"Stain Removal": Measuring the Effect of Violence on Local Ethnic Demography in Kenya*

J. Andrew Harris†

March 25, 2012

Abstract

In this paper, we examine patterns of violent ethnic targeting during Kenya's 2007-2008 post-election violence and estimate the impact of that targeting on local ethnic demography. Specifically, we focus on patterns of arson, one of the key types of violence used to displace certain ethnic sub-groups. We find that patterns of arson are related to the presence of ethnic outsiders in the region, and even more strongly related to measures of land quality, accessibility of targets, and local electoral competition. By comparing ethnic composition before and after the post-election violence, we show that arson caused a significant decrease in the number of Kikuyu registered to vote in the Rift Valley, and that other ethnic groups do not experience this decrease. This result supports narrative accounts of ethnic targeting during Kenya's 2007-2008 post-election violence, and provides systematic evidence of the effects of the post-election violence on local ethnic composition in Kenya. A placebo test confirms the robustness of our results.

*[*Kuondoa madoadoa*, translated as removing spots or stains, was a term used by locals to describe the process of chasing away non-indigenous settlers in Kenya's Rift Valley.]

†Ph.D. Candidate in Government, Harvard University. Email: jaharris@fas.harvard.edu. PLEASE DO NOT CITE OR CIRCULATE WITHOUT AUTHOR'S PERMISSION.

1 Introduction

How does politically-motivated violence affect the neighborhoods in which it occurs? While a large body of literature in political science examines ethnic violence, relatively little research attempts to document the occurence of such violence and its spatial extent and magnitude. In this paper, we demonstrate that a primary mode of violence – arson – reduced the population of ethnic Kikuyu residing in the study area by approximately 60% during Kenya's 2007-8 post-election violence. Arson did not induce similar effects among other ethnic groups. A placebo test shows that these results are unlikely do to chance, and additional analyses support the interpretation that this reduction was due to displacement.

This research makes several unique contributions to the study of ethnic violence and Kenya's recent post-election violence in particular. First, by focusing on how particular violent acts affect particular subpopulations in different ways, we show that violence in Kenya's Rift Valley focused narrowly on one ethnic group – the Kikuyu – and not on other (non-indigenous) ethnic groups. This result stands in sharp contrast to narrative accounts and legal theories suggesting the violence targeted several different ethnic groups in the study area. Second, we tailor our units of analysis and statistical methods to specific contextual details. In particular, we abandon the common practice in the conflict literature where administrative boundaries are used to define units of analysis. Finally, this paper recasts the analytical task of studying ethnic violence as the empirical task of uncovering ethnic targeting. The approach presented here enables the systematic documentation of the extent and magnitude of ethnic targeting, complementing qualitative work on the Kenyan conflict (Anderson and Lochery, 2008; Kanyinga, 2009).

In the next section, we briefly review the relevant literature. Section 3 contextualizes Kenya's 2007-8 post-election violence, describes the incentives underlying the organization and perpetration of the violence, and outlines hypotheses about where the violence occurred and which subpopulations it affected. Section 4 discusses the empirical strategy and data. Section 5 presents results and offers an interpretation. Section 6 concludes.

2 Literature Review

Several recent works attempt to understand the factors motivating violence following Kenya's 2007 General Elections. Anderson and Lochery (2008) and Kanyinga (2009) present historical arguments concerning the causes of the violence. They focus on how colonial land policies and post-colonial resettlement dynamics generated a volatile ethnic patchwork in Kenya's Rift Valley. Together with Lynch (2008), they emphasize the importance of long-standing grievances in generating both political support and an expectation for redress of those grievances through the political process. We discuss these issues in more detail in section 3.

Most relevant to our research is Kasara (2009), a detailed quantitative analysis on the electoral incentives for violence and displacement during that period. Kasara (2009) shows how the relationship between electoral competition in the 2002 parliamentary elections, on the one hand, and measures of violence (counts of fires in administrative areas) and displacement of individuals, on the other hand. Kasara (2009) finds no relationship between electoral competition and violence. However, Kasara (2009) does report a significant positive relationship between electoral competition and the displacement of individuals. Her results echo findings on election-related violence in Kenya's past (Klopp, 2001) as well as other areas (Wilkinson, 2004).

Our work provides several important improvements on Kasara (2009). First, we focus on local government electoral competition in 2007 – proximate to the violence in question – rather than parliamentary electoral competition in 2002. In order for electoral competition to be a motivation for violence, an electoral threat ought be credible and present. An examination of constituency-level parliamentary returns from both 2002 and 2007 finds little evidence of ethnically-aligned political competition in areas that purportedly experienced violence in 2007. An ethnic electoral threat does, however, appear to be substantial for local councillors, who are elected to represent the small wards that make up parliamentary constituencies. This makes sense: an ethnic enclave clustered in a single electoral ward has, by dint of geography alone, a better chance at winning a local council election than a parliamentary election. Thus, we expect the electoral incentives for violence to track local government council electoral competition, rather than parliamentary competition. KNCHR (2008) supports this interpretation, detailing the participation of current, former, and aspiring councillors in the perpetration of violence, providing no instances on actual perpetration of violence by members of parliament.

Second, we focus explicitly on a key motivation of perpetrators during the violence: displacement of non-indigenous from land in Kenya's Rift Valley. As noted above, longstanding grievances related to land were a key political issue motivating Kalenjin political unity. We examine whether patterns of violence follow various measures of land quality in the region of interest.

Third, our units of analysis and measures of violence and its consequences enable a more detailed analysis than previously possible. We examine how specific acts of violence – arson – change local ethnic demography between 2007 and 2010, while Kasara (2009) uses a cross-section measure of the home locations of a subset of displaced individuals to measure the impact of violence. Our difference-in-differences design provides plausibly causal estimates of the effect of arson on local ethnic demography.

Additionally, our measurements are taken and analyzed as geo-referenced polling stations and point-referenced fire locations. Kasara (2009) aggregates measures of violence and displacement to the location level (an administrative unit), which may be problematic if administrative boundaries themselves were drawn to respect local ethnic demography. In

Kenya, this is almost certainly the case (Oucho, 2002; Barkan, Densham and Rushton, 2006; Kasara, 2006). Using ethnically-determined boundaries to delimit units may be problematic, since those boundaries are likely determined by the same processes shaping incentives for violence. Analysis of aggregate data may also induce bias via the modifiable areal unit problem (MAUP) (Openshaw, 1984; Cressie, 1996; Dark and Bram, 2007). Our choice to use the polling station as the unit of analysis guards against MAUP-related bias. Finally, our approach employs a new way of estimating local ethnic composition that is approximately unbiased (Harris, 2011).

This project also represents a natural progression in the study of conflict. Foundational work on civil war and conflict examined variation at the country-level (Collier and Hoeffler, 2002; Fearon and Laitin, 2000). This was soon followed by work focusing on subnational units, studied using regression-based approaches developed in the cross-country literature (Gleditsch et al., 2002; Hegre and Raleigh, 2007). Concomitantly, country-specific research by Wilkinson (2004) and Varshney (2003) began unpacking violence at the subnational level, examining the determinants of violence within a country. Articles in Verwimp, Justino and Bruck (2009) and Cederman and Gleditsch (2009) can be seen to represent the state-of-the-art in the field, examining the local dynamics of violence in specific contexts.

This project makes several contributions to the literature on sub-national violence by improving how we measure violence and estimate its effects. Measuring the effects of violence is difficult for several reasons. First, systematic data collection before and after a conflict (e.g., demographic surveys or censuses) rarely coincides with systematic data on the occurence of violence. As a result, outcomes may be suggestive of conflict-induced change, but without an indicator of conflict related to the measurement, the causal effect of violence on an outcome of interest cannot be estimated.

When an indicator of violence does exist, it is often a function of the consequences of conflict itself (e.g., the number of casualties).[1] For example, self-reporting in survey-based research on the consequences of violence leaves both the determination of the presence or absence of violence, as well as claims about impact, in the hands of the respondent (Bellows and Miguel, 2009).

The use of news reports of conflict leads to a different set of issues surrounding the reported type of violence and the selection process leading to journalistic coverage of an event (Bocquier and Maupeu, 2005, on Kenya). And while *a priori* categorization of conflicts as "ethnic" or "religious" by journalists *may* suffice for understanding the causes of conflict, doing so when attempting to estimate the consequences of conflict simply begs the question: an ethnic conflict should differentially affect some group over other groups. These problems are perhaps most problematic when attempting to evaluate claims of ethnic targeting in local civil conflict, since specific and detailed data are required to identify ethnic sub-populations. Administrative records rarely provide adequate information on population characteristics related to the hypothesis of targeting or cleansing.

3 Background

In this section, we provide contextual background against which the violence plays out. These details highlight the incentives at play in both the organization and the perpetration of Kenya's 2007-2008 post-election violence. By outlining the factors motivating where violence occurred, we can more fully develop a theory of this particular violent episode. Such a theory, in turn, aids in implementing a statistical model of the spatial occurrence of violence.

Work by Kalyvas (2003) emphasizes the dual nature of civil conflict, in which a national "master" cleavage provides a window of opportunity in which to organize local grievances

[1] Indeed, recursive definition of conflict is a hallmark of cross-national work on conflict, where the casualty thresholds for whether or not to include a given episode of violence in a dataset are important (Gleditsch et al., 2002).

towards a larger political goal. During this period, local perpetrators can forcefully resolve these grievances to their material benefit, as well as to the political benefit of national organizers motivated by the master cleavange. This general schematic characterizes the nature of the Kenya's 2007-2008 post-election violence.

We focus on one region of Kenya for this study. The region – approximately bounded by 35.5' to 35.7' longitude and 0.1' to 1.4' latitude – includes the central and upper Rift Valley (along with the towns of Eldoret and Kitale) as well as portions of Western province. The region displays spatial variation in the occurence of violence, containing places experiencing intense violence as well as areas that contained less violence (KNCHR, 2008; ICC, 2010). This variation, along with our high resolution data, facilitates the empirical strategy. Limiting the region of the analysis also makes this project feasible. The models and data employed require significant computing time. Future work will generalize the approach to larger areas of Kenya. Figure (1) shows the extent of the study area, outline in black, with the colored background representing former colonial settlement areas.

3.1 Incentives for Violence

The perpetration of violence can hardly be considered costless. In addition to financing the logistical elements of violence (e.g., weapons and transport), incentives must be in place to motivate people to perpetrate the violence. In this case, colonial settlement patterns and independence-era resettlement policies provide a strong material incentive – land – to perpetrate violence.

Land issues and attendant violence in Kenya's Rift Valley are well-documented (Anderson and Lochery, 2008; Mueller, 2008; Kanyinga, 2009; Rutten and Owuor, 2009; Boone, 2011). The historical roots of these problems lie in the transfer of land from white colonial settlers to African farmers following Independence. In the study area, this transfer created tensions between the indigenous Kalenjin and immigrant Kikuyu. Members of the Kalenjin ethnic

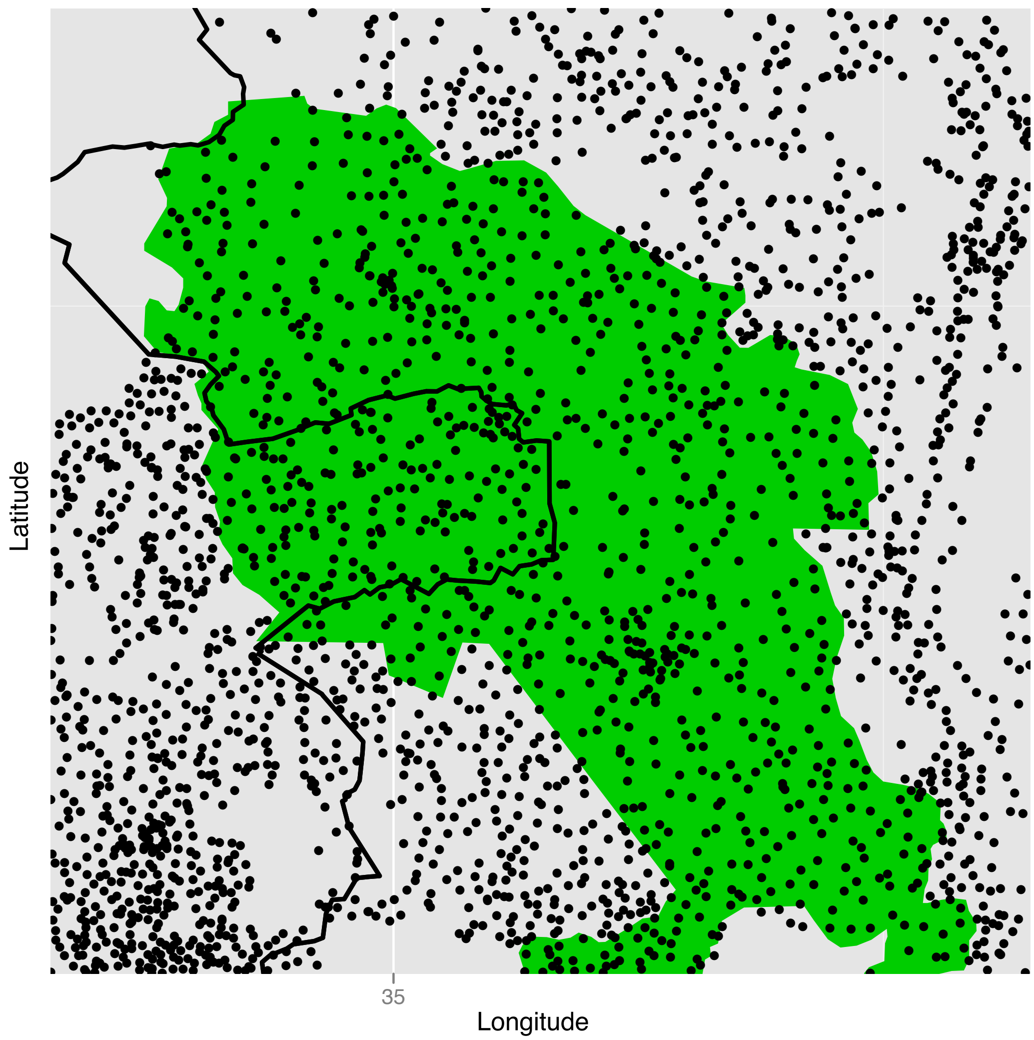

Figure 1: Map of Study Area: Encompassing several thousand square kilometers in Kenya's upper Rift Valley, the study area encompasses a region where much of the post-election violence occurred. Black dots represent the 2596 geo-referenced polling stations in the study area. Black line represents the boundary between Rift Valley and Western Provinces, as well as the boundary with Uganda (upper left corner). Green overlay represents the former colonial "White Highlands."

group feared that landless Kikuyu, many of whom had migrated to the region years before to labor on European farms, would secure title to land in the region. (Gertzel, 1970; Ogot, 1995). As settlement unfolded during the 1960s and 1970s, Kikuyu settlers did gain a foothold in the Rift Valley (Njonjo, 1977). Whether this was due to biased administrative practices favoring Kikuyu over Kalenjin or simply the result of better organization and access to credit on the part of the Kikuyu remains unresolved.[2] More importantly, the perception that Kikuyu "outsiders" illegitimately held Kalenjin land promoted Kalenjin political unity through narratives of land injustice (Lynch, 2008, 2011).

Evidence from pre-trial confirmation hearings at the International Criminal Court provide a window into the factors inducing Kalenjin in the area to participate in the violence. One witness stated that participants were promised a stipend for each Kikuyu killed, as well as a piece of land (ICC, 2012, p. 113). Another tendered a list of individuals that acquired land as a result of their participation in the violence (ICC, 2012, fn. 517 on p. 119). KNCHR (2008, p.79) suggests that this strategy of claiming *de facto* rights over land through force had been successful in the 1990s. This (ostensibly) credible promise of land and money in exchange for the perpetration of violence appears to have served as one incentive to participate in the violence. Kalenjin historical narratives surrounding the "occupation" of Kalenjin land by "outsiders" – Kikuyu, Kamba, Meru, and Kisii settlers – provided clues as to precisely which land could justifiably be targeted by perpetrators (ICC, 2012, p. 45).

The expulsion of ethnic "outsiders" not only opened land for later Kalenjin occupation, but also consolidated the Kalenjin vote in the region (ICC, 2012, p. 67). Given significant local concentration of Kikuyu votes which threatened (and in some wards in 2007, won) local elected councillors' seats, electoral consolidation probably served as a strong motivation for political leaders to encourage violence and drive political competitors from the region. This is especially true given the FPTP system, where a well-organized Kikuyu electorate, for

[2] Kanyinga (2009) suggests the former while Leys (1975) suggests the latter.

instance, could win a councillor seat. This political explanation is not new, as politicians have used ethno-political appeals to galvanize political support to motivate violence and maximize their electoral chances in prior elections (Klopp, 2001; Ajulu, 2002).

The 2007 presidential race also held particular salience with respect to land. The 2007 General Elections saw opposition leader Raila Odinga (an ethnic Luo) challenge incumbent Mwai Kibaki (an ethnic Kikuyu) for the presidency. A large segment of Odinga's electoral support came by way of William Ruto, a prominent Rift Valley politician. The Ruto-Odinga alliance found its roots in the 2005 constitutional referendum, which, among other things, focused on land reform, including redress for historical land injustices and land-related corruption (Lynch, 2006). The importance of redressing historical land injustices remained at the fore in the Rift Valley during the 2007 election season. Willis (2008) argues that Kenyans expected land reform to accompany an Odinga win: "For some, [an Odinga victory] was a threat to be avoided at all costs; for others, it was a prospect so alluring that disappointment could only lead to violence" (p. 270). As Lynch (2008, 561) argues, from the perspective of the average Kalenjin voter, "[t]he stakes invested in the [2007] presidential election were extremely high, change was in sight, and there was general anxiety in 'ODM zones' [Odinga's party] that change might be denied by the incumbent elite [i.e., the Kikuyu]."

Not only were the perceived stakes of winning high, but expectations about the probability of an Odinga victory – and significant land reform for Kalenjin voters – were high. Dissemination of opinion poll results leading up to the 2007 elections provided voters with unprecedented amounts of information about who the likely winner of the contest would be (Wolf, 2009). Kiage and Owino (2010) examines 51 pre-election polls in 2007 and find that, though the race between Kibaki and Odinga was close, the polls indicate a clear Odinga lead. Indeed, the results of the only exit poll completed echo pre-election opinion poll results (Gibson and Long, 2009).

After announcement of the final election results, widespread violence in the Rift Valley was reported. During the following two months, hundreds of thousands of people were displaced and over one thousand killed, and thousands of (purportedly Kikuyu) homes were burnt in the Rift Valley (Anderson and Lochery, 2008; Mueller, 2008; Waki, 2008). From the perspective of the aggrieved Kalenjin voter, one consequence of President Kibaki's "victory" was the continued delay of land reform in the Rift Valley. In this way, the violence that played out in the Rift Valley could be seen as an assertion of *de facto* property rights over land that had been "stolen" from the Kalenjin during post-Independence resettlement, given the continued intransigence of Kikuyu political elites.

ICC records also provide some insight into the logistical plans and challenges to organizing the perpetration of violence. ICC (2012, p. 71-72) reveals that youths perpetrating the violence were instructed "to converge on all trading centres to receive instructions" regarding the violence in the event of an Odinga loss. One criteria for targeted areas surfaced in a meeting that allegedly took place on December 6, 2011. During this meeting, which included organizers of the violence, "maps marking locations densely inhabited by members of Kikuyu, Kamba and Kisii communities were distributed" (ICC, 2012, p. 61). Regarding the scale of the attacks, fewer details exist. One witness describes "a tractor pulling a trailer carrying between 40 and 60 youths armed with arrows and machetes, material which was used to kill people". Another describes "more than 2000 physical perpetrators in the outskirts of Eldoret town before the attack" (ICC, 2012, p. 113).

This discussion provides some insight into the factors motivating the organization and perpetration of violence. In the next section, we outline an empirical strategy, and discuss data on violence, ethnicity, land quality, electoral competition, and target accessibility.

4 Empirical Strategy

In this section, we present our empirical strategy and data. First, we outline our approach to modeling the locations of arson. Then, we discuss estimation of the effect of arson on local ethnic demography at the polling station level. We also briefly describe independent variables used in each analysis data. We focus on a region spanning the border of Kenya's Rift Valley and Western provinces. This study area contained over 1.7 million (1.5 million) registered voters in 2007 (2010) and exhibited some of the most severe violence (KNCHR, 2008; ICC, 2010). Figure 1 above shows a map of the study area.

4.1 Modeling Locations of Arson

We use the presence or absence of fires during the period from December 30, 2007 to February 28, 2008 as indicators of arson during that period. We extracted the latitude and longitude of each detected fire during that period from the MODIS (Aqua and Terra) Thermal Anomalies Fire 5-Minute Level 2 Swath data.[3] The use of fires as a measure of violence is not new. The technique has been applied successfully in Sudan (Prins, 2008; Bromley, 2010). In the Kenyan case, the United Nations used satellite information on fires to locate areas where violence potentially occurred (Anderson and Lochery, 2008).

Kasara (2009) uses aggregated MODIS fire data as an outcome and models *counts* of fires at the location level. This approach has two problems. First, modeling counts of fires that fall within each administrative location aggregates important information on the location of the fire. If the area in question displays a high degree of local ethnic segregation, as shown in (Kasara, 2010), then aggregating fires reduces our ability to discern the relationship between the location of an ethnic group and the location of a fire. Aggregation makes a study of

[3]The MODIS L2 data downloaded from the NASA Land Processes Distributed Active Archive Center (LP DAAC), USGS Earth Resources Observation and Science (EROS) Center, Sioux Falls, South Dakota; `http://lpdaac.usgs.gov/get_data`.

ethnic targeting difficult, and could account for the null results on the relationship between political competition and fires in Kasara (2009). Second, Level 2 MODIS satellite data has a resolution of 1 kilometer. While a point location of a fire denotes the center of a 1 kilometer diameter area where the fire occurred, the actual location of the fire relative to an administrative boundary is uncertain. As a result, fires actually occurring in one location can easily be mis-counted as occurring in another location. As violence may have been most intense near boundaries (Anderson and Lochery, 2008), problems induced by issues of data resolution are especially acute.

We retain all information on where fires occurred by modeling the occurrence of the fire in continuous two dimensional space. Point process models, often used to model phenomena like disease outbreak or earthquake epicenters, provide an ideal vehicle for modeling the locations of fires (Gatrell et al., 1996; Diggle, 2003). Appendix **??** provides technical details. We use a poisson point process model to examine the factors that covary with fire occurrences during the 2007-2008 post-election violence. In effect, we are modeling fire *intensity*: the expected number of fires occurring at spatial location u, for all points in the study area.

4.2 Estimating the Effect of Arson on Ethnic Demography

We use a difference-in-differences (DID) approach to estimate the effect of individual incidents of arson – as detected from the satellite data described above – on ethnic demography at the local level in Kenya. DID compares changes in a treated group across time with changes in a similar control group. The standard DID approach compares one treated group to one control group, assuming that time trends unrelated to treatment and other differences across groups are identical. Here, a treated unit is the number of an ethnic sub-population residing in a local neighborhood that experienced violence. This sub-population is compared to changes in the same ethnic sub-population in an area that did not experience violence. We also include a triple-differenced outcome measure, effectively adding a second control group. This second control group incorporates information on changes in the numbers of the indigenous Kalenjin ethnic group. The rationale for this second control group is that, as perpetrators, changes in the size of the local Kalenjin population would not be due to violence. This approach is not unlike the spatial estimator proposed in McIntosh (2008).

Two kinds of information are required to implement this strategy: information on changes in local ethnic demography before and after the violence and information the occurrence of violence. Each type of information must be spatially referenced, in order to determine which local areas experienced violence and which did not. We proceed by defining and justifying the unit of interest and outcome variable. Then, we discuss the treatment indicator and covariate data.

Our outcome of interest is the change in ethnic demography at the local level. Unfortunately, comprehensive information on local ethnic demography recorded just before and just after the violence is not available. Census data at the enumeration area is not available. Thus, we generated new measures of ethnic demography at the local level using administrative records.

To measure local ethnic demography, we used voter registration records from 2007 and 2010. Voter register audits undertaken by the Institute for Education in Democracy in 2002 and 2007 found that 94% (of 1200 respondents in 2002) and 95% (of 2352 respondents in 2007) registered at the polling station closest to their residence (IED, 2002, 2007). Moreover, IED (2007) reports that over half of registered voters live less than one kilometer from the polling station where they are registered. Thus, ethnic composition measured at the polling station provides a suitable measure of the quantity of interest. Since the registers contains no direct information on ethnicity, we developed a new method for extracting information about ethnicity from person-names (Harris, 2011). This allows us to use person-names in the voter register as signals of each voter's ethnic identity, and combine those signals in a way that provides approximately unbiased estimates of the ethnic composition in each polling station.

For each polling station in the study area, we estimated proportions for the 9 largest ethnic groups in the region, 93% of the population, according to 1999 census data. To estimate the total number of voters of ethnic group j registered at polling station i, we count the number of registered voters (RV_i) at polling station i and then multiply this number by the estimates of the ethnic composition at polling station i (p_{ij}). This gives an estimate of the number of ethnic group j at polling station i (n_{ij}). We calculate this quantity using data from both the 2010 and 2007 voter registers, taking the difference of the logarithm of each estimate. Differencing the outcome in this way avoids serial correlation issues (Bertrand, Duflo and Mullainathan, 2004).

These estimates are then linked to a specific local neighborhood in Kenya using an original dataset of geo-referenced polling stations. This original dataset was assembled from several Kenyan government sources, as well as publicly available GIS data. Approximately 85% of the 20000+ polling stations were geo-referenced schools, which are commonly used as polling stations (Ltd., 2008). 11% were extracted from a comprehensive dataset of market towns and open source point data (Roy Jorgensen Associates, 2006; Map, N.d.; Geonames, N.d.). The remaining 4% of polling stations, for which we could not find suitable point data, were imputed randomly in the electoral ward in which the polling station is located.

Next, we proceeded to determine which polling stations were affected by arson, and which were not. The geo-referenced fire data discussed above provided a ready "treatment indicator" of the occurrence and location of arson; 150 such fires occurred in the study area during the period in question. To estimate the treatment effects, we must decide which polling stations received "treatment" – arson – and which did not. Generally, we could define treatment as any reasonable function of spatial proximity to fire. Given that dichotomous treatments are intuitively attractive and provide a relatively simple inferential target, we choose to model fires as a discrete treatment, defined by a given radius around a polling station.

The radius surrounding a polling station represents the "local neighborhood" or "catchment area" of the polling station. "Treatment" is defined a fire falling within this neighborhood. We tried two different approaches for defining this local neighborhood. First, we tried a fixed radius treatment rule: if a fire falls within two miles around a polling station, then that polling station is considered to have experienced violence. The left panel of figure (2) gives an example of the fixed radius treatment rule. While simple, the fixed-radius method has several undesirable features. In areas where polling stations are tightly clustered, a fixed radius can designate the entire cluster as treated, while in reality, the violence may have been limited to a smaller neighborhood. In sparsely populated areas, with few polling stations, the households registered at a given polling station may come from a much further distance than that defined by a fixed-radius catchment area.

Given these problems, we also implemented a variable radius catchment area defined as the distance between a polling station and its nearest neighbor plus a 500 meter buffer to account for measurement error in fire locations. An example of this approach can be seen in the right panel of figure (2) This approach more accurately captures the clustered nature of human settlements and provides a higher degree of resolution in terms of understanding fires as an activity targeted at a specific ethnic group. If polling stations are clustered together, then radii are small. In sparsely populated areas, where polling stations are further apart, the catchment area becomes larger.[4]

The fire data do present several problems. First, each of the two MODIS satellites passes over the study area twice daily, capturing live fires at that point in time. Fires that occur when a satellite is not overhead are not reported.[5] Second, the algorithms developed

[4]Of course, there are innumerable ways to implement the relationship between fires and polling stations. In future work, we would like to explore catchment areas based on network features (like roads), tessellations of 2-D space, and anisotrophic radii that take population density into account when defining a radius.

[5]That said, future iterations of this research will incorporate data on *burned areas*, which are derived from the same satellite data as the fire data. These may provide more complete coverage of likely arson

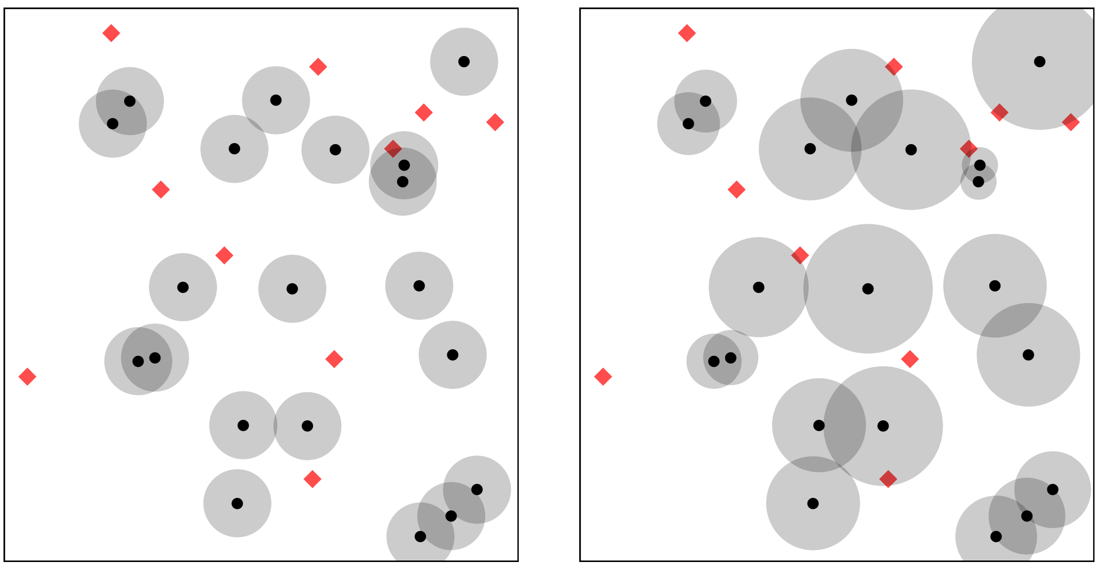

Figure 2: Treatment rules: We tried two different rules for determining the treatment status of polling stations. Under the fixed radius treatment rule (left), a polling station is treated if a fire occurs within a two mile radius. Under the variable radius treatment rule (right), a polling station is treated if a fire falls within a radius determined by a function of the distance between a polling station and its nearest neighbor. Thus, polling stations in urban areas tend to have smaller radii, and polling stations in rural areas tend to have larger radii. Results presented were estimated with the variable radius approach.

to identify fires from the satellite data often miss smaller fires (Wang et al., 2007; Shixing et al., 2010). These two problems imply that incidents of arson will be *under-reported* relative to their true occurrence. The third problem involves the nature of fires that are detected. MODIS data provides no information regarding the etiology of the observed fire, only its location and intensity. As a result, we cannot effectively differentiate between fires related to the violence and fires due to other causes.

How can we expect these problems to affect our estimates? Underreporting of fires will effectively group some "treated" units – neighborhoods actually affected by violence – as "control" units. Categorizing "normal" fires as arson will cause some control units to be grouped with the treated units. Both of these problems will attenuate estimates of the effect of arson, since, intuitively, the average change in the treated (control) group will decrease (increase) due to the measurement error.

In addition to our outcome and treatment variables, we control for number of independent variables that may affect assignment to treatment. Note that, when included in models, independent variables are standardized to have mean zero and standard deviation equal to 1. Summary statistics for these variables on their natural scales are in table **??**.

- Elevation: Land at higher elevation is valued due to more predictable precipitation, which facilitates the production of cash crops. We include 1-kilometer resolution data on average elevation (in meters). The author aggregated and resampled 90-meter resolution raster data from NASA's Shuttle Radar Topographic Mission, supplied as a digital elevation model by CGIAR.[6] Since desirable agricultural land often falls within an ideal band of elevation, we include elevation as a second-degree polynomial.

- Normalized Difference Vegetation Index (NDVI): This variable provides a rough mea-

locations.

[6]See `http://srtm.csi.cgiar.org/`.

sure of land fertility, capturing variation in local vegetation at 1 kilometer resolution.[7]

- Terrain: Flatter land is often better agricultural land. As a result, we might think that, if attacks were perpetrated with the intent to chase inhabitants away from their agricultural holdings, such attacks would be more likely on flatter land. We include a measure of terrain roughness derived from the SRTM digital elevation model described above.

- Roads: Since locations closer to roads may be easier to attack than more distant areas, we include a measure of the distance from a polling station to the nearest road. We use a GIS dataset collected by Kenya's Ministry of Roads and Public Works, measuring the linear distance between a polling station and the nearest road.

- Markets: Given that market centers provide one focal point at which perpetrators can coordinate for violence, we include a measure of the linear distance of a polling station to the nearest market center. This data was taken from a GIS dataset collected by Kenya's Ministry of Roads and Public Works.

- Population density: Cluster of population may affect the ability of perpetrators to easily target other groups. We include a measure of population density taken from the Afripop Project (Tatem et al., 2007).[8]

- % Outsiders: This measure is the percentage of outsider ethnic voters at polling station i in 2007.

- Victory Margin: This captures the victory margin for the local councillor in each electoral ward in the study area during the 2007 elections.

[7] Product MOD13, `http://modis.gsfc.nasa.gov/`.

[8] `http://www.afripop.org/`

5 Results

Table 1 summarizes the total number of registered voters by ethnic group in the 2596 polling stations in the study area. Overall, the population of registered voters declined by 12.5% between 2007 and 2010. This overall decline is likely due to the compilation of a new voter register in 2010. A more detailed analysis of differences between the 2007 and 2010 registers suggests that the 2007 register contained records for many voters who had died since registration. The 2010 registration process effectively purged these voters from the new register. We found no preliminary evidence that the distribution of likely dead voters varied between ethnic groups.

In terms of change across time, table 1 shows larger decreases for Kikuyu and Meru-Kamba ethnic groups, along with a moderate decrease for the Luhya.[9] The changes for the Kikuyu and Meru-Kamba groups align with expectations. However, the sizable decrease in Luhya and Luo (and the lack of a large decrease in Kisii) are puzzling. Closer inspection of table 1 shows a large number of Luo voters, while the study area itself bridges Western Province (which is predominately Luhya) and Rift Valley Province (which is mainly Kalenjin). Closer inspection of person names in the voter register suggests one possible source of this error: much overlap exists between Luhya, Luo, and Kisii names. Moreover, names identified as Luo in the training data used to estimate the ethnic proportions were lower variance than Kisii or Luhya names. As a result, the method over-estimated the proportion of Luo in the region. While these problems do not appear to affect estimates for the Kikuyu, Kalenjin, or Meru-Kamba groups, estimates presented below for the Kisii, Luo, and Luhya groups should be treated with caution. For this reason, discussion below will focus on Kikuyu, Kalenjin, and Meru-Kamba groups.

[9] Relatively few Meru lived in the study area, so they were combined with the Kamba for estimation.

	KIKUYU	LUHYA	KALENJIN	LUO	MERUKAMBA	KISII
2007	98878	912998	654003	36048	483	5300
2010	74911	772959	610517	29857	298	4799
% Change	-24	-15	-7	-17	-38	-9

Table 1: Estimated Voters by Ethnic Group, 2007 and 2010:.

5.1 Locations of Arson

Table 2 presents coefficient estimates from the point process analysis of fire locations in the study area. Coefficients β_i can be interpreted directly as contributions to fire intensity[10] when transformed $\exp \beta_i$. Coefficients remain untransformed in table 2. All variables have been standardized to mean zero and standard deviation of one. Models (1) and (2) contain basic controls for the percentage of outsiders and population density. Model (3) adds variables related to land quality – NDVI, elevation, and terrain. Model (4) adds variables related to the accessibility – distance from markets and major roads. Model (5) includes a variable for electoral competition at the ward level for 2007. Model (6) interacts the electoral competition variable with the percentage outsider variable. Model (7) mirrors model (6), except for the addition of an offset variable that controls for past fire activity.[11]

The first notable result, present in all models, is a positive and significant relationship between the percentage of outsiders (that is, Kikuyu, Kisii, or Meru-Kamba) present in an area and the intensity of fire activity. This is consistent with a hypothesis that acts of arson targeted non-indigenous ethnic groups. This phenomenon does not appear related to population density, suggesting little difference between urban and rural areas.[12]

The coefficients for variables related to land quality also accord with expectations. Denser

[10]That is, expected counts within a small 2-D area.

[11]This offset was generated by estimating the intensity of fires during the same period from 2002 to early 2007.

[12]The coefficient for population density in Model (7) indicates a positive relationship between fire intensity and population density. However, a common model selection statistic (AIC) suggests that the offset term included in model (7) leads to severe over-fitting.

	Models						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Variable:							
Intercept	101.84***	101.86***	71.64***	-0.89	-0.88	-0.89	-1***
	(54.04)	(54.05)	(24.07)	(-1.6)	(-1.53)	(-1.6)	(-4.65)
Outsiders	0.33***	0.33***	0.21***	0.17***	0.15***	0.19***	0.29***
	(8.26)	(8.33)	(5.16)	(4.1)	(3.59)	(4.08)	(5.44)
Population Density		-0.07	-0.02	-0.02	-0.03	-0.03	0.12**
		(-0.66)	(-0.17)	(-0.23)	(-0.41)	(-0.37)	(2.18)
NDVI			0.44**	0.67***	0.54**	0.55**	0.29
			(2.12)	(2.94)	(2.47)	(2.48)	(1.16)
Elevation			2.6***	1.72***	2.26***	2.31***	1.67***
			(5.58)	(4.42)	(4.89)	(4.94)	(3.44)
Elevation^2			-0.52***	-0.45***	-0.47***	-0.47***	-0.36**
			(-3.92)	(-3.38)	(-3.51)	(-3.48)	(-2.05)
Terrain			-0.72***	-0.63***	-0.63***	-0.63***	-0.29
			(-5.44)	(-4.49)	(-4.39)	(-4.41)	(-1.59)
Distance to Market				1.45***	1.31***	1.28***	0.16
				(6.99)	(6.58)	(6.4)	(0.9)
Distance to Road				-1***	-1***	-1***	-0.99***
				(-4.86)	(-4.72)	(-4.76)	(-3.09)
Victory Margin					-0.38***	-0.41***	-0.36**
(Local Gov. Election)					(-2.91)	(-3.01)	(-2.48)
Outsiders × Victory Margin						0.1	0.22**
						(1.43)	(2.52)

Note: $t-$statistics in parentheses. *: $p < 0.1$;**: $p < 0.05$;***: $p < 0.01$.

Table 2: Results – Point Process Model of Fire Locations: Outcome variable is the location of fires in the study area.

green vegetation, measured using NDVI, is positively related to fire incidence. The polynomial elevation corresponds with the notion that an elevation "sweet spot" exists and is positively correlated with fire intensity. The negative terrain coefficient indicates that flatter land (represented when the covariate is negative) leads to higher fire intensity. In summary, flatter, greener, ideally elevated land saw more fires than rougher, less green, and especially low or high land did.

Accessibility to an area also relates to fire activity, though the results are, at first glance, somewhat contradictory. Fire intensity increases with distance to the nearest market center, but decreases with distance to the nearest road. However, the estimated coefficients are consistent with existing arguments regarding the targeting of farms in the region. Settlement farms worked by ethnic outsiders tend to somewhat remote from market centers, though relatively well-served by the road network. If settlement farms were the primary targets of arson, as argued by Anderson and Lochery (2008), Kanyinga (2009) and Boone (2011), the estimated coefficients reflect the spatial reality of the target areas.

Models (5) and (6) in table 2 include the victory margin for local government councillor elections in 2007. The negative coefficient suggests that as a victory margin widens, fire intensity lessens. Put another way, more competitive areas tend to see higher fire activity.

Since all variables are standardized, direct comparisons of the magnitude of the coefficients is possible. One surprising result is that contribution of the percentage of outsiders to fire intensity is consistently smaller than variables related to land, accessibility, and political competition. On the one hand, this may be the result of the ethnicity-related measurement error discussed above. On the other hand, it may suggest that perpetrators' incentives for violence were more closely tied how easy it was to access their targets and the quality of targets' agricultural land.

5.2 Effect of Arson on Ethnic Subpopulations

In this part, we present results from DID and triple difference regressions. Regressions were specified for each ethnic group. Thus, there are six estimates for the DID results, and five for the triple difference results.[13] The tables below present the coefficient of the treatment indicator for each ethnic group. Each row represents one ethnic group.

Columns represent the control variables included in the regression equation along with the treatment indicator. Model (1) in tables 3 to 8 contains only the treatment indicator. Covariates in models (2) to (7) mirror the covariates included in models (1) to (6) of table

[13]Since the Kalenjin were used as the second control group in the triple differences, there is no estimate for that sub-group.

2, except that only the main term of elevation is included in the regressions below. The coefficients can be interpreted as the average percentage change for each ethnic group between 2007 and 2010.

Finally, we include results estimated using three different ways of weighting the data. The first approach weights all polling stations in the study area equally. These results are found in tables 3 and 6. Since many polling stations contain no outsiders in either year, tables 4 and 7 limit analysis only to those polling stations containing the relevant ethnic group in 2007 or 2010. Finally, within these subset, wide variation in the absolute number of each ethnic group at a polling station exists. Since one aim of the perpetrators was to target concentrations of outsiders, we weight by the total number of the relevant ethnic group in tables 5 and 8 to examine changes in ethnic sub-populations across time.

Across all tables, we find striking consistency in the negative and (usually) significant coefficient on the treatment indicator for the Kikuyu ethnic group. This result accords with narrative and legal evidence asserting that perpetrators targeted ethnic Kikuyus during the post-election violence. The effect increases in magnitude when analysis is limited to polling stations where Kikuyu are present (row 1, table 4) and even more-so when observations are weighted by the number of Kikuyu present in 2007 (row 1, table 5).

Estimates using the triple differenced outcomes appear largely similar for the Kikuyu. Unexpectedly, while the magnitude of coefficients for the Kikuyu in models (5-7) of table 7 remain unchanged, triple differencing actually increased the variance of estimates. This may reflect population redistribution between ethnic groups, as Kikuyu flee to IDP camps and urban areas and Kikuyu retain their ground or move into areas previously settled by the Kikuyu.[14]

Inference for the Meru-Kamba group provide some support that arson led to a decrease in that sub-population, though models (4-7) in table 7 show a positive, though insignificant,

[14]Robustness of the Kikuyu result is examined in a placebo test contained in Appendix 6.1

effect. We leave discussion of the estimates for the Kisii, Luhya and Luo groups to future worok, given the measurement issues discusses above.

One final observation from the tables lies in the estimates for the Kalenjin, the group to which most perpetrators purportedly belonged. Of the 21 treatment estimates related to the Kalenjin in tables 3 to 5, only one was significant, showing that, on average, treated polling stations experienced a 5% increase in Kalenjin registrants relative to control polling stations. Other coefficients remain small and insignificant. This result is consistent with the idea that Kalenjin were relatively unaffected by the violence, and may have moved into areas previously Kikuyu-occupied areas.

	1	2	3	4	5	6	7
KIKUYU	-0.3***	-0.22***	-0.22***	-0.2***	-0.19***	-0.19***	-0.19***
	(-6.48)	(-4.6)	(-4.62)	(-4.12)	(-3.86)	(-3.79)	(-3.81)
LUHYA	0	-0.03	-0.03	-0.02	-0.02	-0.01	-0.01
	(-0.08)	(-0.66)	(-0.73)	(-0.44)	(-0.34)	(-0.3)	(-0.29)
LUO	0.02	0.02	0.02	0.03	0.04	0.04	0.04
	(0.51)	(0.57)	(0.63)	(0.69)	(1.03)	(1.01)	(1.04)
KALENJIN	0.02	-0.01	0	-0.01	-0.03	-0.03	-0.03
	(0.72)	(-0.37)	(-0.09)	(-0.24)	(-0.86)	(-0.9)	(-0.89)
MERUKAMBA	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01
	(-0.57)	(-0.44)	(-0.55)	(-0.62)	(-0.59)	(-0.6)	(-0.58)
KISII	-0.02	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01
	(-0.55)	(-0.17)	(-0.14)	(-0.18)	(-0.22)	(-0.15)	(-0.16)

Note: $t-$statistics in parentheses. *: $p < 0.1$;**: $p < 0.05$;***: $p < 0.01$.

Table 3: DID Results – All Polling Stations..

	1	2	3	4	5	6	7
KIKUYU	-0.38**	-0.36**	-0.36**	-0.31*	-0.29*	-0.29	-0.29*
(n = 468)	(-2.5)	(-2.31)	(-2.28)	(-1.83)	(-1.69)	(-1.66)	(-1.67)
LUHYA	-0.02	-0.04	-0.04	-0.01	0	0	0
(n = 1939)	(-0.24)	(-0.67)	(-0.69)	(-0.17)	(0.02)	(0.05)	(0.04)
LUO	0.1	0.09	0.1	0.13	0.17	0.17	0.17
(n = 557)	(0.62)	(0.55)	(0.6)	(0.74)	(0.94)	(0.92)	(0.91)
KALENJIN	0.03	-0.02	-0.01	-0.01	-0.03	-0.03	-0.03
(n = 1642)	(0.79)	(-0.42)	(-0.31)	(-0.31)	(-0.84)	(-0.83)	(-0.87)
MERUKAMBA	-0.33	-0.27	-0.22	-0.24	-0.13	-0.16	-0.17
(n = 298)	(-1.64)	(-1.29)	(-1)	(-1.08)	(-0.56)	(-0.66)	(-0.71)
KISII	-0.12	-0.04	-0.02	-0.01	0	0	0
(n = 655)	(-0.82)	(-0.26)	(-0.1)	(-0.09)	(0.02)	(0.03)	(0.03)

Note: $t-$statistics in parentheses. *: $p < 0.1$;**: $p < 0.05$;***: $p < 0.01$.

Table 4: DID Results – Polling Stations Subset by Ethnicity.

	1	2	3	4	5	6	7
KIKUYU	-0.79***	-0.78***	-0.77***	-0.74***	-0.71***	-0.71***	-0.72***
	(-7.74)	(-7.99)	(-7.53)	(-6.45)	(-5.92)	(-5.94)	(-6.01)
LUHYA	0.05	0.05	0.05	0.14***	0.12**	0.12**	0.12**
	(1.24)	(1.22)	(1.17)	(2.99)	(2.38)	(2.39)	(2.38)
LUO	0.22*	0.24*	0.22*	0.35**	0.41**	0.45***	0.38**
	(1.84)	(1.98)	(1.82)	(2.27)	(2.58)	(2.72)	(2.3)
KALENJIN	0.05**	0.03	0.03	0.03	0.02	0.02	0.02
	(2.42)	(1.39)	(1.31)	(1.23)	(0.73)	(0.78)	(0.79)
MERUKAMBA	-0.95***	-0.82***	-0.67***	-0.6**	-0.56**	-0.61***	-0.51**
	(-18.98)	(-6.51)	(-3.25)	(-2.64)	(-2.47)	(-2.84)	(-2.14)
KISII	0.17*	-0.3***	-0.31***	-0.2**	-0.24**	-0.26***	-0.26***
	(1.96)	(-3.5)	(-3.66)	(-2.17)	(-2.6)	(-2.88)	(-2.9)

Note: $t-$statistics in parentheses. *: $p < 0.1$;**: $p < 0.05$;***: $p < 0.01$.

Table 5: DID Results – Polling Stations Weighted by Ethnic Group in 2007.

	1	2	3	4	5	6	7
KIKUYU	-0.31***	-0.21***	-0.22***	-0.2***	-0.17***	-0.17***	-0.17***
	(-5.75)	(-3.69)	(-3.85)	(-3.35)	(-2.84)	(-2.77)	(-2.78)
LUHYA	-0.03	-0.02	-0.03	-0.01	0.01	0.01	0.01
	(-0.5)	(-0.36)	(-0.59)	(-0.25)	(0.2)	(0.26)	(0.26)
LUO	0	0.03	0.03	0.03	0.07	0.07	0.07
	(-0.07)	(0.66)	(0.53)	(0.67)	(1.31)	(1.32)	(1.34)
MERUKAMBA	-0.03	0	-0.01	-0.01	0.01	0.02	0.02
	(-0.91)	(0.05)	(-0.24)	(-0.16)	(0.37)	(0.39)	(0.39)
KISII	-0.04	0	0	0	0.02	0.02	0.02
	(-0.87)	(0.1)	(-0.05)	(0.01)	(0.37)	(0.44)	(0.43)

Note: $t-$statistics in parentheses. *: $p < 0.1$;**: $p < 0.05$;***: $p < 0.01$.

Table 6: 3DID Results – All Polling Stations.

	1	2	3	4	5	6	7
KIKUYU	-0.43**	-0.4**	-0.41**	-0.33*	-0.29	-0.29	-0.29
	(-2.61)	(-2.31)	(-2.36)	(-1.8)	(-1.49)	(-1.47)	(-1.5)
LUHYA	-0.03	-0.03	-0.04	0	0.06	0.06	0.06
	(-0.47)	(-0.44)	(-0.53)	(0.04)	(0.78)	(0.81)	(0.8)
LUO	-0.06	0.05	0.05	0.13	0.2	0.2	0.2
	(-0.36)	(0.25)	(0.26)	(0.64)	(0.97)	(0.96)	(0.96)
MERUKAMBA	-0.15	-0.06	0	0.05	0.23	0.22	0.19
	(-0.64)	(-0.24)	(0.01)	(0.17)	(0.76)	(0.72)	(0.64)
KISII	-0.02	0.15	0.16	0.18	0.23	0.23	0.24
	(-0.11)	(0.79)	(0.82)	(0.88)	(1.11)	(1.1)	(1.15)

Note: $t-$statistics in parentheses. *: $p < 0.1$;**: $p < 0.05$;***: $p < 0.01$.

Table 7: 3DID Results – Polling Stations Subset by Ethnic Group.

	1	2	3	4	5	6	7
KIKUYU	-0.8***	-0.8***	-0.83***	-0.81***	-0.81***	-0.81***	-0.82***
	(-5.88)	(-5.87)	(-6.44)	(-5.9)	(-5.8)	(-5.84)	(-6.01)
LUHYA	-0.04	0.03	0.03	0.15*	0.21**	0.21**	0.21**
	(-0.51)	(0.49)	(0.45)	(1.94)	(2.57)	(2.54)	(2.59)
LUO	0.94***	0.85***	0.73***	3.79***	4.06***	2.98***	2.73***
	(3.67)	(3.41)	(2.96)	(7.69)	(7.85)	(6.64)	(6.11)
MERUKAMBA	-0.94***	-0.64***	-0.39	-0.23	-0.17	-0.22	-0.14
	(-17.4)	(-3.99)	(-1.52)	(-0.78)	(-0.58)	(-0.77)	(-0.48)
KISII	0.12	-0.32***	-0.32***	-0.19	-0.18	-0.19	-0.19
	(1.13)	(-3)	(-3.03)	(-1.66)	(-1.56)	(-1.58)	(-1.55)

Note: $t-$statistics in parentheses. *: $p < 0.1$;**: $p < 0.05$;***: $p < 0.01$.

Table 8: 3DID Results – Polling Stations Weighted by Ethnic Group.

6 Conclusion

In this paper, we provide the first systematic quantitative evidence of ethnic targeting and the effects of that targeting on ethnic sub-populations following Kenya's 2007 General Elections. The results show that arson – a central method of violence during the episode – was driven by land quality and accessibility of targeted groups in addition to ethnicity. Moreover, fire intensity increases with electoral competition in the 2007 local government elections. This new finding corroborates narrative and legal evidence that local political competition – rather than broader parliamentary or presidential competition – was a key incentive for organizing and perpetrating violence. The results presented suggest that arson caused a significant decrease in the number of ethnic Kikuyu registering to vote, a decrease we believe is consistent with the physical displacement of Kikuyu voters. The methods we present in this paper can be implemented in any context where name information is available before and after the conflict and can be used as a proxy for ethnic identity or religious affiliation. We also demonstrate the effectiveness of using fires as a proxy for violence, which may be appropriate in other settings where arson is used to displace certain populations.

(TBC.)

References

Abadie, Alberto, Alexis Diamond and Jens Hainmueller. 2010. "Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program." *Journal of the American Statistical Association* 105(490):493–505.

Ajulu, R. 2002. "Politicised Ethnicity, Competitive Politics and Conflict in Kenya: A Historical Perspective." *African Studies* 61(2):251–268.

Anderson, David 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.

Barkan, Joel, Paul Densham and Gerard Rushton. 2006. "Space Matters: Designing Better Electoral Systems for Emerging Democracies." *American Journal of Political Science* 50(4):926–939.

Bellows, John and Edward Miguel. 2009. "War and Local Collective Action in Sierra Leone." *Journal of Public Economics* 93:1144–1157.

Bertrand, Marianne, Esther Duflo and Sendhil Mullainathan. 2004. "How Much Should We Trust Differences-in-Differences Estimates." *Quarterly Journal of Economics* 119(1):249–275.

Bocquier, Philippe and Herve Maupeu. 2005. "Analysing Low Intensity Conflict in Africa Using Press Reports." *European Journal of Population* 21(2-3):321–345.

Boone, Catherine. 2011. "Politically Allocated Land Rights and the Geography of Electoral Violence: The Case of Kenya in the 1990s." *Comparative Political Studies* pp. 1–32.

Bromley, Lars. 2010. "Relating violence to MODIS fire detections in Darfur, Sudan." *International Journal of Remote Sensing* 31(9):2277.

Cederman, Lars-Erik and Kristian Skrede Gleditsch. 2009. "Introduction to Special Issue on "Disaggregating Civil War"." *Journal of Conflict Resolution* 53(4):487–495.

Collier, Paul and Anke Hoeffler. 2002. "Greed and Grievance in Civil War." *Centre for the Study of African Economies Working Paper Series* . Working Paper.

Cressie, Noel. 1996. "Change of Support and the Modifiable Areal Unit Problem." *Geographical Systems* 3:159–180.

Dark, Shawna and Danielle Bram. 2007. "The Modifiable Areal Unit Problem in Physical Geography." *Progress in Physical Geography* 31(5):471–479.

Diggle, Peter. 2003. *Statistical Analysis of Spatial Point Patterns*. Second edition ed. Arnold.

Fearon, J. and D. Laitin. 2000. "Ethnicity, Insurgency, and Civil War." *American Political Science Review* 97(1):75–90.

Gatrell, Anthony, Trevor Bailey, Peter Diggle and Barry Rowlingson. 1996. "Spatial Point Pattern Analysis and its Application in Geographical Epidemiology." *Transactions of the Institute of British Geographers* 21:256–274.

Geonames. N.d. "http://www.geonames.org/.".

Gertzel, Cherry. 1970. *The politics of independent Kenya, 1963-8.* Nairobi: East African Publishing House.

Gibson, Clark C. and James D. Long. 2009. "The presidential and parliamentary elections in Kenya, December 2007." *Electoral Studies* 28(3):497 – 502. ¡ce:title¿Special Symposium: Measurement Methods for Better Longitudinal Modelling¡/ce:title¿.
URL: *http://www.sciencedirect.com/science/article/pii/S0261379409000079*

Gleditsch, Nils, Peter Wallensteen, Mikael Eriksson, Margareta Sollenberg and Havard Strand. 2002. "Armed Conflict 1946-2001: A new dataset." *Journal of Peace Research* 39(5):615–637.

Harris, Andy. 2011. "A Method for Extracting Information about Ethnicity from Names." *Working Paper, Department of Government, Harvard University.* .

Hegre, Havard and Clionadh Raleigh. 2007. "Population Size, Concentration, and Civil War: A geographically disaggregated analysis." *Post-Conflict Transitinos Working Paper No. 16* .

ICC. 2010. "ICC-01/09: Prosecutor's Application Pursuant to Article 58 as to William Samoei Ruto, Henry Kiprono Kosgey, and Joshua Arap Sang." http://www.icc-cpi.int/iccdocs/doc/doc985613.pdf.

ICC. 2012. "Decision on the Confirmation of Charges Pursuant to Article 61(7)(a) and (b) of the Rome Statute.".

IED. 2002. "Registration of Voters in 2002: An Audit Report." Nairobi: .

IED. 2007. "Registration of Voters in 2007, An Audit." Nairobi: .

Kalyvas, S. 2003. "The Ontology of Political Violence: Action and identity in civil wars." *Perspectives on Politics* 1(3):475–494.

Kanyinga, Karuti. 2009. "The Legacy of the White Highlands: Land Rights, Ethnicity and the Post-2007 Election Violence in Kenya." *Journal of Contemporary African Studies* 27(3):325–344.

Kasara, Kimuli. 2006. "Ethnic Beachheads and Vote Buying: Explaining the Creation of New Administrative Districts in Kenya, 1963-2001." Paper presented at the annual meeting of the American Political Science Association, Marriott, Loews Philadelphia, and the Pennsylvania Convention Center, Philadelphia, PA, Aug 31, 2006.

Kasara, Kimuli. 2009. "Electoral Geography and Conflict in Kenya: Examining the Local-Level Causes of Violence in Rift Valley Province after the 2007 Election." *Working Paper* .

Kasara, Kimuli. 2010. "Local Segregation and Inter-Ethnic Violence in Kenya.".

Kiage, Odera and Kwame Owino. 2010. History, Politics, and Science of Opinion Polls in Kenya. In *Tensions and Reversals in Democratic Transitions: The Kenya 2007 General Elections*, ed. Karuti Kanyinga and Duncan Okello. Nairobi: Society for International Development and Institute for Development Studies, University of Nairobi chapter 9, pp. 311–371.

Klopp, Jacqueline. 2001. ""Ethnic Clashes" and Winning Elections: The case of Kenya's electoral despotism." *Canadian Journal of African Studies* 35(3):473–517.

KNCHR. 2008. *On the Brink of the Precipice: a human rights account of Kenya's post-2007 election violence*. Nairobi, Kenya: Kenya National Commission on Human Rights.

Leys, C. 1975. *Underdevelopment in Kenya: the political economy of neo-colonialism, 1964-1971*. London: Heinemann.

Ltd., Oakar Services. 2008. Consultancy on development of a GIS database of learning Institutions (School mapping exercise). Inception report Ministry of Education Nairobi: .

Lynch, Gabrielle. 2006. "The Fruits of Perception: "Ethnic Politics" and the Case of Kenya's Constitutional Referendum." *African Studies* 65(2):233–270.

Lynch, Gabrielle. 2008. "Courting the Kalenjin: The failure of dynasticism and the strength of the ODM wave in Kenya's Rift Valley province." *African Affairs* 107(429):541–568.

Lynch, Gabrielle. 2011. *I Say to You*. Chicago: University of Chicago Press.

Map, Open Street. N.d. "http://www.openstreetmap.org.".

McIntosh, Craig. 2008. "Estimating Treatment Effects from Spatial Policy Experiments: An Application to Ugandan Microfinance." *The Review of Economics and Statistics* 90(1):15–28.

Mueller, Susanne. 2008. "The Political Economy of Kenya's Crisis." *Journal of Eastern African Studies* 2(2):185–210.

Njonjo, Apollo. 1977. The Africanization of the "White Highlands" and the Agrarian Class Struggle in Kenya PhD thesis Princeton University Princeton, N.J.: .

Ogot, B.A. 1995. The Decisive Years, 1945-1963. In *Decolonization and Independence in Kenya*, ed. B.A. Ogot and W.R. Ochieng'. London: James Currey chapter 3.

Openshaw, Stan. 1984. *The Modifiable Areal Unit Problem.* Number 38 *in* "Concepts and Techniques in Modern Geography" Kent: Geo Books.

Oucho, John. 2002. *Undercurrents of Ethnic Conflict in Kenya.* Leiden: Brill.

Prins, E. 2008. "Use of Low Cost Landsat ETM+ to Spot Burnt Villages in Darfur, Sudan." *International Journal of Remote Sensing* 29(4):1207–1214.

Roy Jorgensen Associates, Inc. 2006. Classified Roads and Market Town Datasets. Technical report Ministry of Roads Nairobi: .

Rutten, Marcel and Sam Owuor. 2009. "Weapons of Mass Destruction: Land, Ethnicity and the 2007 Elections in Kenya." *Journal of Contemporary African Studies* 27(3):305–324.

Shixing, Liu, Zhang Tongming, Song Weiguo and Xiao Xia. 2010. "An Enhanced Algorithm for Forest Fire Detection Based on MODIS Data." *International Conference on Optoelectronics and Image Processing* 1:200–203.

Tatem, Andrew J., Abdisalan M. Noor, Craig von Hagen, Antonio Di Gregorio and Simon I. Hay. 2007. "High Resolution Population Maps for Low Income Nations: Combining Land Cover and Census in East Africa." *PLoS ONE* 2(12):e1298.

Varshney, A. 2003. "Nationalism, Ethnic Conflict, and Rationality." *Perspectives on Politics* 1(1):85–99.

Verwimp, Philip, Patricia Justino and Tilman Bruck. 2009. "The Analysis of Conflict: A Micro-Level Perspective." *Journal of Peace Research* 46(3):307–314.

Waki, Justice Philip. 2008. *Report of the Commission of Inquiry into Post Election Violence.* Nairobi, Kenya: Government Printer.

Wang, Wanting, John Qu, XianJun Hao, Tongqiang Liu and William Sommers. 2007. "An Improved Algorithm for Small and Cool Fire Detection using MODIS Data: A Preliminary Study in the Southeastern United States." *Remote Sensing of Environment* 108(2):163–170.

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

Willis, Justin. 2008. "What Has He Got up His Sleeve? Advertising the Kenyan Presidential Candidates in 2007." *Journal of Eastern African Studies* 2(2):264–271.

Wolf, Thomas. 2009. ""Poll Poison?": Politicians and polling in the 2007 Kenya Election." *Journal of Contemporary African Studies* 27(3):279–304.

6.1 Appendix: Placebo Test

One concern lies in the quality of the treatment indicator. The MODIS satellite records active fires in the area only twice per day, and some fires may not be due to violence. As a result, our results could be due to simple chance. A method for evaluating the robustness of these results is a placebo test (Abadie, Diamond and Hainmueller, 2010): do control units, if assigned to a placebo treatment, experience zero average treatment effect? A placebo simply means a treatment known to have no effect. To implement a placebo test, we simply categorize some units (which did not received treatment in reality) to the treated group. Then, we estimate the average treatment effect, as in the above analysis. Since we know that the actual effect of the placebo at the unit level is zero, the expected value of the placebo treatment effect should be zero as well.

Like the placebo test in Abadie, Diamond and Hainmueller (2010), our inferential goal is to determine whether our estimated effects for each ethnic group is large *relative* to estimates generated by chance. Given that treatment in this study is explicitly spatial, we can create a placebo treatment by simply generating random "placebo" fires across the study area, and then estimating the treatment effect as above. By repeating this process many times, we can estimate the distribution of placebo effects for each ethnic group.

Figure 3 shows the results of our placebo test. The central finding is in the upper-right panel. For Kikuyu, none of the estimated effects from the placebo treatment were more extreme than the estimated effect of -0.31 (from table 3). This suggests that our estimated result is unlikely to have been generated by chance. Taking into account the treatment indicator issues discussed above, the placebo test also suggests that, absent the treatment indicator issue, our overall results would be reinforced.

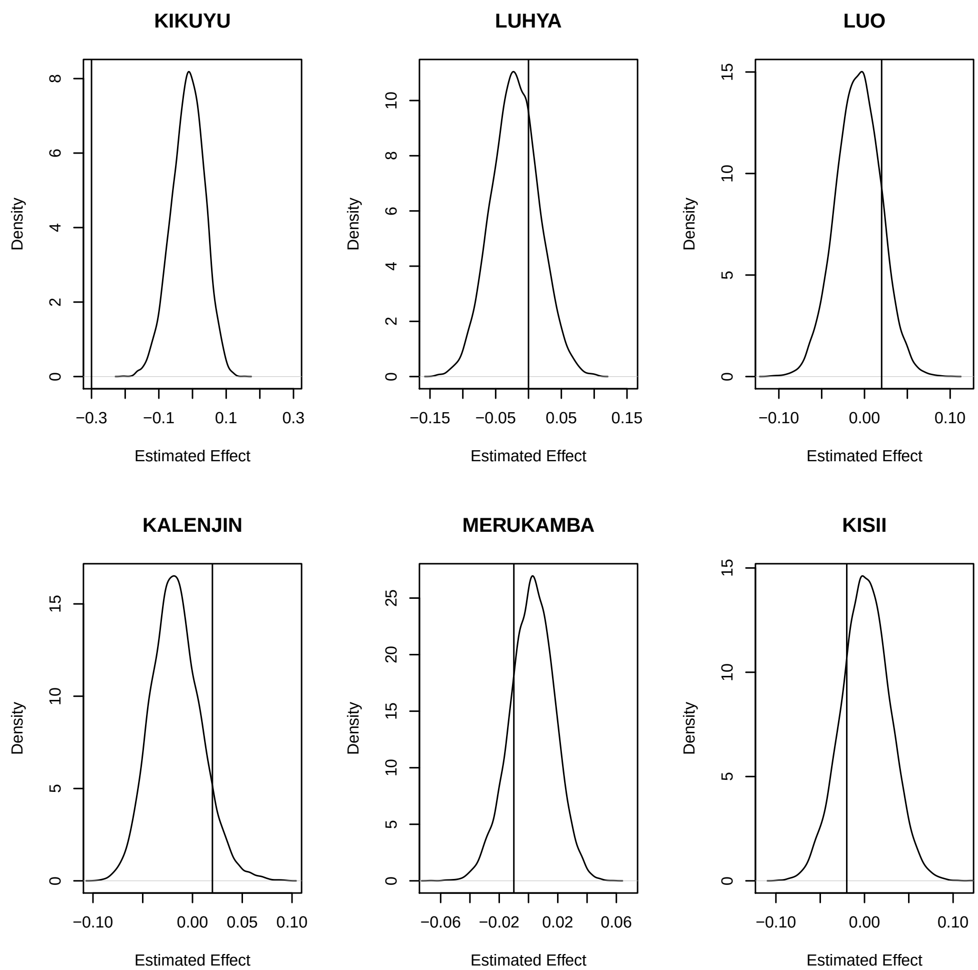

Figure 3: Placebo Test Results: This graph compares the density of 10000 estimated placebo effects with the effect estimated from observed fires for each ethnic sub-group. Vertical black line marks the actual estimated treatment effect in column 1 of table 3. For Kikuyu, none of the estimated effects from the placebo treatment were more extreme than the estimated effect of -0.31 (from table 3).