** ( 0.50)	0.07 ( 0.04)		0.67 ( 0.34)
Log. Population Density (1962)	0.09 ( 0.06)	0.04 ( 0.06)	0.04 ( 0.12)	-0.14** ( 0.07)	-0.17** ( 0.07)	-0.16 ( 0.08)
Log. Area	0.20*** ( 0.06)	0.16** ( 0.06)	-0.28 ( 0.20)	0.24*** ( 0.06)	0.24** ( 0.06)	0.04 ( 0.14)
Log. Rainfall	0.10** ( 0.05)	0.10 ( 0.05)	-0.17 ( 0.13)	0.20*** ( 0.05)	0.20*** ( 0.05)	0.08 ( 0.09)
Alienated Area (1934)	0.39*** ( 0.04)	0.15*** ( 0.06)	0.20** ( 0.10)	0.23 *** ( 0.04)	0.12** ( 0.06)	0.14** ( 0.07)
Prop. Rangeland	-0.09 ** ( 0.04)	-0.10*** ( 0.04)	-0.01 ( 0.07)	-0.01 ( 0.04)	-0.02 ( 0.04)	0.02 ( 0.05)
Log. Altitude	0.16*** ( 0.04)	0.15*** ( 0.04)	0.14** ( 0.07)	0.04 ( 0.04)	0.04 ( 0.04)	0.03 ( 0.05)
Log. Terrain Ruggedness	-0.08** ( 0.04)	-0.08** ( 0.03)	-0.03 ( 0.07)	-0.00 ( 0.04)	-0.00 ( 0.04)	0.02 ( 0.05)
CFDI		0.31*** ( 0.05)			0.14** ( 0.06)	
<b><i>First-Stage: Segregation</i></b>						
CFDI			0.21*** ( 0.06)			0.21*** ( 0.06)
Log. Population Density (1962)			-0.00 ( 0.07)			-0.00 ( 0.07)
Log. Area			0.30*** ( 0.07)			0.30*** ( 0.07)
Log. Rainfall			0.18*** ( 0.05)			0.18*** ( 0.05)
Alienated Area (1934)			-0.03 ( 0.07)			-0.03 ( 0.07)
Prop. Rangeland			-0.06 ( 0.04)			-0.06 ( 0.04)
Log. Altitude			0.01 ( 0.04)			0.01 ( 0.04)
Log. Terrain Ruggedness			-0.03 ( 0.04)			-0.03 ( 0.04)
N	717	717	717	717	717	717
F-stat Inst.			11.02			11.02

*Note:* Continuous variables are standardized to have a mean of 0 and a standard deviation of 0.5. Standard errors in parentheses. \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

Table 2: Accounting for Ethnic Diversity

	<i>Log. No. of IDPs</i>			<i>Log. Fire Brightness</i>		
	[1] OLS	[2] RF	[3] IV	[4] OLS	[5] RF	[6] IV
Segregation (MI Index)	0.16*** ( 0.03)		1.23 *** ( 0.33)	0.10** ( 0.04)		0.55** ( 0.26)
Log. Population Density (1962)	0.11 ( 0.06)	0.04 ( 0.06)	0.19** ( 0.10)	-0.13 ( 0.07)	-0.17 ( 0.07)	-0.10 ( 0.08)
Log. Area	0.18*** ( 0.06)	0.16** ( 0.06)	-0.12 ( 0.13)	0.24*** ( 0.06)	0.24*** ( 0.06)	0.11 ( 0.10)
Log. Rainfall	0.09 ( 0.05)	0.10 ( 0.05)	-0.14 ( 0.10)	0.19*** ( 0.05)	0.20*** ( 0.05)	0.10 ( 0.08)
Alienated Area (1934)	0.35*** ( 0.04)	0.15*** ( 0.06)	0.05 ( 0.11)	0.21 *** ( 0.04)	0.12 ( 0.06)	0.08 ( 0.09)
Prop. Rangeland	-0.09 ** ( 0.04)	-0.10 *** ( 0.04)	-0.05 ( 0.06)	-0.02 ( 0.04)	-0.02 ( 0.04)	0.00 ( 0.05)
Log. Altitude	0.16*** ( 0.04)	0.15*** ( 0.04)	0.16*** ( 0.06)	0.04 ( 0.04)	0.04 ( 0.04)	0.04 ( 0.05)
Log. Terrain Ruggedness	-0.07** ( 0.03)	-0.08 ( 0.03)	0.01 ( 0.06)	0.00 ( 0.04)	-0.00 ( 0.04)	0.04 ( 0.05)
CFDI		0.31*** ( 0.05)			0.14*** ( 0.06)	
<i>First-Stage: Segregation</i>						
CFDI			0.25*** ( 0.06)			0.25*** ( 0.06)
Log. Population Density (1962)			-0.13 ( 0.07)			-0.13 ( 0.07)
Log. Area			0.23*** ( 0.06)			0.23*** ( 0.06)
Log. Rainfall			0.19*** ( 0.05)			0.19*** ( 0.05)
Alienated Area (1934)			0.09 ( 0.06)			0.09 ( 0.06)
Prop. Rangeland			-0.04 ( 0.04)			-0.04 ( 0.04)
Log. Altitude			-0.00 ( 0.04)			-0.00 ( 0.04)
Log. Terrain Ruggedness			-0.07 ( 0.04)			-0.07 ( 0.04)
N	717	717	717	717	717	717
F-stat Inst.			17.79			17.79

*Note:* Continuous variables are standardized to have a mean of 0 and a standard deviation of 0.5. Standard errors in parentheses. \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

Table 3: Exploring Mechanisms

Second Stage:	Log. No. of IDPs				Log. Fire Brightness			
	Terrain Ruggedness		Distance to Road		Terrain Ruggedness		Distance to Road	
	< Median [1]	> Median [2]	< Median [3]	> Median [4]	< Median [5]	> Median [6]	< Median [7]	> Median [8]
Segregation (Theil's Index)	2.36** (1.11)	0.47 (0.38)	2.08 (1.11)	0.30 (0.32)	1.40 (0.74)	-0.25 (0.42)	1.06 (0.70)	-0.54 (0.41)
Log. Population Density (1962)	0.03 (0.30)	0.14** (0.08)	0.29 (0.28)	0.15 (0.09)	-0.27 (0.20)	-0.04 ( )	-0.17 (0.18)	0.13 (0.11)
Log. Area	-0.70 (0.48)	0.10 (0.13)	-0.37 (0.40)	0.06 (0.14)	-0.34 (0.32)	0.36*** (0.14)	-0.11 (0.25)	0.50*** (0.18)
Log. Rainfall	-0.38 (0.31)	0.05 (0.09)	-0.40 (0.35)	0.00 (0.08)	-0.06 (0.21)	0.25*** (0.10)	-0.04 (0.22)	0.24** (0.10)
Alienated Area (1934)	0.18 (0.18)	0.27*** (0.08)	0.24 (0.18)	0.22*** (0.06)	0.12 (0.12)	0.24*** (0.09)	0.18 (0.11)	0.17** (0.08)
Prop. Rangeland	0.10 (0.17)	-0.07 (0.07)	-0.01 (0.16)	-0.02 (0.05)	0.07 (0.12)	0.06 (0.07)	-0.01 (0.10)	0.02 (0.06)
Log. Altitude	-0.02 (0.21)	0.13** (0.05)	0.30 (0.17)	0.09** (0.04)	-0.03 (0.14)	-0.02 (0.06)	0.09 (0.11)	0.05 (0.05)
Log. Terrain Ruggedness	0.12 (0.48)	-0.09 (0.06)	-0.02 (0.16)	-0.02 (0.04)	0.20 (0.32)	-0.01 (0.07)	-0.01 (0.10)	0.04 (0.05)
N	358	359	357	360	358	359	357	360

Note: Continuous variables are standardized to have a mean of 0 and a standard deviation of 0.5. Standard errors in parentheses. \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

## A Data Appendix

### A.1 Estimates of Local Ethnic Composition

Given the unavailability of disaggregated census data on ethnic composition, I construct estimates of ethnic composition at the location level in 2006 by using the 2006 Voter Register and location-level data from the 1962 census.

To match names to groups, one would ideally calculate the probability that a person is a member of each ethnic group ( $g_i$ ) given their last name ( $P(g_i|Name)$ ). However, it is not possible to calculate this probability given the data. Instead, I calculate the probability that a person having that name falls into an ethnically homogenous administrative location in 1962 and then use these probabilities to match names to groups.

For each of the approximately 500,000 unique name strings in the register I calculate the probability that a person holding it is resident in a location ( $s_i$ ) where members of ethnic group  $g_i$  were a *supermajority* in 1962 ( $P(s_i|Name)$ ). This probability is calculated for each of the groups in the dataset and names were matched to groups where this probability is highest.

The probability that a person with some name is resident in an area  $s$  where group  $g$  has a supermajority is

$$P(s_i|Name) = \frac{n_s}{n}$$

where  $n$  is the number of registered voters with the last name and  $n_s$  is the number of registered voters with that name in area  $s$ .

Supermajority areas are defined using the 1962 census, which is the last period for which fine-grained data on ethnic composition is available. Because this ethnicity data was in tabular form, I constructed a map of local-level units in 1962 (Republic of Kenya 1964b).

I use a conservative threshold of 90% to define supermajority areas. There are 310 supermajor-

ity locations, comprising 73% of all locations. I matched names to these groups Embu, Kalenjin, Kikuyu, Kamba, Luhya, Luo, Maasai, Mbeere, Meru, Mijikenda, Orma, Pokomo, Taita, Teso, and Tharaka. Groups were matched to names with the highest value of  $P(s_i|Name)$  only if this probability was over three times larger than the probability for the group with the second highest probability to reduce the possibility of misclassifying ethnically ambiguous names.

## A.2 Polling Stations and Electoral Returns

Frequent changes in administrative jurisdictions uncoordinated across administrative agencies present a major challenge to measuring local-level electoral outcomes. I use polling stations to construct local-level aggregates because they are fixed points in space. I created a map of polling stations drawing on two sources. First, I acquired large scale paper maps (on a scale of 1:50,000 or larger) covering 175 local authorities from the Electoral Commission of Kenya in 2007. The Electoral Commission constructed paper maps for administrative puposes and, at that time, the Electoral Commission was a more credible source of data on both electoral and administrative boundaries than other government agencies. I georeferenced these maps and plotted polling stations from them. Because many polling stations are primary schools, I also used data from a survey of schools done by the Ministry of Education. The final dataset covers 97% of the 14,000 polling stations in existence in 2002 and 83% of the 21,000 polling stations in the 2006 Voter Register.

## A.3 Other Variables

- *Population Density (1962)* is the population density of a location's parent ward in 1962. If a 1999 location spans two or more 1962 locations, these values are area-weighted means.
- The proportion of a location's area covered by rangeland (*Proportion Rangeland*) was calculated using remotely sensed data on landcover (Department of Resource Surveys and Remote Sensing (DRSRS). Ministry of Environment and Natural Resources 2003).

- *Average Monthly Rainfall* in a location was estimated using monthly rainfall data for 48 rainfall stations across Kenya from 1988 to 2007, obtained from the Kenya Meteorological Department. Location-level rainfall estimates were generated in ArcGIS using spatial interpolation.
- *Altitude* is the elevation, in meters, of the center of the location (International Livestock Research Institute (ILRI) 2007).
- *Terrain Ruggedness* is a measure of the value of the “Terrain (or Topographical) Ruggedness Index” of points falling in a location. The terrain (or topographical) ruggedness index measures elevation change between a point on the grid and the eight points surrounding it (Nunn and Puga 2012).