The Geography of Repression in Africa *

Darin Christensen †

Draft Date: December 8, 2013

Abstract

This paper documents two novel patterns in the use of repression in response to protests in Africa: (1) repression is more frequent in response to social conflicts in urban areas, but (2) if the state does employ repression, it more frequently kills dissidents in rural areas. I argue that protests in rural areas pose a smaller threat to governments and, thus, prompt less frequent intervention. However, when governments do decide to repress rural protests, they are less constrained by concerns that lethal repression might incite a backlash, as there are fewer bystanders in these more sparsely populated areas that might join the fray. I formalize these intuitions and test the implications of the model with data on the use of non-lethal and lethal repression in response to protests that occur in localities of varying population density. I find empirical support for my hypotheses that is robust to a number of plausible alternative explanations.

*I am grateful for feedback on earlier drafts from Nick Eubank, James Fearon, Steve Haber, David Hausmann, Francisco Garfias, Grant Gordon, Eric Kramon, David Laitin, Clayton Nall, Ramya Parthasarathy, Jonathan Rodden, Manuela Travaglianti, Jeremy Weinstein, and Kelly Zhang. I owe special thanks to James Fearon for his collaboration in formalizing the model.

†Ph.D. Candidate and Stanford Graudate Fellow, Political Science, Stanford University. Email: darinc@stanford.edu.

1 Motivation: Two Patterns of Repression

In January 2004, university students in Nairobi, Kenya "went on a rampage," destroying property and demanding that the University of Agriculture Technology fire its vice chancellor and reopen (AFP, 2004a). Nine months later 600 University of Nairobi students violently demonstrated in response to proposed fee hikes, blocking streets and stoning passing motorists (AFP, 2004b). In both instances, the Kenyan police used tear gas to quell the riots. In June 2005, students and parents in a village outside of Garissa, Kenya – a small city of 70,000, 350 kilometers northeast of Nairobi – protested the seizing of school land for private development. In this case, the police response was more heavy-handed: Police opened fire on these students and parents, killing at least one and wounding dozens of others (AFP, 2005). All three conflicts occurred under the same regime and involved students speaking out against their schools' administration. If anything, the available reports suggest that the students in the capital were more numerous and violent. What then explains why police opened fire in rural Garissa but employed non-lethal means in Nairobi?

Existing studies of repression provide little insight into why the Kenyan police responded differently to these protests. The literature focuses on cross-national differences in the level of repression and, in so doing, masks subnational variation in whether and how leaders and their agents in the police or military respond to protests. Yet, understanding this geographic variation provides greater insight into executives' decisions about whether to permit or repress (using different degrees of force) public challenges.

This paper describes and offers an explanation for two striking patterns in the use of repression across African countries. Using recently compiled data on social conflicts in 47 African countries between 1990-2010 I find that (1) repression is more frequent in response to social conflicts in urban areas; but (2) if the state does employ repression, it more frequently kills dissidents in more rural areas (defined broadly in the data to include all localities with a population of less than 100,000).[1] Figure 1 illustrates (on the left) that the probability of repression in response to a social conflict is roughly one-third in urban areas but less than one-fifth in rural areas. However, when repression does occur, it is lethal in only a quarter of urban social conflicts but nearly half of rural events.

These patterns are not driven by a small number of countries in the sample. The urban-rural differences displayed in Figure 1 are apparent in the vast majority of African countries for which we have at least 10 social conflicts in both urban and rural areas. For all countries in blue on the left hand side of Figure 2, the probability of repression is higher in urban areas than in rural areas. For all countries in red on the right hand side of the figure, the conditional probability of lethal repression is higher in rural areas.

I proceed by building a formal model to rationalize these regularities, and then empirically evaluating key comparative statics of the model. In short, I argue that governments have little incentive to intervene in small, rural protests; yet, when they do intervene in these settings, they are not constrained by concerns that lethal repression will incite a costly backlash. I find empirical support for these hypotheses: repression is more likely when social conflicts occur in densely populated areas; however, when events in more sparsely populated areas are repressed, repression is more likely to be lethal. These results are robust to the inclusion of plausible omitted variables, including aid dependence, ethnicity, event characteristics, history of armed conflict, proximity to natural resources, and regime type. I conclude the empirical section by also addressing concerns that events in capital cities, particular countries, or reporting bias are driving the results.

This paper adds to our understanding of the conditions under which governments employ repression in several ways: first, it describes two previously undocumented patterns in the use of repression across Africa; second, it offers a formal logic that relates these patterns to governments' desires to suppress dissent without escalating the scale of protests; and third, it uses geo-coded observational data to evaluate the validity of the theory as well as the plausibility of several alternative explanations.

[1] Social conflicts include events, such as riots, strikes, and protests that do not occur during a civil conflict, which is defined by the Uppsala Conflict Database as a conflict over territory or government with more than 25 battle deaths per year. Urban is defined as a city of over 100,000 and the country's capital (Hendrix et al., 2012).

Figure 1: Repression across Urban and Rural Social Conflicts

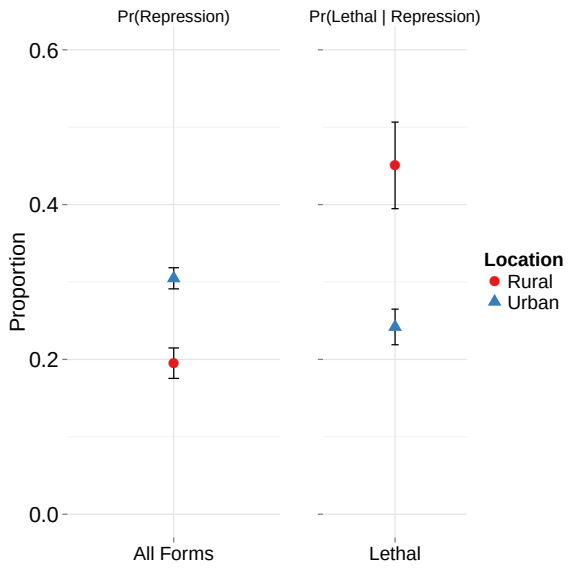


Figure 1: The left panel displays the probability of repression for social conflicts that occur in urban (population $\geq$ 100,000 & capital city) and rural areas; the right panel, the probability of lethal repression in both urban and rural areas within the subset of events that involve some form of repression. The black bars represent 95% confidence intervals around each estimate. Events that overlap with the Uppsala Armed Conflict Database or are represented twice in the SCAD (see codebook for further details) are first excluded.

Figure 2: Uniformity of Patterns across Countries

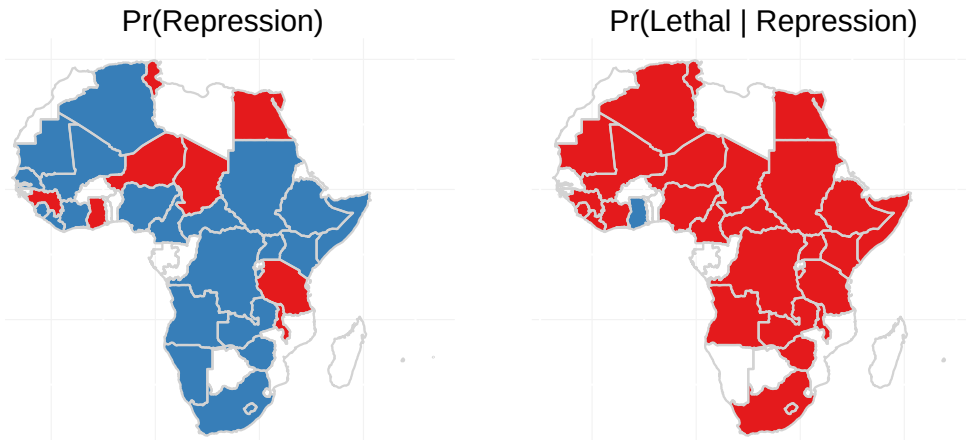


Figure 2: If a country is shaded *blue* then $\Pr(X|\text{Urban}) > \Pr(X|\text{Rural})$; if a country is shaded *red*, then $\Pr(X|\text{Urban}) < \Pr(X|\text{Rural})$; countries in in *white* did not experience at least 10 social conflicts in both urban and rural areas. In the left panel, $X = \mathbb{1}(\text{Repression})$; in the right panel, $X = (\mathbb{1}(\text{Lethal}) \mid \mathbb{1}(\text{Repress}) = 1)$, where $\mathbb{1}(\cdot)$ is simply an indicator function.

2 Background

2.1 Explanations for Repression from Cross-National Studies

Most existing studies of repression focus on cross-national variation (see Davenport, 2007). This work yields two central findings: first, repression is costly and, thus, unlikely to be employed against a docile population (Lichbach, 1984); and second, democracies repress less, because they can peacefully

incorporate opposition groups and oust abusive leaders (Davenport, 1995).[2] Unfortunately, these two prevailing theories do not help explain the patterns noted above: first, I do not consider the repression of docile citizens, but rather restrict attention to repression that comes in response to protests; second, regime type does not vary subnationally in the cross-section, and what inter-temporal variation exists is unable to explain why repression is more frequent in urban areas but more lethal in rural areas.

A more relevant, but also more contested body of literature, does not attempt to explain repression, but rather considers whether repression deters or inflames protests. Backlash theories contend that repression can be counter-productive, and case studies from East Germany, Iran, Palestine, and South Africa have been cited in support of this core claim (Ortiz, 2007). In particular, scholars have argued that repression that fails to suppress opponents, harms the innocent, confirms perceptions of regime vulnerability, or is aimed at groups that the public considers representative and justified is more likely to inspire backlash (see Goldstone, 2001, for a review). Quantitative analysis using either cross-sectional or time-series data on repression and protest has not provided convincing evidence in support of these hypotheses. However, these mixed findings could, in part, be attributable to an unaddressed selection problem: governments are strategic actors whose decisions to employ repression incorporate the probability of backlash. We may then observe repression primarily in contexts where backlash is less likely and, thus, underestimate the potential for repression to inflame dissent. In fact, one of the contributions of this paper is to demonstrate that governments, in deciding how forcefully to repress protests, weigh the removal of protesters against the possibility that repression may incite bystanders.

A second contribution of this project is to add to the small but growing number of recent studies that leverage sub-national variation to study the use of repression (e.g., Bohara et al. (2006); Murtagh et al. (2009); Vadlamannati (2008)). Davenport (2007, p. 18) implores scholars to move beyond country-year analyses: "Such an approach is not only essential for gauging the robustness of the propositions developed in this [cross-national] literature but also allows us to explore other arguments that have previously been ignored." The use of sub-national data not only reduces concerns about omitted variables that plague cross-national comparisons, but also allows for the development and testing of theories (such as the model proposed below) regarding how governments target repression – a line of research that can help identify vulnerable sub-populations.

2.2 Explanations for the More Severe Treatment of Rural Citizens in Africa

This paper also builds on existing work in African political economy, which argues that government policies are biased in favor of urban areas. In his seminal work, Bates (1981) finds that African leaders depress food costs in cities to maintain the purchasing power of workers and avoid destabilizing social conflict in urban areas:

> "The issue that most frequently drives African city dwellers to militant action is the erosion of their purchasing power... Sadat, Nimeiri, Kaunda, Moi, Gowan, and Tolbert are among the other African leaders whose governments have felt the political pressures generated by the erosion of the purchasing power of urban dwellers; in the face of these pressures, several have fallen" (pp. 30-2).

Keeping food prices low in cities requires paying below-market prices to rural farmers. Where farmers resist this expropriation, Bates claims they are coerced: "Through the use of violence, the governments of Africa have forestalled the mobilization of the rural majority against policies that harm their interests" (p. 112).

Bates marshals extensive evidence that agricultural policies benefit urban workers at the expense of rural farmers. However, he does not present comparable evidence to confirm his arguments regarding the use of repression in the countryside. This paper investigates two of Bates's key claims about the handling of urban and rural social conflicts: first, that governments are more responsive to social conflicts

[2] A number of works associate economic variables, such as development, inequality, and openness with levels of repression. However, empirical work has frequently yielded conflicting results: Hafner-Burton (2005), for example, demonstrates that the correlation between openness and repression is highly measure and model dependent.

in urban centers; but second, that repression is more likely to backfire in cities. "Direct attacks on labor movements are open to reprisals; in moments of economic stress, labor movements can join with their urban constituents, paralyze cities, and create the conditions under which ambitious rivals can displace those in power" (p. 33). Left unaddressed, social conflicts in cities pose greater threats to leaders survival; and yet, lethally repressing these conflicts also has the potential to do more harm than good.

Through its analysis of event-level data on the use of repression, this paper contributes both to the literature on why certain leaders use repression and regionally focused work on urban bias. The larger goal is to better understand why, within any given regime, some protests are (lethally) repressed and other are not. Insights into how leaders selectively deploy coercion have policy implications that fall short of requiring regime change to reduce the lethal suppression of protests or riots.

3 A Model of Repression

The model presented below formalizes the trade-offs that a government faces when deciding whether and how forcefully to repress a protest. First, does the protest warrant any intervention? The government may not want to bear the costs of repressing a small group of relatively restrained protesters. Earl and Soule (2006, p. 146) note (unsurprisingly) that "threat is the most widely accepted and empirically supported explanation of repression developed thus far." From the state's perspective, the returns to repression should then be increasing in the threat posed by the initial group of protesters.

Second, while repression imposes a cost on protesters and can, thus, be effective at suppressing dissent, intervening in a protest or riot is a public act. Bystanders (i.e., citizens not involved in the initial protest) observe the government's decision about whether and how brutally to repress and may use this information to update their beliefs about how much the government cares about its citizens. Governments should then worry about how repressing a protest – for example, by firing on demonstrators – will affect their reputation for being benevolent or brutal, as bystanders are more inclined to join protests against a reputedly bad government. In her analysis of urban protests in Iran prior to the Revolution, Rasler (1996, p. 142-7) argues that several "critical events" – most often involving the use of lethal repression – "represent important turning points in collective action... these events propel large numbers of people into collective action." The deaths of earlier protesters were acknowledged in "mourning ceremonies," and "these observances produced violent clashes between security forces and the public and generated new deaths and a new cycle of mourning throughout the country." In Rasler's account, the use of lethal repression incited other citizens to publicly oppose the Shah, and this escalation contributed to his eventual ousting.

To summarize, repression offers the government an opportunity to suppress public dissent, but also carries the risk of revealing that a government cares little about its citizens – a reputation that can incite conflicts that are larger and, thus, costlier than the original protest. I argue that a government is likely to resolve this strategic dilemma differently based on where a protest takes place. In particular, larger initial protests are likely to pose a greater threat to the executive, prompting more frequent repression. However, in densely populated areas the government has to be most concerned about the reactions of bystanders to brutal forms of repression, as there are more people who may be incited by lethal repression to take to the streets. Hence, we are more likely to observe repression in urban areas, where larger demonstrations or riots take place. Yet, when the government does repress urban protesters, brutal types will be constrained by fears that lethal force might prompt a costly escalation and, thus, opt for non-lethal repression in urban settings.

3.1 Setting Up the Model

These intuitions are formalized in a model, where bystanders have incomplete information about the government's type. Consider a protest by a vanguard of dissidents. With publicly known probability, $\alpha \in [0, 1]$, these protesters face the brutal type, who receives greater utility from repression than the good type, $\theta_j \in \mathbb{R}^1_+$ for $j \in \{G, B\}$ such that $\theta_B > \theta_G$. There are a continuum of bystanders in the locality where this protest occurs, each indexed by i and of total measure n.

The sequence of the game is as follows. There has been a protest by a vanguard of dissidents of size p. The game starts with the executive choosing a level of repression, $r \in \{0\} \cup [\underline{r},\ \overline{r}]$, where $r = 0$ corresponds to no repression, $\underline{r}$ to non-lethal repression, and any $r > \underline{r}$ to increasingly brutal forms of repression. Each bystander, i, observes r and decides whether or not to join the vanguard in protesting.

The bystander receives $v_i \in \mathbb{R}^1$ if they join the protest against the bad type, 0 if they join the protest against a good type, and $q \in \mathbb{R}^1_+$ for not joining regardless of the executive's type. v_i captures bystander i's costs of joining and the positive utility they derive from protesting against a brutal executive. Note that v_i can be less than zero for individuals that face very high costs of joining a protest or receive no positive utility from dissent, even against a bad government. v_i is distributed according to a distribution function $F_V(\cdot)$.

Leaders care about maintaining power and the associated stream of benefits. I focus here on a component of their utility that varies with protests and leaders' decisions about how to deploy repression. The executive's payoff is then $u_j = p\theta_j r - cn[1 - F_V]$ for $j \in \{G, B\}$. This function captures the intuition that repression is double-edged sword, which can both quell and inflame dissent. Looking at the first term, the returns to repression are increasing in the size of the initial protest, $p \in \mathbb{R}^1_+$. Large protests are more likely to shut down major roads and disrupt commerce and government activities. Restoring order in these cases through the use of repression, thus, brings greater returns to the executive.[3] However, the executive also pays a cost, $c \in \mathbb{R}^1_+$, for every additional bystander that joins the protest. The cost of any backlash is captured in the second term, which multiplies this marginal cost, c, by the proportion of bystanders that join after observing the executive's decision regarding whether and how forcefully to intervene in a protest.

3.2 Equilibria

Proposition 1. Assuming conditions (1) and (2) in the proof below hold, there exists a Perfect Bayesian Equilibrium, in which both types pool on $r = \underline{r}$ or non-lethal repression. More specifically, the PBE is characterized as follows

$$\left(s_{\theta_B} = s_{\theta_G} = \underline{r},\ s_i = \begin{cases} Join & v_i > q/\alpha \\ \sim Join & o.w. \end{cases} \text{ and } \beta(r) = \begin{cases} 1 & r > \underline{r} \\ \alpha & r = \underline{r} \\ 0 & r < \underline{r} \end{cases} \right)$$

where $\beta(r)$ represents i's belief that the executive is a bad type after observing r.

Proof. If both types pool on $\underline{r}$, then bystander i can not update their prior belief and chooses to *Join* the protest only if their expected return to joining is greater than their reservation utility, $v_i\alpha > q$. The proportion of all bystanders, n, that choose to join is the proportion for which $v_i > q/\alpha$ or $1 - F(q/\alpha)$.

Suppose off-the-path beliefs are such that if a bystander observes repression in excess of non-lethal repression, then they believe that they are certainly facing a brutal type. However, if they witness no repression, then the bystander believes that they are definitely facing a good type. In terms of the notation used above, the posterior belief, $\beta(r)$, equals one for $r > \underline{r}$ and zero for $r < \underline{r}$. More generally, in any pooling equilibrium on r^*, bystanders who observe repression less severe than r^* infer that they face a good type; repression more brutal than r^* conveys to bystanders that they are definitely confronting a bad type.

[3] At first glance, it may appear odd that the executive's utility function is increasing in the size of the initial protest. However, note that we could subtract pK from this utility function, with $K \in \mathbb{R}^1_+ : K > \theta_B \overline{r}$, to ensure that the executive's utility is always decreasing in p, but decreasing at a slower rate when they deploy repression. As p is not endogenously determined, including $-pK$ would only add parameters without affecting any of the results presented below.

Both types choose non-lethal repression ($s_{\theta_B} = s_{\theta_G} = \underline{r}$) in this pooling equilibrium, and receive: $u_j = p\theta_j\underline{r} - nc[1 - F(q/\alpha)]$ for $j \in \{G, B\}$. These strategies are incentive compatible if

$$\begin{aligned} p\theta_G\underline{r} - nc[1 - F(q/\alpha)] > 0 \\ p\theta_G\underline{r} > nc[1 - F(q/\alpha)] \end{aligned} \tag{1}$$

$$\begin{aligned} p\theta_B\underline{r} - nc[1 - F(q/\alpha)] > p\theta_B\overline{r} - nc[1 - F(q)] \\ nc[F(q/\alpha) - F(q)] > p\theta_B(\overline{r} - \underline{r}) \end{aligned} \tag{2}$$

□

Recall that $\theta_B > \theta_G > 0$, so if (1) is satisfied for the good type, it will also be satisfied for the bad type. By the same logic, if (2) is satisfied for the bad type, it will also be satisfied for the good type. These constraints relate back to the intuitions provided above. The first incentive compatibility constraint suggests that, if the vanguard of protesters is large enough relative to the total population, the good type can justify repression at the expense of remaining indistinguishable from the bad type. The second implies that even executives with little regard for their citizens' welfare will not want to lethally repress protesters in settings where revealing their type can touch off a sizable backlash – settings where bystanders hold a low prior belief that the executive is brutal or those where there are a large number of bystanders relative to the size of the vanguard.

Holding the population of bystanders and off-the-path beliefs fixed, as the size of the protest (p) grows, this pooling equilibrium becomes unsustainable. Bad types are the first to defect, and a separating equilibrium occurs in which good types continue to employ non-lethal repression ($\underline{r}$), but bad types now resort to lethal forms of repression ($\overline{r}$). The bad type has no incentive to deviate from this separating equilibrium, so long as the benefits of lethally cracking down on the protesters exceeds the value of disguising their type and, thus, reducing the proportion of bystanders that join. More formally, the incentive compatibility constraint for the bad type is

$$\begin{aligned} p\theta_B\overline{r} - nc[1 - F(q)] > p\theta_B\underline{r} \\ p > \frac{nc[1 - F(q)]}{\theta_B(\overline{r} - \underline{r})}. \end{aligned} \tag{3}$$

Good types continue to repress non-lethally, so long as the cost of inciting bystanders to join the protest exceeds the added benefit of dispersing the vanguard with lethal repression:

$$\begin{aligned} p\theta_G\underline{r} > p\theta_G\overline{r} - nc[1 - F(q)] \\ p < \frac{nc[1 - F(q)]}{\theta_G(\overline{r} - \underline{r})}. \end{aligned} \tag{4}$$

When this last constraint is violated, no pure strategy equilibrium exists. Suppose both types pooled on lethal repression ($\overline{r}$). By deviating to just below $\overline{r}$, a good government can reveal their type to bystanders and prevent a backlash, a highly profitable deviation.

The equilibrium shift characterized above indicates that, as the size of the vanguard grows relative to the population of bystanders, executives (particularly bad types) worry less about provoking a backlash and grow more concerned about the immediate consequences of failing to disperse the threat posed by the mob of initial protesters.

Suppose now that we hold the size of the protest and off-the-path beliefs constant and vary the population of bystanders (n). We can rearrange the four equations numbered above to identify the values of n that sustain different equilibria. Rather than restate those intervals, Figure 3 provides a graphical depiction of the combinations of vanguard size and bystander population that sustain the different equilibria characterized above. To construct the figure, I set all parameters other than p and n to arbitrary values and then plot the equilibrium – if any exists – that prevails for different values in the grid.

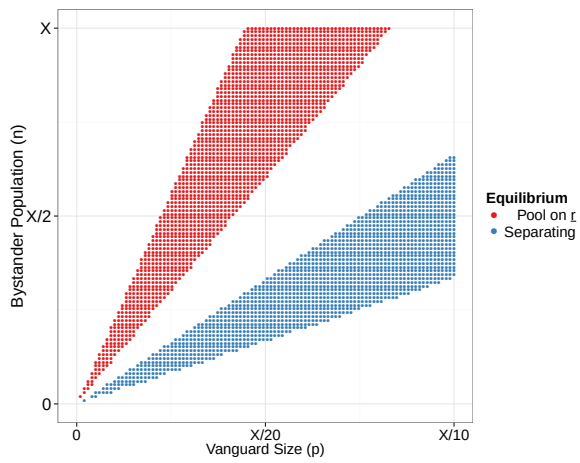


Figure 3: Equilibrium as p and n Vary

3.3 Hypotheses

As the discussion and figure above suggest, a relatively large population of bystanders is more likely to discipline brutal types, while a larger group of initial protesters is more likely compel good types to take repressive action. The empirical portions of the paper focus on these two predictions from the model:

H1: Increasing the size of the vanguard increases the probability of repression.

H2: Increasing the population of bystanders reduces the probability that repression, if it does occur, will be lethal.

4 Illustrative Case: 2008 Food Riots in Cameroon

A good case for evaluating the realism of the proposed model involves an executive facing multiple protests comprised of individuals with similar grievances, but taking place in different locations (i.e., locations that vary in terms of population density).

From 25-29 February 2008 riots broke out in 31 localities across Cameroon. The 2008 riots were all motivated by marked increases in the cost of food and fuel, as well as a recent attempt by President Biya to extend his rule through a constitutional amendment (AFP, 2008). The riots erupted spontaneously, and the composition of rioters does not appear to have differed across localities:

> "Many youths spontaneously descended to the streets to express their disillusionment and loss of hope for a better future. The strike then became a widespread movement. No political organization or trade union instigated the protesters. It was all spontaneous. The cities were totally paralyzed. Peaceful demonstrations could be seen everywhere" (National Human Rights Observer, p. 10).

As is apparent in Figure 4, rioting was concentrated in the five northwestern regions of Cameroon, with an additional riot occurring in the capital, Yaounde.[4] Although rioting did not occur uniformly across Cameroon, the 31 localities involved vary in terms of their population density. That said, even the smallest of these localities would be best characterized as towns, rather than smaller villages. The concentration of riots in the northwestern regions reduces concerns about ethnic heterogeneity driving variation in the use of repression: while these regions contain four major ethnic groups (the Baileke,

Figure 4: Location of 31 Reported Food Riots

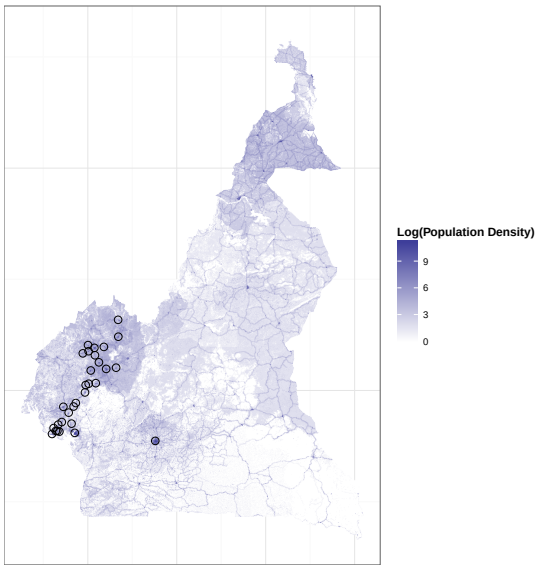


Figure 4: Riot locations are circled on the map. These locations are based on government reports reproduced by the National Human Rights Observer (p. 10). The color gradient represents the log of population density at a resolution of one square kilometer as measured by Oak Ridge National Laboratory (2010). NA values in the raster are colored white for presentation purposes.

Duala, Tikar, and Tiv), none of these groups share ties with the Bulu, who comprise much of the country's political and military elite, including the President (GREG, 2010; MRGI, 2007).

Across these 31 localities, Biya faced choices about whether to repress protesters and, if so, how to confront them. In Yaounde and Douala, the country's largest cities and political and commercial capital respectively, Biya rapidly deployed police and other security forces to repress protesters. Initially, these forces deployed non-lethal means – arrests, batons, tear gas, and water canons. Remarkably, only two deaths were reported in Yaounde, a police officer and another individual whose cause of death is unknown (National Human Rights Observer). In Douala, police did resort to using lethal force, but reports suggest that this change in tactics occurred after they lost control of the situation. By contrast, in more rural areas police did not attempt to use these same non-lethal means before killing rioters: in Bafou looting was punished with "summary executions in the plantations"; in Loum, "many rioters were shot and killed, including six young people" (National Human Rights Observer, pp. 12-13). Biya's responses to these riots are largely consistent with the predictions of the model presented above: riots in urban centers provoked an immediate, initially non-lethal response; in the countryside, reports suggest that repression was more erratic and more likely to involve the use of live ammunition than the tear gas or water canons deployed in the larger cities.

Unfortunately, we know little about why Biya made these decisions. As the model suggests, Biya's attempts to avoid killing demonstrators in major cities may have been motivated by a concern that lethal repression could signal disregard for rioters' legitimate concerns about rising food and fuel prices and, thus, expand protests in densely populated areas. However, Biya's actions are also consistent with an alternative explanation – namely that Biya ordered that non-lethal means be used in Douala and Yaounde to avoid raising the ire of international observers, who are concentrated in these cities. The International Crisis Group notes that regime increasingly relies on mass arrests or "judicial repression, which attracts less criticism from international human rights organizations and the international community" (ICG, 2010, p. 14). While this short case study can only speculate as to Biya's decision calculus, it suggests

[4]The NRHO report contains a reference to Muea, a town much further east than all of the other riots. Because this reference is ambiguous, I do not feel confident about this riot's location and exclude the event from Figure 3.

that the predictions of the model are consistent with the actions taken by an executive that faced multiple protests prompted by the same grievances but staged in different locations. The empirical tests undertaken in the next section add further evidence in support of the model and help to rule out alternative explanations, including monitoring by international observers.

5 Social Conflict in Africa Database (SCAD)

The event-level data on social conflicts used in the remainder of the analysis are taken from the Social Conflict in Africa Database Version 3.0, compiled by (Hendrix et al., 2012). The SCAD uses Lexis-Nexis to query all *Associated Press* and *Agence France Presse* news wire stories between 1990 and 2011 for African countries with a population of over 1 million. The union of several keyword searches provides an initial pool of reports.[5] This pool is then sorted, read, and hand-coded. If a story contains information on multiple events, these events both enter the data; if multiple stories cover the same event, the event only enters the data once. If an event takes place within the context of an armed civil conflict (as defined by the start and end dates in the Uppsala Armed Conflict database), violent events associated with the civil conflict are not coded.[6] Events that take place in multiple locations (e.g., a protest that occurs simultaneously in multiple villages) receive separate entries with distinct latitude and longitude coordinates. However, the other variables are not uniquely coded for each individual location. But for location, these events receive the same coding for repression, duration, etc. (Hendrix and Salehyan, 2012, p. 7).

Per the recommendations of Hendrix and Salehyan (2012), I avoid double-counting by only retaining the first locality listed for each event.[7] In the base sample, I also exclude all events that overlap with the Uppsala Armed Conflicts Database. This results in 6,975 events. Unless otherwise noted, the unit of analysis is the social conflict event. While I extract information about the geography surrounding each event (e.g., population density), I am not using cities or towns as units.

Below I provide definitions for several of the variables in the SCAD that are used in the subsequent analysis:

Lethal Repression. The government's use of violence results in the death of a protester.

Non-Lethal Repression. The government's use of violence does not result in the death of a protester.

Note: The coding of whether repression is lethal or not is based on whether a participant is killed by security forces and not the means deployed by the government. The authors of the data indicate that this coding decision eliminates ambiguity about the intentions of security forces, for example, whether police fire live rounds at protesters as warning shots or with the intent to kill (Personal Communication with I. Salehyan, 2012).

Urban. Urban includes events that occur in the capital city, other major urban areas, and multiple urban areas. Urban is defined as a population exceeding 100,000.

Rural. Rural includes events that occur in one or more rural areas.

Number of Participants. Events are placed in one of seven categories defined by the intervals: $[0, 10)$, $[10, 100]$, $(100, 10^3]$, $(10^3, 10^4]$, $(10^4, 10^5]$, $(10^5, 10^6]$, $(10^6, \infty)$.

Grievance. Events are coded as: political, economic, ethnic/religious, or other (see figure 6). These codings are based on the first issue mentioned as the source of tension in the SCAD.[8]

[5] Search terms include: "protest" OR "strike" OR "riot" OR "violence" OR "attack" (Hendrix and Salehyan, 2012).

[6] The Uppsala Conflict Data Program (UCDP) defines an armed conflict as "contested incompatibility which concerns government and/or territory where the use of armed force between two parties, of which at least one is the government of a state, results in at least 25 battle-related deaths" (Department of Peace and Conflict Research).

[7] I also exclude all events that are coded as having occurred "nationwide."

[8] The full articulation of issue categories provided in the SCAD includes 14 categories. The decision to aggregate these categories is based on the apparent similarity of several categories (e.g., "economy" and "economic resources/assets").

Organization. Was the event spontaneous i.e., can clear leadership or organization can be identified? The SCAD includes information about event type. I classify "organized demonstrations," "organized violent riots," "general strikes," "limited strikes," "anti-government violence," "extra-government violence," and "intra-government violence" as organized events, because their coding rules explicitly mention the participation of organized groups i.e., groups with clear leadership (Hendrix and Salehyan, 2012). The two event types coded as spontaneous are "spontaneous demonstrations" and "spontaneous violent riots."

Violence. Was the event violent i.e., did the participants intend to cause physical injury and/or property damage? This variable was also coded based on the event type information included in the SCAD: event types that explicitly include violence in their coding rules (violent riots and pro-, anti-, extra-, and intra-government violence) were coded as violent.[9]

Government Target. Was the central or regional government the target of the event?

Based on the geo-coding for each event in the SCAD, I am able to join the SCAD with other geo-referenced datasets. Rather than listing them in full here, I discuss these different data sources as they are employed in the analysis.

5.1 Examples of Social Conflicts in Rural Areas

The words protest or strike evoke images of students marching on the State House or civil servants picketing outside of ministries. The association between social conflicts and major cities is understandable: these events are more common in major urban areas in Africa, and urban protests have received more scholarly attention (Arriola, 2012, being an excellent recent example). As a result, for the roughly 1,800 social conflicts that occur in more rural areas, many of us have a poorer sense of who is involved in these events and what they are protesting. This section briefly describes a selection of SCAD events in rural Kenya to provide a better sense of what social conflicts look like outside of major cities.

No Repression. Roughly 100 protesters blocked the drilling of four geothermal wells in the Rift Valley in May 2010, complaining that existing wells contaminated local water sources and produced a hissing noise that made it difficult to sleep. There are no reports of police intervention, and officials working on the geothermal project indicated that they were working on an agreement to resettle the affected communities. This protest exemplifies small scale conflicts between larger companies involved in agribusiness or the extractives sector and communities affected by these operations and upset about companies' environmental or social impacts.

Non-Lethal Repression. Several thousand agricultural workers at tea plantations in Western Kenya went on strike in September 1998 demanding higher wages. Newspaper reports suggest that most of the striking workers were employed by major tea exporting companies and were not union members. Four days into the strike, a plantation official was attacked and seriously wounded by striking workers. This prompted an armed intervention by police, which did not claim any lives but left twenty people injured. The strike ended the day after this clash between police and striking workers. While strikes in rural areas are not a regular occurrence – only 51 rural events are coded as limited or general strikes in the base sample I use – it is more common to see actions initiated by farmers, protesting low prices or disadvantageous trade policy.

Lethal Repression An estimated three thousand farmers from the Kirinyaga district in Central Kenya rioted in January 1999, accusing the government of buying rice at below production costs. The farmers refused to deliver their harvest, leaving large quantities in the fields. One person was killed and two others hospitalized with bullet wounds when police fired on the farmers, who, by some reports, were armed with stones, petrol bombs, and bows and arrows. Repression followed by the arrest of the MP, Chairman of the local cooperative society, and several farmers appears to have quelled the riot.

While these events (and the protest in Garissa mentioned in the introduction) are taken from a single country in the sample, they are indicative of protest events that occur outside of major urban areas.

[9] To ensure that none of the results presented below are sensitive to these coding decisions, I also employ the original event type variable included in the SCAD.

5.2 Visualizing the SCAD

The SCAD is a rich data source for scholars interested in African politics and government repression. Figure 5 maps the events in the SCAD over a raster of population density (logged). This figure understates the prevalence of social conflict, as events that occur in the same location are plotted on top of each other.

Figure 5: Map of Social Conflicts in Base Sample

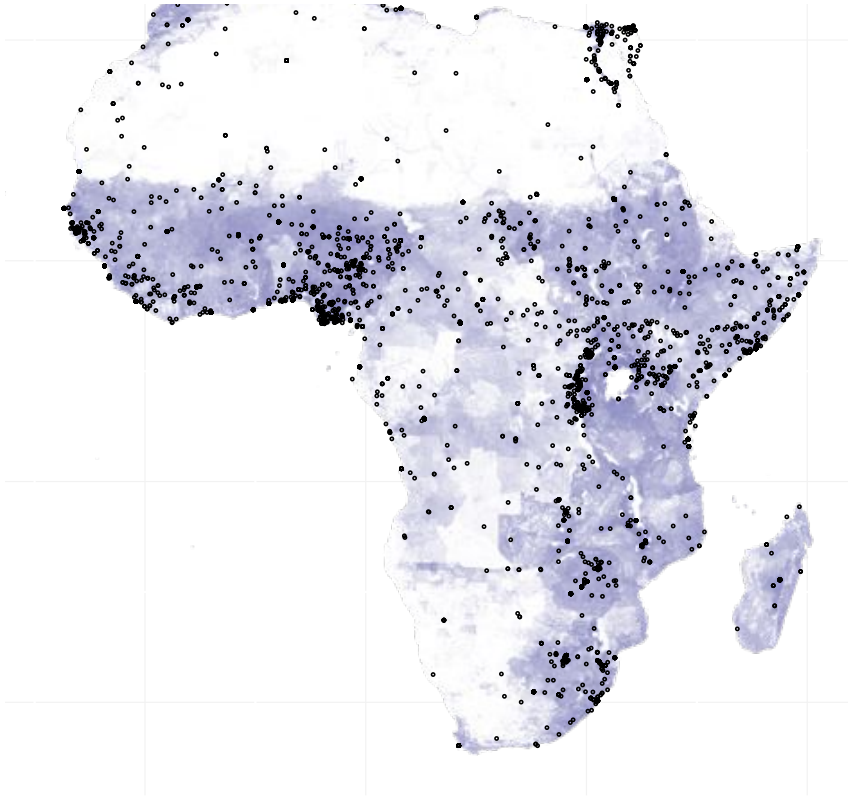


Figure 5: Each event is represented as a black point. These are plotted over a raster of population density (logged) from LandScan (Oak Ridge National Laboratory, 2010).

Several of the variables noted – the nature of protesters' grievances and their tactics (i.e., their level of organization and use of violence) – are potentially relevant to the executive's decision about whether to repress a protest or riot. Figure 6 plots the proportion of events in urban and rural areas that fall into each category of grievances, as well as the proportion that are organized or involve violence. A higher proportion of events in urban areas cite political issues as the primary grievance, while a lower proportion cite ethnic or religious issues. Looking at the last two columns of the plot, a higher proportion of events in rural areas are organized and involve violence. To address the possibility that event characteristics could account for variation in the use of repression, I include these variables as covariates in the subsequent analysis.

Table 1 provides summary statistics for the variables beyond these event characteristics that are employed in the next section. Figure 7 visualizes the frequency of social conflicts in the base sample by both country and year. Darker tiles correspond to more events in a given country year. As is apparent from this figure, social conflicts are not distributed evenly across countries or across time within any given country.

Figure 6: Characteristics of Social Conflicts across Urban and Rural Areas

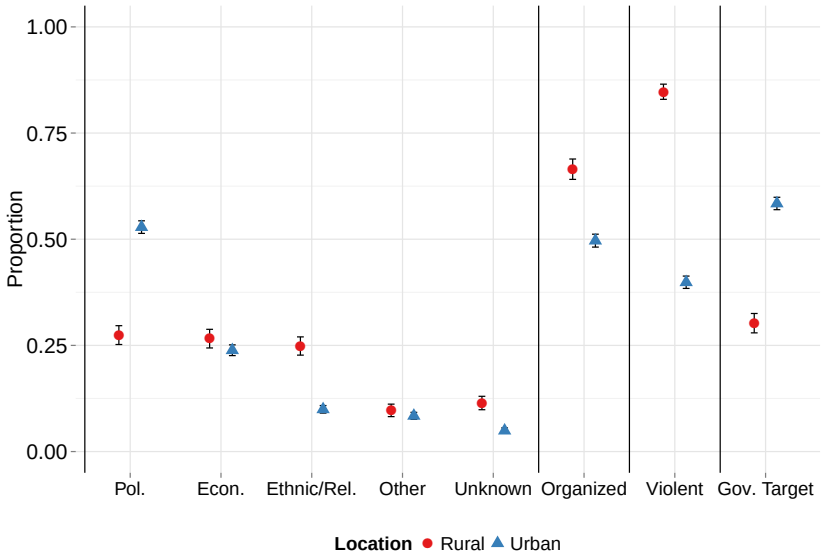


Figure 6: The first five columns plot the proportion of events in urban and rural areas that fall into each category of protesters' grievances ($N = 5,920$), the sixth column displays the proportion of events that are organized as opposed to spontaneous ($N = 5,663$), the seventh column plots the proportion that involve violence ($N = 5,925$), and the final column, the proportion targeted at the central or regional government ($N = 5,925$). While all events are drawn from the base sample, the number of events varies, because different subsets contain information on the grievances, organization, violence, and urban variables. In all cases, the estimate of p is from a multinomial or binomial process, and $SE(p_L) = [p_L(1 - p_L)/N_L]^{(0.5)}$ where $L \in \{\text{Rural, Urban}\}$.

Variable	n	Min	$\tilde{x}$	$\bar{x}$	Max	IQR	s	#NA
$\mathbb{1}$(Repression)	6975	0	0.0	0.3	1.0	1.0	0.4	0
$\mathbb{1}$(Lethal)	6975	0	0.0	0.1	1.0	0.0	0.3	0
$\mathbb{1}$(Urban)	5925	0	1.0	0.7	1.0	1.0	0.4	1050
Log(Pop. Density)	6962	0	8.9	7.7	11.2	4.2	3.2	13
# Participants	2901	1	3.0	3.2	7.0	1.0	1.1	4074

Table 1: Summary Statistics for events in the base sample.

6 Evaluation of Hypotheses

The empirical strategy pursued in this paper is to offer correlations consistent with the hypotheses presented above that are robust to the inclusion of other plausible explanatory variables. Absent control over the data generating process or plausibly exogenous variation in the key independent variables, I refrain from making causal claims.

I rely heavily on least squares regressions in subsequent sections, interpreting $\boldsymbol{\beta}$ the best linear predictors of $\boldsymbol{y}$ given the regressors ($\boldsymbol{X}$) included in the model. I employ linear probability models for ease of interpretation. The data is well-behaved, such that none of the models predict probabilities below zero or above one over the support of the independent variables.

6.1 H1: Pr(Repression) Increasing in Size of Vanguard

The first comparative static I evaluate is whether the probability of repression is increasing in the size of initial protests. I operationalize the size of initial protests in two different ways. I first use data from the SCAD on the number of participants involved in a social conflict. This is the most direct measure that exists for the size of a protest or riot. Unfortunately, there are a number of problems with this measure. First, this measure is missing for a large number of cases, particularly for events that are in more rural areas or are not repressed (see table 1). Second, the coding rules use very broad

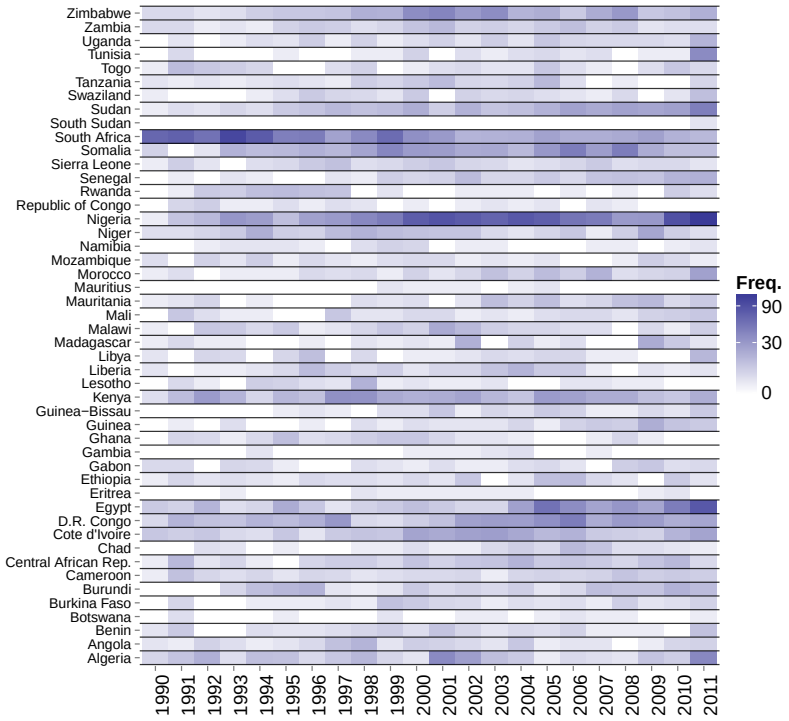


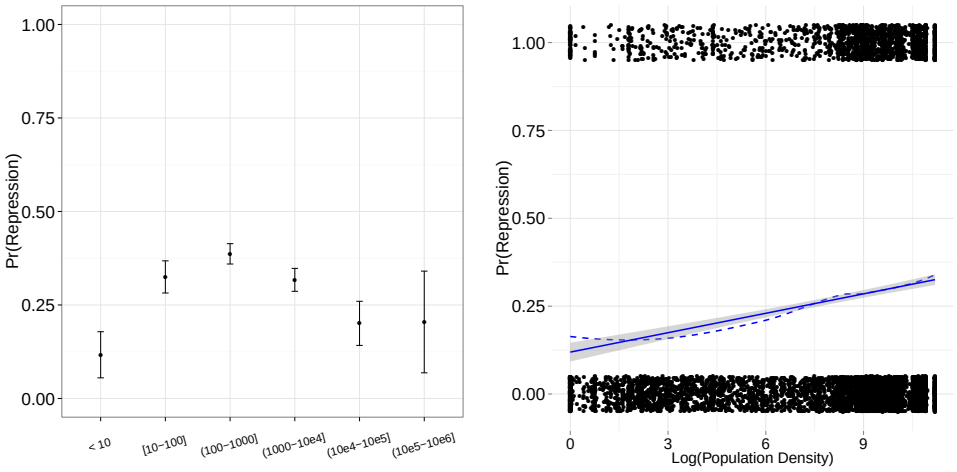
Figure 7: The fill of each tile corresponds to the number of social conflicts in the base sample in each country-year in the SCAD.

categories, and the intervals that define categories dramatically change in size. For example, events with less than 10 participants form one group; events with between 100,000 and 1,000,000 participants form another. Lastly, the measure most likely captures the final size of a social conflict, which may also include bystanders and not just the vanguard.

The second approach uses logged population density as a proxy for the size of initial protests. Given the lower costs of collective action in densely populated areas and a larger population of potential vanguard members, I expect protests in cities to be bigger than those that occur in more rural areas. The data confirms this common-sense prediction: the number or participants variable from the SCAD is positively and significantly correlated with both logged population density and whether an event occurred in an urban area. Using logged population density at the site of each event as a proxy allows me to recover over 4,000 events from the SCAD that have no data on the number of participants. Furthermore, as suggested by the model, the initial and final size of a protest may differ depending on whether repression is used. Unlike the number of participants variable from the SCAD, population density provides a proxy for initial protest size that can not be affected by the decision to use repression. Population density is extracted from the 2010 LandScan raster file, which provides population density measures at 1 kilometer resolution.[10]

Figure 8 displays the predicted values from a bivariate regression of whether the government uses repression on (a) the number of participants in a social conflict and (b) logged population density. As expected, both plots suggest that small events and events in sparsely populated areas are least likely to be repressed. The bivariate relationship between logged population density and the probability of repression appears to be positive and approximately linear (i.e., the non-parametric estimate closely follows the linear fit). By contrast, plotting the probability of repression against categories for the

[10] I use a bilinear interpolation, which uses the four raster cells nearest to each social conflict to calculate population density (Hijmans and van Etten). The decision to interpolate is not consequential: simply extracting the cell value for raster cell in which a social conflict occurs yields a nearly identical measure (ρ(Interpolated, Simple) = 0.98).



(a) H1: Pr(Rep.) Increasing in # Participants (b) H1: Pr(Rep.) Increasing in Log Pop. Density

Figure 8: (a) Points correspond to the mean of $\mathbb{1}$(Repress) (and associated 95% CI) for each category of the number of participants variable. The final category of the number of participants variable is omitted, as its CI spans [0,1]. (b) Points represent events and can take values of zero or one on the y-axis. Random noise is added to provide a sense for the density of observations. The solid lines plot the fitted values from a least squares regression with the associated 95% CI. The dashed lines plot the fitted values from a non-parametric (loess) regression with a span of 0.75.

number of participants reveals unexpected non-linearity: the probably of repression appears highest for events involving between 100 and 10,000 individuals. A couple of points are worth noting: first, over 90 percent of the sample falls in the first four categories of the number of participants variable, and over this range the probability of repression appears to be roughly increasing in the number of participants. Second, as noted above, events involving repression are more likely to have data on the number of participants. I expect that the relationship between missingness and repression is strongest for smaller events, as the scale of large protests is likely to get reported regardless of whether repression is used. Hence, the averages in figure 8(a) likely overstate the use of repression in response to smaller events – a bias that would contribute to the observed non-linearity.

While suggestive, the bivariate relationships displayed in figure 8 could be driven by other omitted variables related to the countries or time period in which these social conflicts occur or differences in the event characteristics described in figure 6. The least squares regressions presented in table 2 try to account for these other explanatory variables. Models 1 and 3 add a full set of country and year fixed effects to the regressions that produced figure 8. The country fixed effects account for time-invariant features of each country (e.g., geography), while the year fixed effects soak up variation attributable to the year in which an event takes place. Leveraging the remaining within-country variation, I find further evidence in support of the first hypothesis: small social conflicts or social conflicts that occur in sparsely populated areas are significantly less likely to be repressed. Moving from the 25th to the 75th percentile of logged population density increases the probability of repression by seven percent – roughly 27 percent of the mean of the DV or the difference in the probability of repression in South Africa (0.14) vs. Lesotho (0.22).

These estimates appear robust: even after adding regressors for whether an event is violent, organized, targeted the government, or organized around a specific type of grievance, the point estimates on both the number of participants and logged population density remain stable. Model 5 is a notable exception: when I enter both logged population density and number of participants variables, the latter remains stable and highly significant, but the coefficient on logged population density decreases in size and is less precisely estimated. This is not unexpected, as population density is being used as a proxy for the size of protests, and its effect should attenuate when the number of participants is entered directly. In the final two models, I substitute ethnic-homeland for the country fixed effects to ameliorate concerns that

$\mathbb{1}(\text{Repression}) = \beta \text{ IV } + \boldsymbol{\gamma X} + \varepsilon$

DV: $\mathbb{1}$(Repression)	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
(Intercept)	0.250^{**}	0.045	0.181^{**}	0.176^{**}	-0.038	0.055	0.072
	(0.096)	(0.099)	(0.056)	(0.061)	(0.109)	(0.134)	(0.081)
Participants	0.198^{***}	0.227^{***}			0.229^{***}	0.240^{***}	
$[10, 100]$	(0.036)	(0.032)			(0.031)	(0.038)	
Participants	0.268^{***}	0.293^{***}			0.294^{***}	0.291^{***}	
$(100, 1000]$	(0.027)	(0.022)			(0.022)	(0.032)	
Participants	0.216^{***}	0.270^{***}			0.269^{***}	0.251^{***}	
$(1000, 10^4]$	(0.026)	(0.028)			(0.027)	(0.033)	
Participants	0.103^{***}	0.185^{***}			0.186^{***}	0.179^{***}	
$(10^4, 10^5]$	(0.031)	(0.041)			(0.041)	(0.047)	
Participants	0.087	0.154^{**}			0.151^{*}	0.140^{*}	
$(10^5, 10^6]$	(0.056)	(0.060)			(0.059)	(0.069)	
Participants	0.201	$0.211^{\cdot}$			$0.202^{\cdot}$	0.207	
$> 10^6$	(0.204)	(0.123)			(0.118)	(0.133)	
Log(Pop. Density)			0.017^{***}	0.011^{***}	0.008^{*}		0.014^{**}
			(0.003)	(0.002)	(0.004)		(0.004)
$\mathbb{1}$(Organized)		-0.096^{***}		-0.183^{***}	-0.093^{***}	-0.091^{***}	-0.163^{***}
		(0.021)		(0.016)	(0.020)	(0.025)	(0.020)
$\mathbb{1}$(Violent)		0.246^{***}		0.107^{***}	0.258^{***}	0.272^{***}	0.115^{***}
		(0.036)		(0.025)	(0.034)	(0.040)	(0.028)
$\mathbb{1}$(Issue:Econ)		-0.001		-0.011	0.004	0.012	-0.002
		(0.037)		(0.022)	(0.036)	(0.035)	(0.022)
$\mathbb{1}$(Issue:Ethnic/Rel.)		-0.004		-0.024	-0.003	-0.014	-0.018
		(0.040)		(0.019)	(0.042)	(0.042)	(0.024)
$\mathbb{1}$(Issue:Other)		-0.027		-0.014	-0.024	-0.045	0.016
		(0.042)		(0.026)	(0.042)	(0.044)	(0.027)
$\mathbb{1}$(Issue:Unknown)		-0.142^{**}		-0.093^{***}	-0.141^{**}	-0.147^{*}	-0.093^{***}
		(0.053)		(0.023)	(0.053)	(0.065)	(0.026)
$\mathbb{1}$(Gov. Target)		0.190^{***}		0.196^{***}	0.189^{***}	0.190^{***}	0.203^{***}
		(0.029)		(0.017)	(0.029)	(0.034)	(0.022)
Country + Year FE	✓	✓	✓	✓	✓		
Ethnic Group + Year FE						✓	✓
Adj. R^2	0.071	0.148	0.086	0.173	0.150	0.147	0.170
Num. obs.	2901	2821	6962	6640	2817	2075	4577
Num. regresors	74	81	70	77	82	134	169
Mean(DV)	0.32	0.33	0.26	0.27	0.33	0.34	0.29

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\cdot}p < 0.1$

Robust standard errors are clustered on country for models 1-5 and ethnic group for models 6-7.

Table 2: Linear Probability Models to Evaluate H1

specific ethnic groups are being targeted for repression.[11] In these latter two models, standard errors are clustered on ethnic homeland; otherwise, I cluster the standard errors on country.

6.2 H2: Pr(Lethal | Repression) Decreasing in Population Density

Looking at the bivariate regression plot in figure 9, both the linear and non-parametric regression lines suggest that the conditional probability of lethal repression is decreasing in population density, which is supportive of H2. If executives fear that lethal repression can backfire in more densely populated urban areas, they are more likely to opt for non-lethal means of putting down protests.

Yet, the increased likelihood of lethal repression in less densely populated rural areas could be driven by variables omitted from this simple model. Social conflicts or protesters may differ systematically across cities and rural areas (figure 6 suggests as much), and these sources of variation could account for the observed pattern. I attempt to account for these omitted variables in table 3. Models 1 and

[11] The shapefile provided by (Weidmann et al., 2010) allows me to place social conflicts from the SCAD in ethnic groups' territories. In order to complete this spatial join, I was forced to drop three duplicate polygons: Toubou (Chad) pre-1973, Berbers (Morocco) pre-1977, and Luba Kasai (DRC) pre-1994.

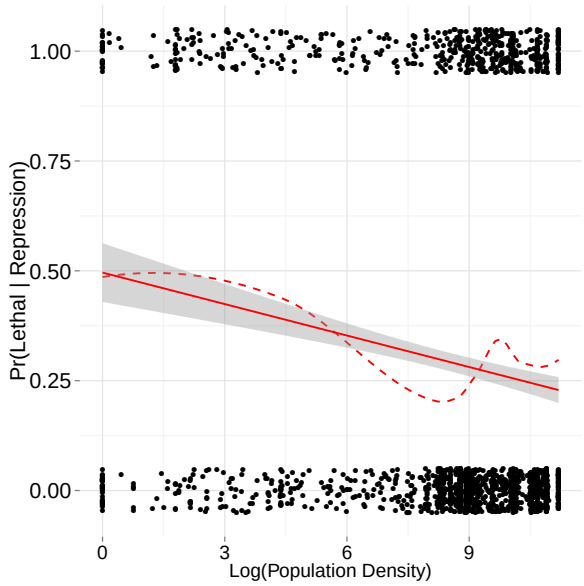


Figure 9: This figure was created using the same method as Figure 8. The y-axis now reflects whether lethal repression occurred, and the sample has been reduced to only those events that included some form of repression.

$\mathbb{1}(\text{Lethal} \mid \text{Repression}) = \delta \text{Log(Pop. Density)} + \boldsymbol{\gamma X} + \varepsilon$

DV: $\mathbb{1}$(Lethal | Repression)	Model 1	Model 2	Model 3	Model 4	Model 5
Log(Pop. Density)	-0.031^{***}	-0.028^{***}	-0.026^{***}	-0.025^{***}	-0.025^{***}
	(0.005)	(0.005)	(0.007)	(0.007)	(0.007)
$\mathbb{1}$(Organized)		$0.041^{\cdot}$		0.016	0.016
		(0.021)		(0.027)	(0.027)
$\mathbb{1}$(Violent)		0.125^{***}		0.105^{***}	0.104^{***}
		(0.028)		(0.030)	(0.030)
$\mathbb{1}$(Issue:Econ)		0.003		0.005	0.005
		(0.031)		(0.033)	(0.033)
$\mathbb{1}$(Issue:Ethnic/Rel.)		0.037		0.022	0.022
		(0.031)		(0.038)	(0.039)
$\mathbb{1}$(Issue:Other)		0.036		0.041	0.043
		(0.032)		(0.045)	(0.045)
$\mathbb{1}$(Issue:Unknown)		$0.133^{\cdot}$		$0.154^{\cdot}$	$0.158^{\cdot}$
		(0.071)		(0.080)	(0.081)
$\mathbb{1}$(Gov. Target)		0.161^{***}		0.170^{***}	0.172^{***}
		(0.034)		(0.033)	(0.032)
(Intercept)		$0.041^{\cdot}$		0.016	0.016
		(0.021)		(0.027)	(0.027)
Country + Year FE	✓	✓			✓
Ethnic Group + Year FE			✓	✓	✓
Adj. R^2	0.120	0.150	0.160	0.188	0.187
Num. obs.	1818	1808	1359	1350	1350
Num. Regressors	69	76	125	132	134
Mean(DV)	0.29	0.30	0.26	0.26	0.26

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\cdot}p < 0.1$
Robust standard errors are clustered on country in models 1-2 and ethnic group in models 3-5.

Table 3: Linear Probability Models to Evaluate H2.

2 enter country and year fixed effects, with the latter also including the available event characteristics. Models 3-5 include ethnic homeland and year fixed effects, with the final specification also including the country fixed effects that are not perfectly collinear with ethnic homeland. This final specification

leverages variation within-ethnic groups' territories and conditions on the protesters' tactics and primary grievance. Across all specifications the point estimate for population density remains between -0.025 and -0.031 and highly significant. Based on the estimates from model 1, moving from the 25th to 75th quantiles of Log(Population Density) is associated with a 6 percent decrease in the conditional probability of lethal repression, about 20 percent of the mean of the DV.

The results from table 3 lend support to the hypothesis that low density is associated with an increased probability that repression is lethal when it does occur. While one can not draw causal inferences from this analysis, the relationship between population density and the conditional probability of lethal repression appears robust: including additional covariates to account for event characteristics or heterogeneity attributable to country or ethnic homeland does not shift the point estimate or diminish its significance.[12]

6.3 Addressing Alternative Explanations

Sanctioning by International Actors

Anecdotally, INGO workers and foreign diplomats spend much of their time in major cities and, thus, receive less information about what is happening in the countryside. If true and governments believe that the detection of lethal repression by these actors can lead to sanctions, then this uneven monitoring across urban and rural areas could contribute to a lower conditional probability of lethal repression in urban areas (consistent with figure 1). Unfortunately, I am not able to measure this variation in international monitoring across countries' territories. I rely instead on a country-level indicator of the potential costs of international sanctioning, foreign assistance as a percentage of gross national income. The potential costs of international sanctions in response to (lethal) repression should be higher in those countries where foreign aid flows comprise a larger share of the economy.

If these arguments are correct, then we should expect increased aid dependence to be associated with a lower (lethal) probability of repression in response to social conflicts? As an independent variable, I use time-series data from the World Development Indicators on net official development assistance as a percentage of gross national income (GNI) (Bank, 2012). I aggregate the social conflicts in the base sample to the country-year level, calculating the share of events that were repressed in a given country-year over the total number of events in that country-year. As a basic specification, I regress the proportion of events that are repressed (model 1) or repressed lethally (model 2) on development assistance and country and year fixed effects. The right-most column of table 4 uses the conditional probability of lethal repression as the dependent variable.[13]

As is apparent in the first row of table 4, the correlation between development assistance and the frequency of repression is negligible: the point estimates are small and insignificant across the models. Even if the data is pooled (i.e., the fixed effects are removed), there is no discernible relationship between aid dependence and the probability that an executive opts for repression in response to a social conflict. Overall, there does not appear to be any evidence that countries more dependent on aid engage in less repression, suggesting that a fear of international sanctioning (in the form of decreased aid flows) does not strongly affect how executives respond to protests and riots.

Regime

Regime type is a robust predictor of levels of repression in the cross-national literature. To guard against the possibility that variation across leaders drives the results reported above, I extend the Archigos dataset through 2011 for all countries in the sample (Gleditsch and Chiozza, 2009). I substitute leader fixed effects for the country fixed effects and reestimate the models in tables 2 and 3. These specifications then leverages within-leader variation. Employing the full set of controls for event characteristics and year fixed effects, I find that the coefficient on Log(Population Density) is 0.011 when used to predict repression (identical to table 2 model 4) and -0.034 when used to predict the conditional probability of lethal repression (slightly stronger than the effect reported in table 3 model 2). The effects remain significant in both cases, suggesting that idiosyncratic variation in the repressive strategies of certain African leaders does not account for the effects reported above.

[12] As a further check of robustness, I reran model 1 with a country, year and country×year indicators and the coefficient remains the same magnitude and highly significant.

[13] If no event occurs in a country-year, I can either code Pr(Repression) as missing or assign it zero. I have tried both, and (fortunately) this coding decision has no impact on the results presented in table 4.

$\Pr(\text{Repression}) = \eta\text{ODA} + \boldsymbol{\gamma X} + \varepsilon$

DV:	Pr(Repression)	Pr(Lethal Rep.)	Pr(Lethal | Rep.)
ODA (% GNI)	0.000	−0.001	−0.001
	(0.001)	(0.001)	(0.001)
(Intercept)	$0.158^{\cdot}$	0.167^{**}	0.117
	(0.083)	(0.053)	(0.093)
Country + Year FE	Yes	Yes	Yes
Adj. R^2	0.093	0.034	0.097
Num. obs.	932	932	932
Num. Regressors	67	67	67

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\cdot}p < 0.1$

Table 4: Unlike prior tables, the unit of analysis in these regressions is the country-year. The fixed effects specifications leverage within-country variation in official development assistance as a percentage of gross national income.

History of Armed Conflict

While the SCAD explicitly excludes events associated with armed conflicts, it could be that rural social conflicts are more likely to occur in areas with histories of armed conflict. The state's greater willingness to deploy lethal force in response to these events (when they do violently intervene) could then be explained by past armed confrontations between the state and insurgents in the region. Fortunately, geocoded data on over 24,000 events associated with armed conflicts in Africa (defined as over conflicts resulting in over 25 deaths in any year in the data) are available from 1989-2010 from UCDP (UCD).[14] With this data, I am able to measure the number of armed conflict events that precede each event in the SCAD, occur in the same country, and take place within a given radius (e.g., 50 or 100 km) of the social conflict. The data, however, do not suggest that armed conflict is much more prevalent near social conflicts that occur in rural areas: the correlation coefficient between the two variables is only 0.09.[15] The small and positive correlation coefficient alleviates concern that rural SCAD events are low-level continuations of armed conflicts. Given the very small positive correlation between Log(Population Density) and the number of armed conflicts within 100 km of a social conflict, the results presented in tables 2 and 3 are not affected by the inclusion of this new variable.

Proximity to Natural Resources

Perhaps it is not a history of armed conflict, but rather the natural resource wealth of a region that affects whether and how the executive represses social conflicts. Regimes may respond particularly swiftly or harshly to protests or riots that threaten the continued extraction of valuable natural resources. We might then expect social conflicts occurring near these resources to be repressed more frequently or lethally. If this claim is correct and the occurrence of these natural resources is also correlated with population density, then we should be concerned about omitted variable bias in the earlier estimates.

Walter (2006), however, provides good theoretical reasons to be skeptical of this claim: Even if regimes are especially concerned about conflict in resource-rich regions, they do not accommodate challengers in resource-poor regions for fear of developing a reputation for weakness and emboldening potential future challengers. Nonetheless, I employ data on diamond occurrences (i.e., sites of production or confirmed discovery) and oil and gas fields (as of 2003) to measure the minimum distance between each event in

[14] The UCDP defines an event as, "The incidence of the use of armed force by an organised (sic) actor against another organized actor, or against civilians, resulting in at least 1 direct death in either the best, low or high estimate categories at a specific location and for a specific temporal duration." To be included, the event must be part of a conflict between two actors that crosses the 25 death threshold in any year of the UCDP data (Melander and Sundberg, 2012, p. 8). Given the limited temporal scope of the geocoded data, it is not possible to evaluate the effects of armed conflicts pre-1989.

[15] If we reduce the radius to 50 km, the correlation coefficient increases slightly to 0.12; expanding the radius to 200 km reduces the correlation coefficient to 0.04. The results are robust to the use of any of these three radii.

the SCAD and known deposits of these resources (Gilmore et al., 2005; Lujala et al., 2007). In these distance calculations, I only include resource deposits within the same country as the social conflict. As a result, a minimum distance can not be computed for events that occur in countries without diamond or oil deposits.[16]

The associations between population density and the frequency and severity of repression are not affected by the inclusion of variables measuring the minimum distance (in kilometers) between an event and diamond occurrences, oil and gas fields, or either natural resource. The coefficients reported in tables 2 and 3 remain stable despite the inclusion of these measures for proximity to natural resources. I find no evidence that the earlier empirical tests were confounded by failing to include measures of how close a social conflict occurs to a diamond or oil deposit.

6.4 Further Robustness Checks

Dropping Capital Cities. To ensure that the results are not driven by events in capital cities, I rerun earlier models, excluding events that take place in a capital city. While this drops 2,486 events from the base sample, the coefficient estimates on logged population density remain stable and highly significant: when $\mathbb{1}$(Repression) is the dependent variable, the coefficient estimate is 0.013 (s.e. = 0.003); when $\mathbb{1}$(Lethal | Repression) is the dependent variable, the coefficient decreases in magnitude slightly to -0.018 (s.e. = 0.006). These regressions are otherwise identical to table 2 model 4 and table 3 model 2, respectively.

Dropping Individual Countries. I also exclude each country individually and rerun the analysis. The coefficient estimates on logged population density remain stable and highly significant: when $\mathbb{1}$(Repression) is the dependent variable, the minimum coefficient estimate is 0.01 (s.e. = 0.002) and the maximum is 0.012 (s.e. = 0.002); when $\mathbb{1}$(Lethal | Repression) is the dependent variable, the minimum coefficient is -0.030 (s.e. = 0.005) and the maximum is -0.025 (s.e. = 0.005). Again, besides reducing the sample, these regressions are otherwise identical to table 2 model 4 and table 3 model 2.

Dropping Events with Less Precise Geocodings. The SCAD includes an indicator for whether an event could be precisely geocoded or whether its location is estimated by the coders. Fortunately, the vast majority of SCAD events received precise codings based on place names. However, to be safe, I drop all events with an estimated geocoding and rerun table 2 model 4 and table 3 model 2. The results remain unchanged.

6.5 Reporting Bias

A final confounder relates to the construction of the dataset. If newswires underreport non-lethal repression in rural areas relative to lethal repression, this could partially account for the observed difference in the conditional probability of lethal repression across urban and rural areas. It is possible to bound any such reporting bias – that is, to determine how large the bias would have to be to account for the observed difference in the conditional probability of lethal repression across urban and rural areas reported in figure 1. To calculate this bound, I first make the most penalizing assumptions:

A1: All events in urban areas are reported with probability 1.

A2: All events involving lethal repression are reported with probability 1 regardless of location.

A3: There is no difference in the probability of *non*-lethal repression across urban and rural areas given that some repression occurs. That is, for the subset of events involving repression, Pr(Non-lethal Rep.|urban)=Pr(Non-lethal Rep.|rural).

[16] While the diamonds dataset is comprise of geolocated points, the oil and gas fields are mapped as polygons. I measure the shortest great-circle distance between each event and the boundaries of these polygons using the `geosphere` package in `R` (Hijmans et al.).

I then solve for the number of events involving non-lethal repression that would have to have occurred in rural areas to satisfy this equality.

$$\begin{aligned} \Pr(\text{Non-lethal}|\text{rural}) &= \Pr(\text{Non-lethal}|\text{urban}) \\ \frac{x}{137+x} &= \frac{1009}{1331} \\ x &= 429 \end{aligned}$$

If there were no difference in the conditional probability of lethal repression across urban and rural areas, then we should observe 429 events in rural areas involving non-lethal repression. Yet, we only observe 167 in the sample.

First, This level of reporting bias implies that the newswires miss over 60 percent of the rural events in which the state violently represses protesters but no participant dies. Recall that the threshold for "urban" in the SCAD is 100,000 residents. This means that a number of "rural" localities are well-sized towns and small cities in which this level of underreporting seems unlikely. Second, if events involving no repression suffer from the same reporting bias, this would imply that a nearly equal number of social conflicts "actually" occurred in urban and rural areas. Given what we know about the relationship between population density and the costs of collective action, it is difficult to believe that the incidence of social conflict is unrelated to population. Third, if both non-lethally repressed and unrepressed events are underreported at equal rates, the probability of repression in response to rural events would drop even further. Put differently, if this reporting bias exists, it would strengthen the first pattern noted in the paper – namely that the unconditional probability of repression is higher in urban areas. Finally, if we relax either of the first two assumptions and allow some lethally repressed events in rural areas to be missed or social conflicts in urban areas to be unreported (both of which occur), then the reporting bias would have to be even larger still to account for the observed patterns.

We can also partially address concerns about reporting bias by looking only at the subset of rural events. If, among these events, we still observe the hypothesized relationships, we can be more confident that reporters' reluctance to cover rural events does not drive the results reported above. Table 5 replicates the first two models from tables 2 and 3 using only events in the base sample that occurred outside of urban areas (defined as localities with more than 100,000 residents). These results are consistent with the earlier findings: the coefficients all take signs consistent with the first three hypotheses and are of the same order of magnitude; the coefficients on population density are significant when regressed on the incidence of repression (top panel). However, population density loses significance when regressed on the incidence of lethal repression among events involving some form of repression. This largely results from a lack of statistical power: the regressions in the bottom panel of table 5 only include 300 observations but around 70 regressors. This represents only 12 percent of the observations in models 1 and 2 from table 3.

Finally, if the proposed reporting bias exists, we might expect reporters to neglect relatively tranquil rural events when the opportunity cost of covering such events increases. If reporting resources are scarce (which is the case throughout most of Africa), then reporters may reduce their coverage of rural protests that do not involve lethal repression when other major news stories demand their attention. From April to July 1994, over a half million people were brutally slaughtered in the Rwanda Genocide. This tragedy captured the world's attention: the number of AP newswire stories filed in Rwanda increased over 11-fold between 1993 and 1994. In order to devote adequate attention to what was occurring in Rwanda, newswires may have reduced their coverage in the countries bordering Rwanda (Burundi, Democratic Republic of Congo, Tanzania, and Uganda). This reduction would only be a concern for the present paper if it disproportionately affected rural areas. If that were the case, it would suggest that the coverage of rural events is more sensitive to changes in the costs of reporting than urban events.

Figure 10 plots the number of social conflicts *not* involving lethal repression that occurred in the countries bordering Rwanda between 1992 and 1997 (aggregated by month). While there does appear to be a drop in the number of such social conflicts that occurred during the months of the genocide, this drop is actually more pronounced in urban areas. Moreover, the ratio of events involving no or

DV: 𝟙(Repression)	Model 1	Model 2
Log(Pop. Density)	0.015^{***}	0.009^{*}
	(0.004)	(0.004)
Country + Year FE	Yes	Yes
Event Characteristics	No	Yes
Adj. R^2	0.174	0.304
Num. obs.	1551	1481
Num. Regressors	67	74
DV: 𝟙(Lethal — Repression)	Model 1	Model 1
Log(Pop. Density	−0.014	−0.012
	(0.016)	(0.015)
Country + Year FE	Yes	Yes
Event Characteristics	No	Yes
Adj. R^2	0.136	0.228
Num. obs.	301	299
Num. Regresors	63	70

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\cdot}p < 0.1$
Robust std. errors clustered on country.

Table 5: The third and fourth models from tables 2 and first and second models 3 are replicated using only events from the base sample that occurred in rural areas (defined by the SCAD as localities with less than 100,000 people).

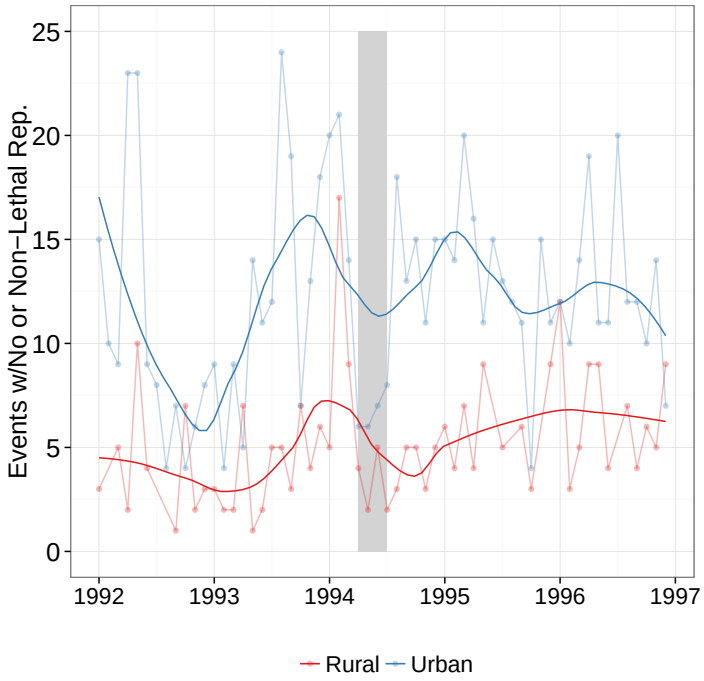


Figure 10: The y-axis measures the number of social conflicts that occurred in Burundi, Democratic Republic of Congo, Tanzania, and Uganda (i.e., the countries that border Rwanda) that did not involve lethal repression. These events are aggregated by month. The smoothed darker lines represent the fitted values from a loess regression with a span of 0.4. (I choose the reduced span to allow for the detection of sharper changes in trends.) The lighter lines simply connect the monthly counts, and the grey rectangle corresponds to the months of April to July 1994, during which the Rwandan Genocide occurred.

non-lethal repression in rural and urban areas appears relatively stable despite the likely diversion of reporting resources to Rwanda. The data in this particular case do not suggest that newswires neglect non-lethal social conflicts in rural areas when their available reporting resources become limited.

Unfortunately, no dataset exists that captures the full universe of social conflicts that have occurred in Africa over the last two decades. However, these checks suggest that the best available resource, the SCAD, does not suffer from reporting biases that would confound the statistical inferences drawn earlier in the paper.

7 Conclusion

Past studies of repression have relied heavily on cross-national comparisons and, thus, missed subnational variation in how leaders respond to social conflicts. This paper demonstrates that this subnational variation is systematic: the probability or repression is increasing in population density, while the conditional probability of lethal repression is decreasing in population density. These findings are consistent with a theory, in which initial protests in urban areas pose a greater threat to the executive but are also more likely to escalate if confronted with lethal repression.

The findings are not driven by observed characteristics of the event and are robust to ethnic homeland, country, and year fixed effects. Further checks do not suggest that regime, a history of armed conflict in a particular region, or other plausible omitted variables account for the effects. The observed differences in the handling of social conflicts across urban and rural areas are substantively meaningful: a one standard deviation increase in logged population density is associated with a roughly five point increase in the probability of repression but a ten point decline in the conditional probability of lethal repression.

Moreover, the theory and results help shed light on an ongoing debate in the literature as to whether repression does or does not incite a backlash. If, as I argue in this paper, governments are factoring the expected cost of a backlash into their decisions regarding how severely to repress protests, then we are less likely to observe a correlation between repression and escalation.

References

UCDP Georeferenced Event Dataset.

AFP. Kenyan university closed as students protest lack of teachers. *Agence France Presse*, 2004a.

AFP. Kenyan university closed after student riot over planned fee hike. *Agence France Presse*, 2004b.

AFP. One killed as students, villagers battle private developers in Kenya. *Agence France Presse*, 2005.

AFP. Hike in World Prices Sparks Deadly Riots in Africa. 2008.

Leonardo Arriola. Suppressing Protest: The Geographic Logic of Mass Arrests. *WGAPE Working Paper*, 2012.

The World Bank. World Development Indicators. 2012.

Robert H Bates. *Markets and States in Tropical Africa: The Political Basis of Agricultural Policies*. University of California Press, 1981.

A K Bohara, N J Mitchell, and M Nepal. Opportunity, democracy, and the exchange of political violence. *Journal of Conflict Resolution*, 2006.

Christian Davenport. Multi-Dimensional Threat Perception and State Repression: An Inquiry into Why States Apply Negative Sanctions. *American Journal of Political Science*, 1995.

Christian Davenport. State Repression and Political Order. *Annual Review of Political Science*, 2007.

Department of Peace and Conflict Research. Definition of Armed Conflict.

Jennifer Earl and Sarah A Soule. Seeing Blue: A Police-Centered Explanation of Protest Policing. *Mobilization: An International Journal*, 2006.

Elisabeth Gilmore, Nils Gleditsch, Päivi Lujala, and Jan RØd. Conflict Diamonds: A New Dataset. *Conflict Management and Peace Science*, 2005.

Kristian S Gleditsch and Giacomo Chiozza. Introducing Archigos: A Data Set of Political Leaders. *Journal of Peace Research*, 2009.

Jack A Goldstone. Toward a Fourth Generation of Revolutionary Theory. *Annual Review of Political Science*, 2001.

International Crisis Group. World Directory of Minorities and Indigenous Peoples - Cameroon: Overview. 2007.

E M Hafner-Burton. Right or robust? The Sensitive Nature of Repression to Globalization. *Journal of Peace Research*, 2005.

Tristen Hayfield and Jeffrey Racine. The np Package.

Cullen S Hendrix and Idean Salehyan. SCAD 3.0 Codebook. 2012.

Cullen S Hendrix, Idean Salehyan, Christina Case, Christopher Linebarger, Emily Stull, and Jennifer Williams. The Social Conflict in Africa Database: New Data and Applications. *Working Paper, University of North Texas*, 2012.

Robert J Hijmans and Jacob van Etten. Raster: Geographic Data Analysis and Modeling.

Robert J Hijmans, Ed Williams, and Chris Vennes. Geosphere: Spherical Trigonometry.

International Crisis Group ICG. Cameroon: The Dangers of a Fracturing Regime. *Africa Report*, 2010.

Mark Lichbach. An economic theory of governability: Choosing policy and optimizing performance. *Public Choice*, 1984.

Päivi Lujala, Jan Ketil Rod, and Nadja Thieme. Fighting over Oil: Introducing a New Dataset. *Conflict Management and Peace Science*, 2007.

Erik Melander and Ralph Sundberg. UCDP Georeferenced Event Dataset (GED) Codebook. 2012.

F Murtagh, M Spagat, and J A Restrepo. Ultrametric wavelet regression of multivariate time series: Application to Colombian conflict analysis. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, 2009.

National Human Rights Observer. Cameroon: A Discreet and Bloody Crackdown.

Oak Ridge National Laboratory. LandScan. 2010.

David G Ortiz. Confronting Oppression with Violence: Inequality, Military Infrastructure and Dissident Repression. *Mobilization: An International Journal*, 2007.

K Rasler. Concessions, repression, and political protest in the Iranian revolution. *American Sociological Review*, 1996.

Peter C Reiss. Empirical Work in Economic Research.

K C Vadlamannati. Socioeconomic, Institutional and Political Determinants of Human Rights Abuse: A Subnational Study of India, 1993-2002. *William Davidson Institute Working Paper*, 2008.

Barbara F Walter. Building reputation: Why governments fight some separatists but not others. *American Journal of Political Science*, 2006.

Nils Weidmann, Jan Ketil Rod, and Lars-Erik Cederman. Representing Ethnic Groups in Space: A New Dataset. *Journal of Peace Research*, 2010.