Predicting Local Violence: Evidence from a Panel Survey in Liberia

Robert A. Blair* Christopher Blattman† Alexandra Hartman‡

keywords: forecasting, violence, machine learning, surveys, Africa

*Corresponding author. Department of Political Science and Watson Institute for International and Public Affairs, Brown University, robert_blair@brown.edu

†Harris School of Public Policy, University of Chicago

‡Department of Political Science, University College London

Abstract

Riots, murders, lynchings and other forms of local violence are costly to security forces and society at large. Identifying risk factors and forecasting where local violence is most likely to occur should help allocate scarce peacekeeping and policing resources. Most forecasting exercises of this kind rely on structural or event data, but these have many limitations in the poorest and most war-torn states, where the need for prediction is arguably most urgent. We adopt an alternative approach, applying machine learning techniques to original panel survey data from Liberia to predict collective, interpersonal and extrajudicial violence two years into the future. We first train our models to predict 2010 local violence using 2008 risk factors, then generate forecasts for 2012 before collecting new data. Our models achieve out-of-sample AUCs ranging from 0.65 to 0.74, depending on our specification of the dependent variable. The models also draw our attention to risk factors different from those typically emphasized in studies aimed at causal inference alone. For example, we find that while ethnic heterogeneity and polarization are reliable predictors of local violence, adverse economic shocks are not. Surprisingly, we also find that the risk of local violence is higher rather than lower in communities where minority and majority ethnic groups share power. These counterintuitive results illustrate the usefulness of prediction for generating new stylized facts for future research to explain. Ours is one of just two attempts to forecast local violence using survey data, and we conclude by discussing how our approach can be replicated and extended as similar datasets proliferate.

Introduction

Riots, murders, lynchings and other forms of local violence are an urgent concern for police and peacekeepers, especially in weak and war-torn states. Local violence is more common, and possibly even more costly, than war- or terrorism-related violence (?). Local violence can also shape, and be shaped by, national conflict dynamics (?). Resources for prevention are often scarce, and any information that helps identify risk factors and predict where local violence is most likely to occur should have large practical and, potentially, theoretical returns.

Many studies have attempted to predict national-level conflicts—e.g. civil war, political instability (?) or "irregular regime change" (?).[1] Recently, however, the most active frontier of conflict research has focused on local-level incidents, including murders (?), riots (?), domestic violence (?), "low-intensity" sectarian clashes (?), conflicts between states and "excluded, downgraded and underrepresented groups" (?), and killings of suspected witches (?). Like its cross-national counterpart, most sub-national conflict research has been descriptive or causal. Forecasting has been relatively rare.

We complement and extend this literature by combining original survey data from Liberia with existing machine learning techniques to test the feasibility of predicting violence at a lower unit of analysis (the town or village level) than is typically feasible in exercises of this kind. In the process we gauge the relative predictive power of 56 potential risk factors for local violence, identifying relationships that might be used to validate and refine existing theoretical models of conflict, or to suggest new puzzles and stylized facts for future research to assess.

Our dataset covers 242 towns and villages in Liberia in 2008, 2010 and 2012. We used surveys of both residents and leaders to measure three categories of local violence (collective, interpersonal, and extrajudicial) as well as dozens of demographic, social, political and economic predictors. We then used the 2008 data to predict local violence in 2010, simulating

[1] For a literature review see ?.

forecasts using cross-validation. We relied on three common machine learning techniques in particular: lasso, random forests and neural networks. We also included logit for purposes of comparison. Finally, we locked in the parameters of our models, generated forecasts for 2012, and collected new data to compare our predictions to reality.

We highlight two sets of findings. First, our models perform considerably better than chance and are robust to changes in specification. This is especially true of lasso, our most parsimonious model. The area under the Receiver Operating Characteristic (ROC) curve is between 0.65 and 0.73 for the 2012 lasso forecast, depending on our coding of the dependent variable. Given the novelty of this exercise, we view these results as a promising first step. Moreover, the lasso model includes just 5 of our 56 risk factors, some of which are either slow-moving (e.g. ethnic heterogeneity) or time-invariant (e.g. land lost during the Liberian civil war). While we cannot be sure that these risk factors will continue to predict violence in the future, our results raise the possibility of updating the model's forecasts at relatively low cost in terms of new data collection. (One finds similar parsimony in some cross-national models, e.g. **?**.)

Second, the models draw our attention to risk factors different from those typically emphasized in studies aimed at causal inference alone. Many of these studies focus on the relationship between local (or national) violence and adverse economic shocks, and existing evidence suggests a strong causal relationship between shocks and crime (**?**), witch killings (**?**) and domestic violence (**?**), among other outcomes (see **?** for a review). Yet we find that shocks such as droughts, floods and pest infestations have little predictive power, at least in our sample. As others have noted (**?**), statistically significant causes of conflict often prove weak or unreliable predictors. Forecasting can help focus our attention on substantively as well as statistically significant risk factors—those that not only cause conflict, but also predict it.

Some of these risk factors are intuitive. For example, variables related to land loss or ethnic heterogeneity, polarization and fractionalization consistently predict local violence in our sample—an unsurprising result given that Liberia remains divided along ethnic lines, and given that disputes over property rights are common and prone to escalation, especially in rural areas (**?**). Other results are more unexpected. Perhaps most striking, the best predictor of local violence in our most reliable model (lasso) is an indicator for communities in which minority and majority ethnic groups share power. Counterintuitively, power-sharing predicts *higher* rather than lower levels of local violence. While this correlation is not evidence of causality, it is consistent with recent studies warning of the unintended consequences of power-sharing (e.g. **?**). We discuss possible interpretations and implications of this finding for future research, and include additional results and robustness checks in the online appendix, available at [URL].

As with any single case study, we cannot be sure how our results will generalize. Our goal is to be more foundational. To our knowledge, ours is one of just two studies that use survey data to forecast conflict (the other being **?**). Most researchers use either structural or event data instead (or, increasingly, some combination of the two, e.g. **???**). Unfortunately, these data are scarce or unreliable in weak and war-torn states, where the urgency of forecasting is arguably greatest, but where journalists and NGOs—the sources typically used to populate datasets of this kind—are often confined to large, relatively urban areas.

Our survey covers regions of Liberia that are systematically underrepresented in structural and event datasets. It also captures a wider range of incidents at a lower level of aggregation and with higher reliability that most event datasets, as we are able to "ground truth" our quantitative data through qualitative follow-up. Moreover, the survey features a large number of theoretically- and contextually-relevant predictors that are typically unavailable at this unit of analysis,[2] measured using questions that were extensively pretested for appropriateness and comprehensibility.

Of course, our dataset is not without limitations. The most obvious is the relatively small number of communities (242) and time periods (three) that it covers; another is the

[2]Though innovation in passive measurement should make data collection at this level increasingly reliable, even cross-nationally; see, e.g., **?** in this volume.

potential for misreporting of the dependent variable (even with qualitative follow-up), and for sampling error in our risk factors. While we address these issues below and in the online appendix, replication is needed in more settings, ideally at higher frequency in larger and longer panels. Fortunately, as datasets on local violence and its correlates continue to proliferate, opportunities for replication will proliferate as well. Leveraging these datasets for purposes of forecasting is a promising frontier, not just to facilitate conflict prevention and mitigation but also to identify theoretically interesting patterns for future research to explore.

Setting

Liberia is a small West African nation still recovering from two civil wars fought between 1989 and 2003. Four features of the setting are especially relevant to this study.

First, in many respects Liberia stabilized over the study period, 2008-12. It held reasonably free, competitive elections in 2005 and 2011, which brought to power a largely legitimate and professional regime. The state has been supported by ample aid and a large United Nations (UN) peacekeeping operation. Second, despite that stability, ethnic and religious cleavages remain deep. Liberia is home to fifteen indigenous ethnic groups (or âĂIJtribesâĂİ) as well as an Americo-Liberian elite that has historically held political power. Tensions between these groups are often implicated in incidents of local violence.

Third, while the prevalence of local violence declined over the years of the study (see the online appendix), it remains endemic. In 2010 alone, 10% of the towns in our sample reported a violent strike or protest or a violent confrontation between tribes; 15% reported a murder or rape; and 9% reported a lynching or trial by ordeal. Fourth and finally, the Liberian government's ability to prevent or mitigate these incidents is limited: the courts remain largely inaccessible, and the army and police are underfunded and ill-equipped. Meanwhile, the UN Mission in Liberia officially concluded its mandate in this year, and the country is struggling to restore economic growth after the devastating Ebola epidemic of 2014-15. This has increased the urgency of creating early warning systems to anticipate local violence before it occurs, and to prevent peaceful disputes from escalating into violent ones.

Data and measurement

We collected survey data from residents and leaders in 242 communities across three of LiberiaâĂŹs 15 counties: Lofa, Nimba and Grand Gedeh. Our unit of analysis is the smallest administrative unit in the country, whose size typically ranges from a few hundred to a few thousand residents. Most are villages or small towns. Fifty are neighborhoods, called "quarters," within larger towns, with their own quarter chiefs. For simplicity, we refer to all of these settlements as "towns." See the online appendix for maps and further details on sampling.

The first round of data was collected as part of randomized controlled trial evaluating a UNHCR-sponsored alternative dispute resolution program (?), and the towns in our sample are not representative of the three counties. Rather, they were selected because government officials believed them to be especially prone to disputes. Comparison to a nationally representative survey conducted at the same time suggests the towns in our sample were not, in fact, much more conflicted than the average Liberian community in these three counties, or nationwide (?). Nonetheless, we interpret our results as conditional on some minimum pre-existing level of risk.

In each town we surveyed a representative sample of roughly 20 randomly-selected residents and four purposively-selected leaders—typically a town chief, women's group leader, youth group leader, and minority ethnic group leader. We collected data in three rounds: March to April 2009, November 2010 to January 2011, and February to April 2013. We selected a new cross-section of residents each time, aggregating individual-level responses into town-level variables to construct a panel.

Dependent variable

We coded our dependent variable using the leaders survey, focusing on seven types of local violence in three categories:

Collective violence includes violent strikes or protests and violent confrontations between tribes. Strikes and protests are common forms of collective action in rural Liberia; not all turn violent, but some result in property destruction, fights or lynchings. If weapons are involved, they tend to be limited to sticks or machetes. Common proximate causes include ambiguities in property rights, elders' restrictions on youths, violations of ethnic or religious customs, and disagreements over political decisions or institutions.

Interpersonal violence includes murders, rapes and assaults with weapons (i.e. aggravated assault). Many of these crimes begin as disputes between families, or between men over a woman. Killings over land also occur, though with less frequency. Murders can also include cases of manslaughter where foul play is unproven but suspected.

Extrajudicial violence includes trial by ordeal and beatings or killings of suspected witches. Trial by ordeal is an informal mechanism of criminal adjudication or dispute resolution. While efficient (**?**), it is usually coercive and almost always illegal. The most common forms of trial by ordeal are "hot cutlass," whereby suspects are made to withstand a heated machete pressed against their skin, and "sassywood," whereby suspects are made to eat the bark of the poisonous sassywood tree (or ingest some other potion). Mobs sometimes beat or kill the accused, especially in cases of witchcraft.

Coding

Enumerators asked each of four leaders whether each of the seven types of local violence described above occurred in their community in the past 12 months. We then used the modal response across the four leaders to code an indicator for each type of local violence before aggregating by category (collective, interpersonal and extrajudicial) and overall (i.e. any local violence). This is a conservative coding rule: if different leaders classify the same event in different ways, it is possible that no incident will be recorded. We report descriptive statistics based on this coding rule in the online appendix, and examine alternatives below.

Validation

We validated our coding in 2010 and 2012 through qualitative follow-up. In 2012, enumerators recorded and transcribed leaders' accounts of all incidents of all types while conducting the survey. In 2010, enumerators returned to towns that reported at least one incident of collective violence a few months after the survey to investigate further. For simplicity and consistency, we do not recode our dependent variable based on qualitative follow-up for most of our analyses; rather, we use leaders' accounts to validate our approach to aggregation, and to better understand the dynamics of local violence in our sample. We do, however, test the robustness of our results to alternate coding rules. The online appendix discusses the qualitative follow-up in more detail.

Advantages and disadvantages of the data and coding

Most conflict forecasters use structural data, event data, or some combination of the two. Surveys allow us to capture incidents that these datasets generally do not, and to validate them through qualitative follow-up.[3] Moreover, many event datasets do not include geolocation below the country (or, at best, district) level, precluding local prediction. Those that do tend to rely on newspaper or NGO reports, which are often incomplete and biased, especially in the poorest and most war-torn countries, where media coverage is usually spotty and NGOs are often confined to capital cities and other large, relatively urban areas. Survey data also allows us to capture a wide variety of theoretically- and contextually-relevant

[3] **?** and **?** both stress that there are probably more incidents missing than present in event datasets, including ICEWS and GDELT, though comprehensiveness will likely improve over time. We discuss these advantages and disadvantages in detail in the online appendix.

predictors using extensively pretested questions.

Our approach has at least three limitations, however, which could cause either under- or over-reporting of local violence. First, we focus on a non-random sample of towns from a non-random sample of Liberian counties. These communities were purposively selected because they were believed to be at disproportionately high risk of conflict. This is disadvantageous in that it limits the generalizability of our results, but it is also potentially advantageous in that it focuses our attention on the areas of most immediate concern to police and peacekeepers. This is similar in spirit to cross-national models that focus on particularly "high-risk" regions—e.g. sub-Saharan Africa—or that drop or separately model countries believed to be "immune" from conflict (**?**). Furthermore, comparison to a nationally representative survey suggests that our towns were not much more conflicted than the average Liberian community (**?**). Even in our sample, the actual prevalence of violence reported in the survey was far from uniform, alleviating potential concerns about selection on the dependent variable. Finally, while most of our towns are small and rural, the sample also includes quarters of several larger cities, including Voinjama (the 6th largest city in Liberia) and Zwedru (the 8th largest).

Second, categorization of local violence was more ambiguous than we expected. In a setting like Liberia, a murder can provoke accusations of witchcraft, which can incite collective violence, which different leaders may interpret in different ways (e.g. as a violent confrontation between tribes, or as a violent strike or protest). We are not the first to note this ambiguity (see, e.g., **?**), which afflicts any approach to coding conflict, survey- or news-based or otherwise. Ambiguity motivates our decision to aggregate multiple categories of violence into a single indicator, and also motivates our exploration of alternative coding rules below.

Third and relatedly, leaders may disagree on what does and doesn't constitute "violence." For example, in our qualitative follow-up, leaders described two of the seven violent strikes or protests recorded in the 2012 survey in ways that suggested they may have been non-violent; a third incident was more ambiguous. Three of the 16 trials by ordeal may have been conducted semi-voluntarily, though they were still illegal, and still involved the threat of death or injury. And one of the five murders in 2012 appears to have been a case of manslaughter, though neither we nor the leaders know for sure. As we discuss below and in the online appendix, performance either improves or remains the same when we use alternative coding rules to correct for possible over- or under-reporting.

Predictors

We trained our models using 56 potential risk (and protective) factors for local violence, organized into the following categories:

Demographics includes estimates for the per capita population of men, youths, ex-combatants, foreigners and returned internally displaced persons (IDPs) in each town. Conventional wisdom has long portrayed young men, and especially ex-combatants, as the most likely participants in crime and collective violence (**?**). Conflict between "locals" and perceived "foreigners" (including returned IDPs and refugees) is common across sub-Saharan Africa (**?**), and has been a recurring source of instability in the counties we study.

Social capital and civic life includes proxies for membership in civil society organizations, contributions to public goods and expressions of self-reliance in social service provision. Social capital and civic life have been posited as protective factors against crime (**?**) and collective (especially ethnic) violence (**?**).

Quality of governance includes residents' perceptions of corruption in the state security and justice sectors, and of fairness and impartiality among the leaders of their towns. A large literature has found a positive correlation between grievances and collective violence (see **?** for a review); extrajudicial violence may also be more common where citizens distrust the police and courts (**?**).

Accessibility and social services includes an indicator for cell phone coverage, an index of facilities available in each town (e.g. wells, latrines, schools and clinics) and an estimate of the distance from each town to the nearest usable road. Accessibility increases the likelihood of public goods provision, and of third party intervention to prevent or mitigate violence. We include estimates for the frequency of police patrols and visits from NGO personnel as well.

Natural resources includes indicators for any rubber plantations or diamond, gold or iron mines within an hour's walk of each town. Cross-national studies have found a positive correlation between natural resources and the onset and duration of civil conflict (**?**), though it is unclear whether the mechanisms underlying this relationship manifest at the sub-national level as well. Even if they do not, proximity to natural resources may attract transient and potentially disruptive populations—ex-combatants, for example (**?**)—as well as multinational corporations, whose presence may provoke conflict with communities.

Ethnic heterogeneity, fractionalization and polarization includes estimates for the number of tribes cohabiting in each town; the proportion of residents belonging to the majority tribe in each town; and the proportion of residents expressing prejudice against other tribes (e.g. the belief that other tribes are "violent" or "dirty"). Ethnic fractionalization and polarization have long been viewed as risk factors for both local (e.g. **?**) and national violence (e.g. **?**), especially in the absence of strong state institutions. Cross-national studies have similarly found a correlation between ethnic heterogeneity and violent crime (**?**).

Employment, wealth and inequality includes an indicator for employment, an additive index of asset ownership, and an estimate for the variation in asset ownership across individuals within the same town (a proxy for inequality). While hotly debated, scholars have long argued that unemployment, poverty and inequality tend to correlate with crime and collective violence (see **?** and **?** for reviews), as well as with persecution of accused witches (?), among other outcomes.

Access to land includes estimates for the proportion of residents reporting an ongoing land dispute, the proportion of residents without access to land, and the proportion of residents who lost their land during the Liberian civil war. Access to land is a recurring driver of conflict across sub-Saharan Africa in general (?), and in Liberia specifically (?).

Adverse economic shocks includes estimates for the proportion of residents affected by crop failures (due to droughts, floods or pest infestations) or human or livestock diseases. Previous studies have found a causal effect of shocks of this sort on crime (?), domestic violence (?) and witch killings (?), among other outcomes (see ? for a review). More recent research has focused on testing the predictive power of climate change more generally (e.g. ? in this volume).

History of violence includes estimates for the proportion of residents that witnessed, participated in or were victims of violence during the Liberian civil war, and the proportion of residents that fled their homes. Intuitively, we might expect exposure to wartime violence to exacerbate the risk of peacetime violence, though recent studies have found that experiences of victimization may have counterintuitive salutary effects on cooperation, altruism and civic life (see ? for a review). Finally, we also include a lagged dependent variable in all models, as well as estimates for the proportion of residents reporting a burglary, robbery or simple assault in the previous year.

The online appendix provides descriptive statistics for all of the risk factors listed above.

Methods

Statisticians have developed a wide variety of tools for forecasting. We focus on three in particular: lasso, random forests and neural networks. These techniques have proven track records, and—consistent with our inductive approach—can accommodate many highly collinear regressors and non-linear or interactive relationships simultaneously. They also capture much of the variation across existing classes of machine learning models: selection and shrinkage techniques (lasso), ensemble and tree-based methods (random forests), and non-linear adaptive weighting algorithms (neural networks). For purposes of comparison, we include results from a logit model as well.

We describe our models briefly here, with additional details in the online appendix.

Models

Least absolute shrinkage and selection operator (lasso) is among the most widely-used variable selection techniques in statistics (**?**). It is analogous to logit, with the crucial difference that it penalizes model complexity, shrinking all coefficients and reducing some to zero. It thus discards predictors automatically, using a transparent, replicable selection mechanism. Lasso models also tend to be parsimonious, which is advantageous for avoiding multicollinearity and preventing over-fitting—problems that often afflict logit models, especially when the number of predictors is large and the dependent variable is rare—as well as for the more practical purpose of deciding which risk factors to track over time.

Random forests are collections of decision trees (**?**). A decision tree sorts observations into subgroups, or "nodes," by first identifying the risk factor that most accurately distinguishes positives (in our case, violence) from negatives (no violence), then identifying secondary and tertiary risk factors to further reduce mean-squared error within each node. Each observation is passed down the tree until it reaches a terminal node, at which point a prediction is made based on the modal predicted outcome at that node. Random forests are simply ensembles of decision trees grown within many random subsamples of the data using many random subsets of predictors.[4] The forest's prediction is the average prediction across

[4]In our case, each tree is fit to 24 randomly-selected observations (roughly 10% of the sample) using 7 randomly-selected predictors and a maximum of five terminal nodes. Each random forest is composed of

every tree, increasing stability.

Neural networks were inspired by the structure of the human brain, and have been applied (albeit sparingly) in political science in the past (e.g. **?**). They make no parametric assumptions, instead using an iterative algorithm to approximate the underlying structure of the data. The algorithm begins by constructing different nonlinear weighted sums of the available predictors, each called a node. It then uses that layer of weighted sums to generate another weighted sum, which maps onto the prediction space. In principle there can be many layers and nodes, but our model is more parsimonious, with one layer and 5 nodes. The weights are initially chosen at random, and then tuned iteratively to minimize mean-squared error.

Logit has the virtue of simplicity and familiarity, but has potential drawbacks for our purposes. While lasso, random forests and neural networks can accommodate many highly collinear predictors simultaneously, logit generally cannot, and the statistical consequences of attempting to do so can be severe. Unlike lasso, logit provides no systematic or transparent mechanism for pruning risk factors. Unlike random forests and neural networks, while logit in principle allows an unlimited number of non-linearities and interactions, in practice it quickly becomes overdetermined, especially when the dependent variable is rare.

Spatial autocorrelation

One common concern in regression-based models of conflict is spatial autocorrelation—i.e. the tendency for incidents to "cluster" within particular geographical areas. As we show in the online appendix, local violence in our communities does exhibit some clustering within counties. We also show, however, that proximity to towns that report incidents of local violence in the past is a poor predictor of new incidents in the future, and that including a

1,000 trees constructed in this manner. For a given observation, the algorithm generates a "final" forecast by taking the average of these 1,000 predictions.

spatial lag does not noticeably change the performance of our models.[5] We conclude that incorporating spatial spillover may improve predictive performance, but probably not by much.

Cross-validated forecasts

We trained our models and simulated a forecast through 5-fold cross-validation, using 2008 risk factors to predict local violence in 2010. Cross-validation has been shown to approximate out-of-sample accuracy and is widely used to evaluate model performance without new data (?). We proceeded in four steps. First, we randomly partitioned our sample into five equally sized subsets, each of which contained data on both 2008 risk factors and 2010 local violence. Second, we trained our models on four of the subsets and generated predictions for the fifth, reshuffling until we had a prediction for each observation based on a model that was not fit to that observation. We standardized all predictors to have zero mean and unit standard deviation.[6] Third, because a single cross-validation can yield idiosyncratic results, we repeated this process 200 times for each model. Finally, we identified optimal model parameters across the 200 trials, applied them to another 200 trials, and estimated average model performance. We call these within-sample tests our "cross-validated forecasts."

True forecasts

We then applied the optimal parameters from the cross-validated forecasts to the 2010 risk factors to generate predicted probabilities of violence in 2012. We published these forecasts online prior to 2012 data collection (?).[7] Afterwards we made only minor changes, described in the online appendix. We call these out-of-sample tests our "true forecasts."

These tests set a high bar, as Liberia was (and is) a country in flux. Between 2010 and

[5] The only exception appears to be the "true forecast" AUC on the neural networks model, though even here the Brier score remains unchanged.

[6] Some predictors are skewed. We show robustness to non-linear transformations in the online appendix.

[7] This earlier version of the article can be downloaded at `http://cu-csds.org/wp-content/uploads/2012/11/BBH_CAPERS.pdf`

2012, thousands of refugees returned to Liberia from neighboring countries; large numbers of police officers were deployed to rural areas for the first time; the UN withdrew several contingents of peacekeepers; and a presidential election was held. We sought to determine whether our models would perform well despite these changes.

Evaluating model performance

Forecasting involves a trade-off between sensitivity (true positive rate) and specificity (true negative rate). For continuous predictions, this trade-off is regulated by the threshold above which we predict our dependent variable will occur; in general, the higher the threshold, the lower the true positive rate, and vice versa. A ROC curve plots this trade-off over the entire range of possible thresholds. We plot ROC curves for each of our models, and also estimate the corresponding area under the curve (AUC), which captures each model's overall predictive power relative to chance; the higher the AUC, the better the model. We also include the Brier score, computed as the average squared deviation of each predicted probability from the true observed value of the dependent variable (0 or 1). The lower the Brier score, the better the model. For the cross-validated forecasts, we average both the AUC and the Brier score across the 200 trials.

While AUCs and Brier scores are useful benchmarks, they can be misleading because they weight false positives and false negatives equally. This may be sensible in some applications, but less so in others. For example, local violence may be sufficiently costly that a policymaker would be willing to tolerate many more false positives for the sake of fewer false negatives. We therefore also provide performance metrics at the point on the ROC curve that maximizes sensitivity while maintaining accuracy of at least 50% within each cross-validated trial. (Because performance varies across the 200 trials, not all models achieve an accuracy rate of 50% or higher on average after applying a single set of optimal model parameters.) This choice is principally a discursive device, however, capturing what we believe to be a sensible trade-off between sensitivity and specificity.[8]

Results

We focus on forecasting our aggregate dependent variable, which takes a 1 for any town in which leaders reported any incident of collective, interpersonal or extrajudicial violence in the past year. We focus on this coding of the dependent variable for two reasons. First, as noted above, our qualitative investigation revealed that a single incident of local violence could easily be classified in multiple ways. Second, aggregation makes rare events less rare, reducing the risk of overfitting and increasing the stability of our results. This is especially important in cross-validation, which requires splitting the data into subsets. Many cross-national models aggregate for similar reasons, e.g. by combining multiple events into a single indicator for âĂIJpolitical instabilityâĂİ (**?**) or "adverse regime change" (**?**). We discuss forecasts of our disaggregated dependent variables below and in the online appendix, though these results are unstable and should be interpreted with caution.

Cross-validated forecasts

Figure **??** plots the ROC curve for each cross-validated forecast of 2010 local violence using 2008 risk factors. The solid circles indicate the thresholds that maximize sensitivity while maintaining accuracy at or above 50%. Moving left on the ROC from each solid circle reduces the number of towns in which we predict local violence will occur, lowering sensitivity but increasing specificity and (in some cases) accuracy. This would be appropriate, from a policy perspective, if it were difficult or costly to expend resources on prevention. (Incidentally, the solid circles correspond to the points of maximum distance from the diagonal for both lasso and neural networks. This suggests that further improvements in sensitivity beyond our

[8] By way of comparison, **?** opt for the threshold that equalizes sensitivity and specificity, weighting false positives and false negatives equally. Note that the surest way to increase accuracy would be to predict peace in every town. This approach would achieve sensitivity of 0%, specificity of 100%, and overall accuracy of 83% in both 2010 and 2012.

Figure 1: ROC curves for cross-validated forecasts of 2010 local violence using 2008 risk factors

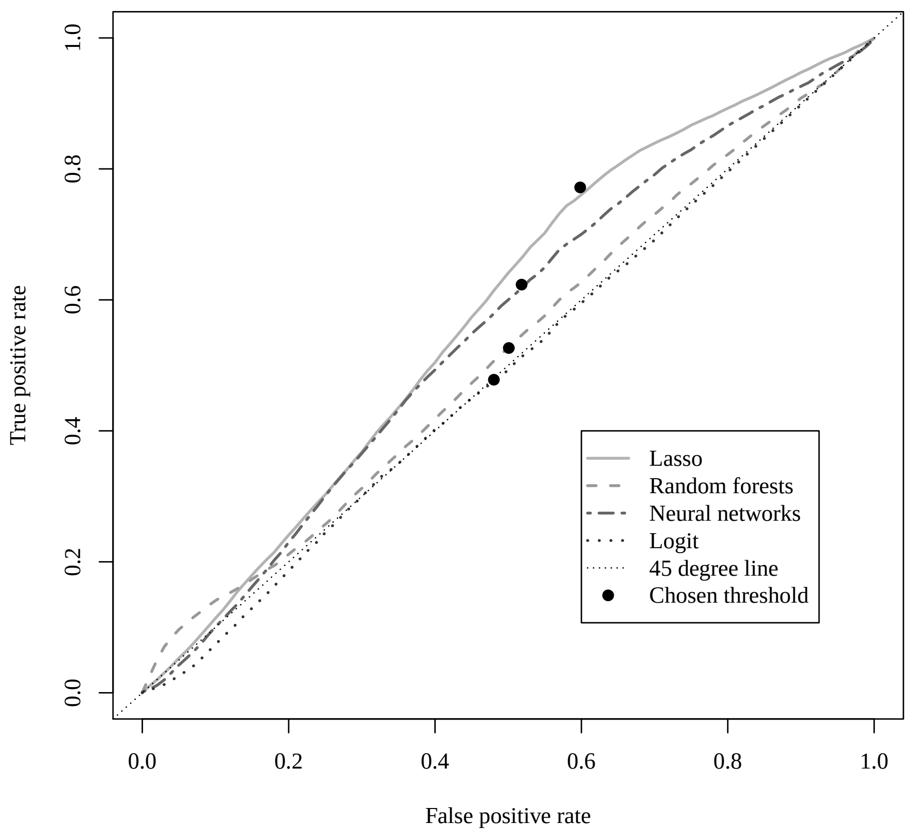


Table I: Performance metrics for cross-validated forecasts of 2010 violence using 2008 risk factors

	Aggregate dependent variable				
Performance metric	Lasso	Random forests	Neural networks	Logit	Recurrence
AUC	0.58	0.52	0.53	0.52	
	(0.03)	(0.02)	(0.04)	(0.03)	
Brier score	0.14	0.15	0.20	0.26	
	(0.00)	(0.00)	(0.01)	(0.02)	
True positive rate (sensitivity)	77%	52%	62%	48%	48%
	(0.05)	(0.05)	(0.06)	(0.08)	
True negative rate (specificity)	40%	50%	48%	52%	65%
	(0.02)	(0.02)	(0.03)	(0.04)	
Accuracy	47%	50%	51%	51%	62%
	(0.02)	(0.02)	(0.03)	(0.03)	
Ratio of false + to true +	3.71	4.62	3.99	4.88	3.50
	(0.26)	(0.46)	(0.44)	(0.73)	
Ratio of false - to true +	0.30	0.94	0.62	1.15	1.10
	(0.09)	(0.19)	(0.16)	(0.36)	

preferred threshold would have required increasingly large penalties in terms of specificity and accuracy.) In addition to the AUC and Brier score, Table **??** reports five performance metrics at the thresholds indicated by the solid circles in Figure **??**. We also report the standard deviation of these metrics over the 200 cross-validations. Finally, as a benchmark we report results from a "recurrence" model in which we predict local violence in 2010 wherever it occurred in 2008.

While none of our models performs especially well in cross-validation, lasso exceeds the alternatives. It achieves the highest AUC overall—0.58, compared to 0.53 or less for the other models—and also outperforms the alternatives at virtually every threshold on the ROC curve (though these differences are not always statistically significant). It also achieves the lowest Brier score. Lasso's performance is especially noteworthy given its parsimony: as we will see in Section **??**, it relies on just five of our 56 risk factors. Random forests is in a sense parsimonious as well, in that it assigns importance scores of near zero to most predictors.

Lasso outperforms at our preferred threshold as well, achieving a true positive rate of 77% and an accuracy rate of 47% overall. The former is considerably higher than the sensitivity rates achieved by the other models, while the latter is only slightly lower than the other models' accuracy rates. High sensitivity comes at the cost of a high ratio of false to true positives, however, which reflects our intuition that false negatives are far costlier than false positives, and that policymakers might be willing to tolerate more of the former for less of the latter. (The threshold can, of course, be adjusted to accommodate different preferences.) At our preferred threshold, lasso achieves high sensitivity without sacrificing much in the way of accuracy relative to the other methods (though its specificity is lower—a result of over-predicting violence). The standard deviations on these metrics are small, suggesting the results are relatively stable over cross-validations. They are also robust to most modeling choices (see the online appendix).

Nor are the results driven by serial autocorrelation alone, as the lagged dependent variable predicts less than half of 2010 local violence, with a true positive rate of just 48%. More important, as we will see, the lagged dependent variable does not appear among the top 10 predictors for any model. While local violence is thus to some extent autoregressive, there is room for improvement with the inclusion of additional risk factors. It is possible that the degree of serial autocorrelation would increase in a longer panel, or that further lags of the dependent variable would improve performance and displace other covariates. We cannot say.

These results do not imply that lasso is superior to random forests, neural networks or logit in any universal sense. The standard deviations on the AUCs appear to overlap, and further optimization may have improved the performance of the other models. For example, dropping a subset of poorly-performing predictors before running the cross-validated forecasts tends to increase the sensitivity of the random forests and neural networks models to levels similar to lasso (see the online appendix). Nonetheless, given the relatively small number of towns and time periods and the relatively large number of predictors, lasso's approach to variable selection appears to be advantageous.

True forecasts

Figure **??** plots ROC curves for our "true forecasts" of 2012 local violence using 2010 risk factors. Table **??** reports the AUC and Brier score for each model, as well as five performance metrics at our preferred threshold. Standard errors for the AUC are estimated via stratified bootstrapping.

All of our models achieve higher AUCs and lower Brier scores in the true forecasts than in cross-validation. They also achieve higher sensitivity at our preferred threshold. Moreover, lasso no longer dominates the other models. It still outperforms random forests and neural networks in terms of both AUC and Brier score, but the random forests and neural networks ROCs cross the lasso ROC at the lower and upper ends of the distribution, suggesting that lasso's dominance depends on the threshold above which we predict violence will occur. Despite logit's poor performance in cross-validation, it now outperforms lasso in terms of

Figure 2: ROC curves for true forecasts of 2012 violence using 2010 risk factors

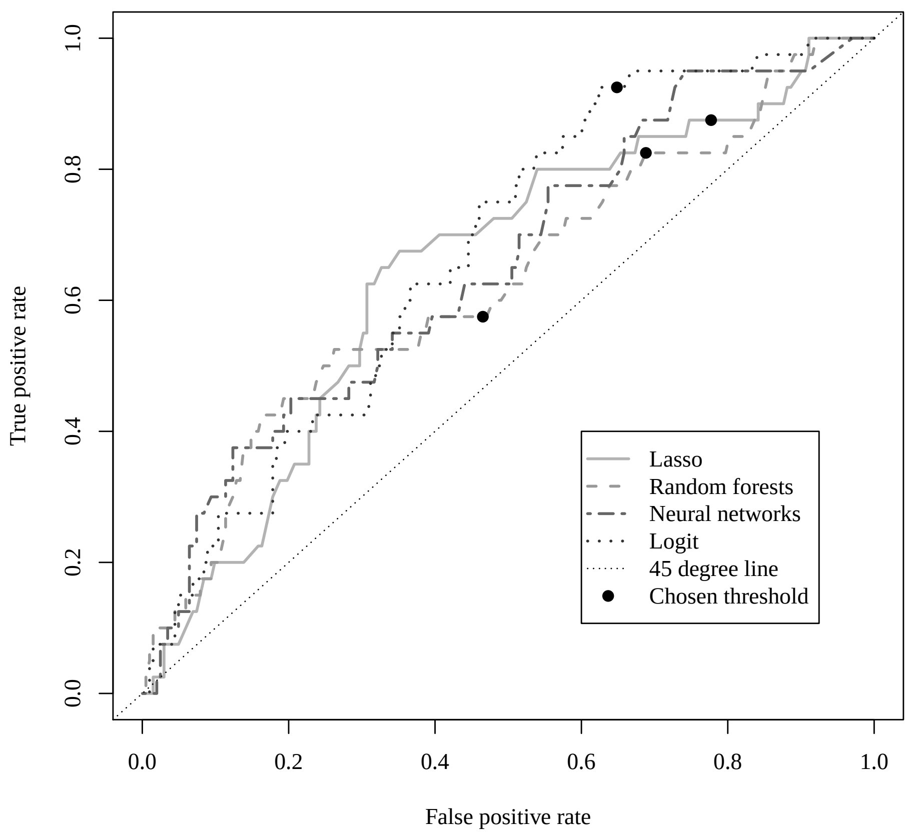


Table II: Performance metrics for true forecasts of 2012 violence using 2010 risk factors

	Aggregate dependent variable				
Performance metric	Lasso	Random forests	Neural networks	Logit	Recurrence
AUC	0.65	0.63	0.62	0.67	
	(0.05)	(0.05)	(0.05)	(0.05)	
Brier score	0.13	0.14	0.17	0.15	
True positive rate (sensitivity)	88%	83%	58%	93%	38%
True negative rate (specificity)	22%	31%	53%	35%	87%
Accuracy	33%	40%	54%	45%	79%
Ratio of false + to true +	4.49	4.21	4.09	3.54	1.8
Ratio of false - to true +	0.14	0.21	0.74	0.08	1.7

AUC, though lasso continues to outperform in terms of Brier score. In general, each model's AUC and Brier score are robust to modeling choices, but specificity and sensitivity at our preferred threshold vary considerably, especially for random forests and neural networks (see the online appendix).

Alternative coding rules and model specifications

Here we briefly consider four alternative coding rules and model specifications, and include figures, tables and robustness checks in the online appendix.

Using a less conservative coding rule

In our analyses above we use a conservative coding rule for our dependent variable, potentially undercounting local violence if different leaders disagree about the appropriate categorization of the same event. As a less conservative alternative, we code the dependent variables as a 1 if at least two leaders reported any incident of any type in the past year (even if their categorizations of the incident were different). This increases the prevalence of local violence in the sample—from 37% to 40% in 2008; from 17% to 27% in 2010; and from 17% to 20% in 2012—and improves model performance significantly. AUCs increase by 6 to 9 points, to a high of 0.67 for the cross-validated forecasts (lasso) and 0.74 for the true forecasts (random forests, though lasso's AUC is only slightly worse at 0.73). Lasso and random forests are the best performers overall; logit is the worst.

Dropping ambiguous incidents

In our analyses above we do not correct the coding of potentially non-violent incidents. Our qualitative follow-up revealed seven such incidents, but the performance of our models is virtually unchanged if we recode these as 0s.

Disaggregating incidents by category

In the online appendix we reproduce our cross-validated and true forecasts with the dependent variable disaggregated into its three component parts: collective, interpersonal and extrajudicial violence. Model performance varies across time periods and dependent variables, and no model unambiguously dominates the others. As mentioned above, disaggregating in this way makes the dependent variable rarer, increasing the risk of over-fitting[9] and reducing the stability of the results, which should thus be interpreted caution.

Averaging across models

Finally, recent research suggests that aggregating forecasts across multiple models can yield more accurate results (e.g. **?**). We consider several approaches to aggregation in the online appendix. Overall, the performance of these models is similar or marginally superior to that of our best-performing models above.

Identifying the most reliable predictors of local violence

Identifying the most reliable predictors of local violence may be equally if not more informative than predicting where it is most likely to occur. Table **??** ranks our 56 risk factors by (1) the absolute value of their coefficients in the lasso model and (2) their importance scores in the random forests model, where "importance" is calculated as the average decrease in mean squared error achieved by the addition of each predictor to the model. Because neural network weights cannot be meaningfully ranked in this way, and because of the limitations of logit discussed above, we focus on lasso and random forests alone. A comparison to logit, and a list of all 56 risk factors and their corresponding coefficients, is provided in the online appendix.

We observe some model dependence in the rankings, which is unsurprising given that

[9] For example, while the lasso model relies on just five risk factors to predict the aggregate dependent variable, it uses dozens of risk factors to predict each of the disaggregated categories.

Table III: Rankings of risk factors by model

	Lasso		Random forests	
Risk factor	Rank	Coeff.	Rank	Importance
Minority tribe in town leadership	1	0.31	7	0.0004
Town population	2	0.16	1	0.0024
% believing other tribes are violent	3	0.08	23	0.0000
% in majority tribe	4	-0.05	2	0.0008
% who contribute to public facilities	5	0.01	47	-0.0001
Mean educational attainment			3	0.0007
Number of tribes in town			4	0.0006
Number of households in town			5	0.0005
S.D. of wealth index			6	0.0004
% reporting loss of land during civil war			8	0.0004
% reporting burglary or robbery			9	0.0003
% non-native (Residents)			10	0.0003

lasso tends to favor uncorrelated regressors while random forests tends to favor correlated ones. Nonetheless, the highest ranked predictors do seem to capture some similar town characteristics across models. Some of these patterns are intuitive. For example, given that our dependent variable is binary, the correlation between violence and town population may be mechanical: more people means more potential disputants, and therefore more potential violence. Other patterns are consistent with existing theories about local violence in general, and about Liberia specifically. For example, four of the top 10 predictors—"minority tribe in town leadership," "% believing other tribes are violent," "% in dominant group" and "number of tribes in town"—are related to ethnic heterogeneity, polarization and fractionalization. This is unsurprising given that Liberia remains divided along ethnic lines. An estimate for the proportion of residents reporting land loss is among the top predictors in random forests (though not in lasso) as well, which is unsurprising given that disputes over property rights are a persistent source of tension in rural Liberia (and across sub-Saharan Africa), and that many of these disputes revolve around land lost in the civil war.

Other results are more unexpected. For example, while risk factors related to wealth and inequality are highly ranked in the random forests model (though again not in lasso), adverse economic shocks—droughts, floods, pest infestations and diseases—are not. This is striking given the number of studies focused on estimating the impact of shocks on violence of various kinds (see **?** for a review). Also unexpected is the absence of the lagged dependent variable from the top 10 predictors for either model. This is especially relevant from a policy perspective. In settings where information about violence is limited or unreliable, police, peacekeepers and NGOs may use past incidents as heuristics for allocating scarce resources in future. The absence of the lagged dependent variable from Table **??** reveals the potential limitations of this approach (though serial autocorrelation may play more of a role in a longer panel, and we do find some evidence of serial autocorrelation even here).

Perhaps most striking is the relationship in the lasso model between violence and an indicator for whether or not majority and minority tribes share political power. This indicator is the most reliable predictor in our most reliable model, with a coefficient twice as large as the second highest-ranked righthand-side variable. (Power-sharing appears among the top ten random forests predictors as well, but is lower ranked.) Yet, counterintuitively, the coefficient is *positive*, meaning that power-sharing heightens rather than reduces the predicted probability that local violence will occur. We return to this result in the discussion and conclusions below.

Identifying the most reliable predictors after disaggregating incidents by category

Of course, different categories of local violence may have different predictors. In the online appendix we replicate the analysis in Table **??** for each of our three categories: collective, interpersonal and extrajudicial. Caution is warranted when interpreting these results, since the appropriate categorization of incidents is often ambiguous, and since the top risk factors vary dramatically across cross-validated trials—a result of an increasingly rare dependent variable. With these caveats in mind, our results suggest that, in general, different categories of violence do indeed have different predictors, though some recur.

Discussion and conclusions

Overall, our results suggest that prospects for leveraging survey data to forecast local violence are promising. We find that a relatively simple, parsimonious model (lasso) outperforms most alternatives, and predicts local violence reasonably accurately using few risk factors, especially when we apply a less conservative coding rule to our dependent variable. Lasso's parsimony, and its reliance on several risk factors that are either slow-moving or time-invariant, suggests that future models may achieve similar or superior results at lower cost in terms of new data collection, though this proposition awaits further testing.[10] Fortunately,

[10] One important open question in Liberia is whether our models will continue to perform well in the wake of the Ebola epidemic—a far more destabilizing event than any that occurred during our study period.

the continued proliferation of survey-based datasets on local violence and its correlates offers many opportunities for replication. Where datasets are limited or non-existent, or where the cost of conducting surveys is prohibitively expensive, cell phone polls or web-scraping may serve as substitutes. Administrative data compiled by police and peacekeepers—e.g. arrest records and criminal complaints—can further complement these efforts.

Even when we relax our coding rule for the dependent variable, however, the AUCs of our micro-level models, which never exceed 0.74, still underperform many of their macro-level counterparts. What might explain this disparity? One possibility is that the outcomes we study—riots, murders, lynchings—are inherently more difficult to predict than, say, civil wars or battles between armed groups. Another is that the slow-moving "structural" variables that we measure in the survey need to be combined with faster-moving "process" variables, such as changes in leadership or influxes of migrants at the community level. (Indeed, some cross-national models are moving in this direction; see, e.g., **?**, **?** or, in this special issue, **?**). Another possibility, discussed above, is that our panel is too short—shorter than those used in most cross-national studies. Yet another is that we simply failed to measure the optimal predictors of local violence, and so failed to achieve optimal results. While we were aimed for comprehensiveness in the variables we measured, oversights were inevitable—there are only so many questions a survey respondent can answer—and we cannot know how these gaps may have affected our results. These caveats notwithstanding, given the novelty of the exercise, we view the performance of our models as a promising first step.

Beyond their potential practical applications, our models generate interesting substantive patterns for future research to explore. Perhaps most notable is the strong positive correlation between local violence and power-sharing. Previous studies have found that exclusionary institutions foment conflict between dominant and marginalized groups (**?**), and that power-sharing helps mitigate the risk of violence (e.g., in this special issue, **?**). Others, however, have warned of the dangers of power-sharing (**?**) or of cohabitation between equally dominant ethnic or religious groups (**?**), and have suggested that violence often accompanies minority access to political power (?). Our results are more consistent with these latter intuitions.

Of course, the correlation we observe is not evidence of causality, and interpretation remains ambiguous. One possible explanation is that once minority groups gain access to power, majority groups resort to the use of force to suppress and intimidate them. Another is that current power-sharing arrangements are responses to past conflicts (?), and that past conflicts continue to predict future ones. A third is that the correlation is spurious—an artifact of over-fitting. We view this as unlikely given the strength of the association, and given that it emerges most strongly in a model with only five predictors. Nonetheless, we are careful not to generalize too much from this one case, and we are doubly careful not to imply a relationship of cause and effect. Whatever the nature of the relationship, we believe it merits further study, especially at the local level.

More generally, we believe our results argue for a more even balance between forecasting and hypothesis-testing in conflict research. Currently the balance of this literature—and of comparative politics and international relations more generally—is strongly skewed towards hypothesis-testing, with relatively few attempts at prediction. At this extreme, the marginal gains from forecasting are probably large. A myriad of existing micro-level datasets could be harnessed for purposes of prediction, and new datasets are being constructed every day. We view this unexplored frontier as one of the disciplineâĂŹs most promising.

Data replication

The dataset, codebook, and do-files for the empirical analysis in this article can be found at `http://www.prio.org/jpr/datasets`.

Acknowledgements

This research was supported by grants from the United Nations Peacebuilding Fund in Liberia (via UNHCR), Humanity United, the MacMillan Center for International Affairs at Yale University, and the World BankâĂŹs Italian Children and Youth Trust Fund. For comments we thank Peter Aronow, Kate Baldwin, Neal Beck, Grant Gordon, Guy Grossman, Stathis Kalyvas, Jay Ulfelder, Elisabeth Wood, the anonymous referees, and seminar participants at Yale University, the 2011 Northeast Methodology Program meeting, a 2012 Carnegie Mellon University Machine Learning Department colloquium, a 2014 CAPERS meeting at Columbia University, a 2014 APSA Annual Meeting panel, a 2014 BREAD conference, and a 2015 World Bank/Households in Conflict Network conference. Bryan Plummer, Mathilde Emeriau, Tricia Gonwa, Rebecca Littman, Benjamin Morse, Grant Gordon, Daniel Masterson, Johnny Ndebe, Richard Peck, Gwendolyn Taylor, Prince Williams, and John Zayzay provided superb research assistance.

Biographical statements

ROBERT A. BLAIR, PhD in Political Science (Yale University, 2015); Joukowsky Family Assistant Professor, Brown University, Political Science and Watson Institute for International and Public Affairs.

CHRISTOPHER BLATTMAN, PhD in Economics (University of California, Berkeley, 2007); Ramalee E. Pearson Professor of Global Conflict Studies, University of Chicago, Harris Public Policy.

ALEXANDRA HARTMAN, PhD in Political Science (Yale University, 2015); Lecturer in Qualitative Research Methods, University College London, Political Science.

Appendix for online publication

Survey sites

Our unit of analysis was the smallest unit of local administration in Liberia: the village, town, or (in slightly larger towns) the town quarter. 36 larger towns with quarters are in the sample, with 5 quarters at the median. For simplicity we refer to these all as "towns". In 2008 they had populations of 439 at the 10th percentile, 4,045 at the 90th percentile, and 1,811 at the median.[11] Figure **??** below maps them, along with the geographical distribution of violence in each year of the survey.

There were originally 246 towns in the sample in three counties (a "county" is akin to a state or province in other countries). We have survey data on 242 of these, as surveyors could not reach two extremely remote villages for any survey round, one tiny village disbanded before the first follow-up, and one town is missing data.

The data were collected in the context of a randomized evaluation of a government-sponsored alternative dispute resolution training intervention, and the towns are not a representative sample of towns in these counties. Rather, county officials nominated these towns because they were thought to be more dispute-prone than others. It is difficult to say how this sampling approach affects our predictions and predictive power. It could improve or reduce performance. One can imagine that focusing on the subset of places perceived to be most risky is useful from a policy perspective, and reasonable under budget constraints. Nonetheless, one would like to be able to test these assumptions using a representative sample of towns, or stratified random sample. Unfortunately such a sample is not available in this case, and is recommended for future research.

No census frame existed at the time of the first round of data collection. To create a representative sample, a team walked each town and divided it into blocks, chose a random pathway, counted all houses along that pathway, and randomly chose a set number of households to survey. Household members were selected randomly. Non-response was typically less than 5 to 10% per town.

Measurement of local violence

Descriptive statistics for local violence

Table **??** reports the prevalence of violence in each survey round. In 2008, 37.2% of towns reported at least one major incident. By 2010, that proportion roughly halved, to 17.4%.[12]

[11] There are 50 town quarters in 36 larger towns. The largest town has 16 quarters, but most have far fewer—a median of 5.

[12] At baseline the "fights with weapons" question was less specific, asking only about "serious fights." This accounts for some of the decline in fights from 2008 to 2010, but the decline in other categories is similar. If we omit all fights (with and without weapons) from the aggregate indicator, prevalence rates in 2008, 2010 and 2012 are 29%, 16% and 15%, respectively. Thus the fall from 2008 to 2010 is still precipitous.

Table A.1: Number of towns reporting any major incident of crime or violence, 2008–12 (n=242)

Dependent variable	2008	2010	2012
Any major incident of crime or violence	90	42	40
Any collective violence	25	9	7
Any violent strike or protest	8	5	4
Any violent confrontation between tribes	17	6	3
Any interpersonal violence	64	32	21
Any murder	9	14	5
Any rape	33	17	11
Any serious fight with weapons	43	9	6
Any extrajudicial violence	23	8	16
Any trial by ordeal	23	7	16
Any killing or beating of witches	1	1	0

By 2012, however, the rate of decline had slowed, and the proportion of towns experiencing at least one major incident remained high (16.5%).

Comparison of local violence in our sample relative to a nationally representative one

Our selection of towns is not representative of Liberia, or of the counties from which they were sampled. They were identified by government officials and other stakeholders because they were believed to be at disproportionately high risk of local violence. Nonetheless, comparison to a nationally representative survey (**?**) conducted at the same time as our second wave of data collection suggests that our towns were not much more conflicted than the average Liberian community, either in these three counties or nationwide.

Since **?** do not survey leaders, we cannot directly compare our measures of collective, interpersonal and extrajudicial violence to theirs. We can, however, compare rates of armed violence as reported by citizens themselves. 4% of respondents in Vinck et al.'s survey reported being victims of armed violence in Lofa, 5% in Nimba and 4% in Grand Gedeh, compared to 7% nationwide. In our survey, 1% of respondents reported being victims of armed violence in Lofa, 3% in Nimba and 3% in Grand Gedeh. These rates are comparable across surveys in all counties. Moreover, since our question is more specific than Vinck et al.'s—we ask about armed robberies and aggravated assaults specifically, while Vinck et al. ask about any violence involving a weapon—it is likely that our results are underestimates relative to theirs. Though less directly relevant, rates of robbery and burglary are similar as well,[13] as are complaints of witchcraft—indeed, if anything the latter are less common in our sample than in Vinck et al.'s[14].

[13] In Vinck et al., 12% of respondents reported a robbery or burglary in Lofa, 11% in Nimba and 18% in Grand Gedeh, compared to 15% nationwide. In our survey, 14% of respondents reported a robbery or burglary in Lofa, 25% in Nimba and 13% in Grand Gedeh. Except for Nimba, these rates are comparable across surveys.

[14] 19% of Vinck et al's respondents reported being victims of witchcraft in Lofa, 21% in Nimba and 14% in Grand Gedeh, compared to 17% nationwide. In our survey, 6% of respondents reported being victims of witchcraft in Lofa, 10% in Nimba and 11% in Grand Gedeh. If anything, complaints of witchcraft appear to be less common in our sample than in Vinck et al.'s.

Of course, it is possible that our towns are more conflicted along other dimensions, including potentially the three dimensions of local violence we attempt to forecast. But it is not obvious why this would be the case, and we interpret the parallels between our survey and Vinck et al.'s as evidence that the towns in our sample are not much more conflicted than the average Liberian community.

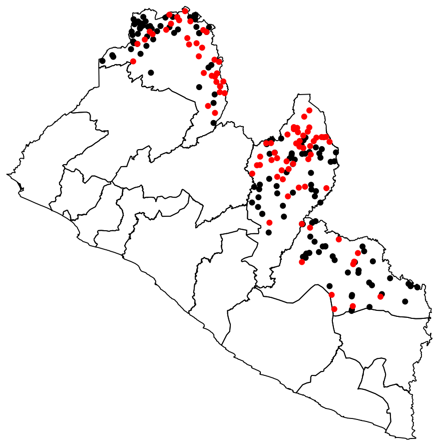
Spatial autocorrelation in local violence

Figure **??** shows the spatial distribution of local violence in 2008, 2010 and 2012. The decline in violence in Table **??** appears to be most pronounced in Lofa County, and least so in Nimba. Because these towns are not a representative sample, however, we cannot be sure whether violence is spatially concentrated in this way throughout the three counties, rather than just among the towns in our sample. Areas that appear relatively unpopulated may include towns that are simply absent from our dataset, and where we therefore do not know the prevalence of violence.

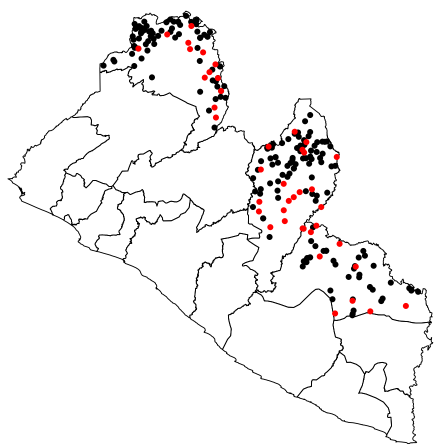
While we do find evidence of spatial autocorrelation, proximity to towns that report violence in one period seems not to be correlated with the onset of violence in the next. Table **??** investigates this relationship through descriptive statistics and OLS regression. Compared to towns that reported violence in 2010, those that did not seem to be further away on average from towns that reported violence in 2008. This difference is not statistically significant, however, and may be explained by the fact that towns that reported violence in 2010 tend to be further away from *all* other towns. Controlling for this distance, there is no correlation between nearby violence in the past and violence in the present.

Figure A.1: Communities experiencing violence

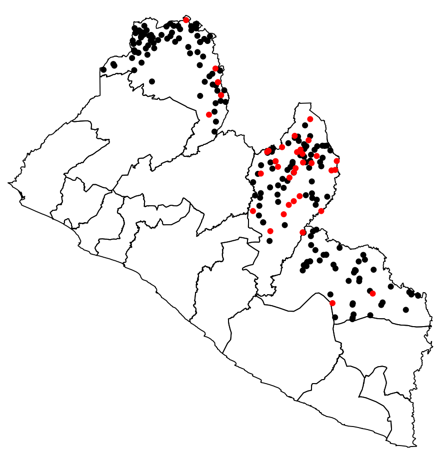
(a) 2008



(b) 2010



(c) 2012



Distribution of violent events over time. Study towns experiencing violence are red. Study towns that do not experience violence are black.

Table A.2: Effect of past violence in neighboring communities

	Summary statistics				Regression	
	Towns with violence in 2010		Towns without violence in 2010		D.V. : Any violence in 2010	
	Mean	S.D.	Mean	S.D.	Coef.	S.E.
	(1)	(2)	(3)	(4)	(5)	(6)
Miles to nearest town	3.96	(2.87)	2.73	(2.22)	0.0367	(.012)***
Miles to nearest town reporting violence in 2008	6.32	(5.04)	5.92	(4.46)	-0.0057	(.005)
					R^2	0.043

Robust standard errors in parentheses. ***p<.01,**p<.05,*p<.1.

Qualitative follow-up on survey-based reports of local violence

Quantitative data

We collected three types of qualitative data:

1. Between the first and second phases of data collection (2008–10), we and three Liberian research assistants conducted 104 formal interviews with respondents in 20 purposefully selected towns. We selected towns with high and low levels of conflict, as well as those showing variation along potentially important correlates of conflict (exposure to wartime violence, remoteness and size).
2. Following the second wave of data collection (2010–11), we sent our Liberian research assistants to investigate and verify all incidents of collective violence reported in the survey through interviews and written notes. While we did not have the resources to back-check other types of violence, these interviews helped us validate the survey data and explore the interconnections between apparently disparate violent events.
3. Finally, during the third wave of data collection, enumerators sought to record a short qualitative interview with any leader who reported an incident of violence (with the exception of serious fights, which we excluded).[15] Excluding serious fights, we have qualitative information on 125 (74%) of all leader reports of violence.[16] This exercise served two main purposes. First, it helped us to further validate the survey data, building our confidence that the dependent variable was measured with as little reporting bias as possible. Second, along with the earlier interviews, it informed our decision to aggregate different categories of violence into a single indicator.

This latter validation data provides two main insights: into survey categorization and potential for non-violent events to be captured in the survey.

Exploring variation in survey-based categorization

In general, respondents who reported incidents in the survey continued to do so during qualitative follow-up. However, the interviews also suggested that grouping incidents into non-overlapping categories would be challenging. Respondents described the same incidents in strikingly different ways. Much of this ambiguity resulted from the dynamics of conflict escalation. The police and courts in Liberia are notoriously inept, inaccessible and corrupt.

[15] Unfortunately, these were accidentally excluded by the field data collection team without the authors' awareness.

[16] There are several explanations for missingness. In some cases enumerators failed to record leaders' accounts altogether. In others, enumerators recorded leaders' accounts, but the recordings were lost over the course of data collection. In still others, enumerators recorded leaders' accounts but the recordings could not be transcribed because the audio was too poor. 12 leader accounts fall into this third category, and 32 fall into the first two.

Because victims cannot rely on these institutions to resolve disputes, violence easily mutates from one form to another (e.g., a murder turns into a riot or mob justice). Riots, lynchings and trials by ordeal often serve as extrajudicial mechanisms for adjudicating other types of crime (e.g., rape, murder, or suspected witchcraft). The distinction between violent ethnic clashes and violent strikes was also quite murky.
There are numerous specific examples from our fieldwork, and we outline a handful here for illustration.

- In our largest study town—Voinjama, Lofa County—the mysterious disappearance and killing of a girl provoked a peaceful protest which quickly turned violent. The (Lorma and Christian) mother of the missing girl accused the town's Mandingo (Muslim) population of abducting and murdering the girl in a ritual killing. Traditional healers ("Zoes") were called to attempt to divine the identity of the perpetrator, fomenting allegations of witchcraft. Riots ensued, killing four. In our interviews, respondents varied dramatically in how they categorized the sequence of events—as a riot, a murder, a violent confrontation between tribes, a fight between men, a lynching of suspected witches, or several or all of the above.
- In one of the larger towns, a hit-and-run accident provoked a violent protest by the motorbike union. In their descriptions of the incident, some respondents focused on the hit-and-run, others on the violent protest.
- In several towns, respondents described how trials by ordeal had been used to identify suspected murderers, typically in cases involving an unusual or mysterious death. As one local leader explained: "A little girlâĂępassed away within this town and everybody was surprised of that particular death, so the parents of that little girl decided to go for sassy wood [trial by ordeal]âĂę The sassy wood man came and he...used hot cutlass—they put the cutlass on the fire and...if you ainâĂŹt part of it will just be like water on your skinâĂę They started to do it going around all the people in the neighborhood.... They started touching them with the hot cutlass...and the cutlass was able to grab one personâĂębecause that particular person was the doer of the act."
- In a small town, we directly observed a seemingly intoxicated woman attack a man. Shortly thereafter another woman—a female relative or friend of the man—attacked the first woman. As the two women grappled, male family members and friends gathered and began to exchange insults. A physical fight between two of the men ensued. The crowd continued to grow, and several bystanders began agitating to join the fray. While the incident was eventually diffused, its interpersonal and collective dimensions remained difficult to disentangle.

Identifying potentially non-violent incidents

The 2012 qualitative accounts are not exhaustive, but a review suggests some potential for misreporting of "nonviolent" events as violent ones, though this depends on what types of coercive acts one considers "violent."

1. In three of the 16 cases of extrajudicial violence, individuals may have elected to take sassywood oaths. The qualitative follow-up suggests that these individuals were not forced to consume a potion or burn themselves with a hot cutlass, but instead willingly participated in a milder procedure. In one community, both accounts of extrajudicial violence suggest that the sassywood process involved the accused voluntarily swearing an oath in a non-violent manner. In another town, the two accounts of the case suggest that the accused woman volunteered to participate in the sassywood (although whether she took an oath or engaged in riskier procedure such as poison drinking is less clear). She survived. Finally, in another community, there are two reports of extrajudicial violence. One of the recordings is too poor to understand, while the other account suggests that the oath was voluntary and no one was hurt.

2. Three of our 7 cases of collective violence may have actually been relatively peaceful protests. In one community, we have two reports, one of a protest against an international mining company, and one of a protest of the town commissioner's leadership. It is unclear if these accounts are describing the same incident or related incidents, but both interviews say explicitly that the protests were not violent. In another community, we collected two reports of a protest related to a land dispute, and both mentioned that the dispute was not violent. In still another community, there are two reports of a violent strike. For one of these reports, the recording quality is too poor for us to extract any qualitative data. Qualitative data from the other report describes a dispute over land brushing, but the respondent repeats "they did not spoil nothing", which we interpret to mean there was no violence. [17]

3. In one of our five communities coded as experiencing a murder, the case might be better described as either an accidental or negligent killing, from a hunting accident. All three leaders who confirmed the event in the qualitative data suggest that the victim could have died by accident. Nevertheless, the case did involve a violent death and was of interest to the community authorities, and there was uncertainty regarding what exactly transpired.[18]

Overall, five of the 40 incidents (13%) of violence in 2012 could be coded as zeroes if we adopted a more conservative coding rule based on the qualitative data. All are of significant police interest, so it's not clear we would ant to ignore them in an early warning system. Nonetheless, we consider this scenario below.

[17] In addition to these two communities, there are 2 communities for which we collected 1 report of a non-violent protest that we coded as violent, but, because there was only one report in each community, the towns were never coded as having strikes.

[18] There are two other reports that we coded as murders, while the qualitative data reported accidental deaths while hunting. These reports occur in separate communities and are the sole reports of murder for those communities, so our dependent variable is unaffected by these according to our original coding. Finally, there is one town where there are 3 reports of the same murder happening in the bush. One of the three reports suggests the death could have been an accident, but the other two call it a murder, and one mentions an arrest of the shooter.

Table A.3: % of towns with any major incident of crime or violence, 2008–12 (n=242)

Coding rule	2008	2010	2012
By event type (as in Table 1 of main paper)	37%	17%	17%
By event category (modal leader reports event in same category)	39%	20%	18%
By any event (modal leader reports any kind of seven events)	40%	27%	20%

There is also one case where we coded an account as extrajudicial violence, but the qualitative data makes it clear that a murder occurred. However, this is the only account of extrajudicial violence in this community, so this miscoding does not affect our final dependent variable. In principle this could be an undercounted incident.

All other qualitative accounts were consistent with our survey data and increase out confidence in our quantitative measures of violence and our conservative coding of the dependent variable. Nevertheless, it is worth noting that the data are sometimes challenging to interpret, especially for cases of extrajudicial punishment. By definition these cases are always associated with other events, which range from stealing to suspected witchcraft to murder. There is variation in whether individuals report that the accused participated willingly in extrajudicial procedures, or whether they were forced to do so. In addition, the procedures themselves vary, from being forced to drink poison to being burned with a hot knife to swearing an oath in front of the community. Nevertheless, all extrajudicial punishments are illegal in Liberia and are of interest to local authorities responsible for maintaining order and security.

Alternative coding rules

Difficulty of categorization raises the possibility that our coding rule (modal leader reports of each of the seven event types) could underreport the least well-known and most ambiguous incidents of violence. For instance, if two leaders were unaware of an event, and the other two differed in their description of the event—one calling it a trial by ordeal and the other referencing the rape that preceded the trial by ordeal, for instance—then we would code no event. In principle this biases our coding towards better-known events with established narratives.

Table **??** reports descriptive statistics for the dependent variable under alternative coding rules. The first row corresponds to the coding rule reported in the paper. For the second row, rather than take the modal leader report for each of seven types of violence, we instead take the modal leader report for each of three categories of violence. For the third row, we take the modal leader report of *any* violence. The prevalence of violence increases with each of these changes to the coding rule. However, the increase is generally modest, and no more than a few percentage points. (The exception is the increase in 2010 violence between the second and third rows). We report predictive performance for these alternatives in Appendix **??**.

Relative reliability of survey- vs. news-based data

We validate our events qualitatively, as described above. Unfortunately, it is not possible to further validate our data against third party sources, such as the Armed Conflict Location and Event Data Project, or ACLED (?). First, the events we measure are inherently smaller-scale than those reported in ACLED and other news-based datasets. Second, the majority of the events in the ACLED dataset occurred during, not after, the Liberian civil war, so the datasets do not overlap much in time. Third, as we discuss below, news-based databases are incomplete and biased, especially in fragile states such as Liberia.

While we cannot validate our data against other sources, we can make more general comparisons between them. We can compare our interview-based (and qualitatively validated) data to at least three alternatives: the Integrated Crisis Early Warning System or ICEWS (?); ACLED (?); and (in Liberia) the Liberian Armed Violence Observatory or LAVO (?). These three datasets represent prominent approaches to sub-national data collection on violence. ICEWS is a machine-coded events database constructed exclusively from web-based news reports (via Factiva). ACLED is human-coded, and draws on a combination of web-based news, NGO and research reports. LAVO is human-coded and draws on a similar combination of sources, but focuses exclusively on Liberia, and includes both web- and print-based news, as well as police and hospital records. Each of these datasets illustrates the challenges of capturing local violence in a setting like Liberia.

- Many news-based events do not have specific location information. **?** notes that news-based datasets like ICEWS "will often fail to identify the specific location (i.e. city) of reported events up to 80% of the time." For instance, 68% of all ICEWS events in Liberia do not specify a county (akin to a state or province), and 81% do not specify a town or city. Of events with a location, more than half are in the capital. ICEWS has the advantage that non-geolocated events are included, so the extent of incompleteness is clearer than for sources such as ACLED.[19]

- News-based datasets likely underreport conflicts. **?** note, for example, that news reports constitute only a "tiny, tiny fraction" of the events that occur in a given location over a given period of time, and are "non-randomly selected by reporters and editors." They also note that the dictionaries used to populate news-based datasets are "very generic" and tend to "bin together events that may not always belong together." In this sense, questions of aggregation that we view as a choice in our models are answered by default in many news-based datasets. We see two notable examples of this underreporting in Liberia and the DRC.

  - **?** collected detailed village histories of rebel attacks and occupation in 380 com-

[19] Our understanding is that ACLED does not include events with incomplete information such as location, and so the extent of selection bias is unclear.

munities in Eastern DRC. He finds that only about 5 to 10% of attacks are recorded by ACLED. The difference is greatest in the most intense periods of war (when there may be too many incidents to report) and also varies systematically with rebel occupation (as these groups likely limit the flow of information).

– In Liberia, a major limitation is the capacity and reach of the written press. Liberian reporters have limited to no funds to travel outside the capital and a few major towns except in response to the most serious crises. Rural reporting is mainly limited to community and national radio, which do not enter news-based datasets. Thus reporting is heavily biased to the most populous areas. Indeed, in 2012 just 0.1% of LAVO violent events were taken from media sources. The rest were gleaned from police and hospital records (which, of course, have biases of their own). Similarly, 63% of the 92 ACLED incidents reported between 2008 and 2012 in Liberia are identified as occurring in the capital. Finally, news-based datasets tend to focus on higher profile events. Thus, even if they were complete, our data and coding of events would be different and complementary.

Measurement of risk factors

Descriptive statistics for risk factors

Table **??** reports sample means for all 56 risk factors in 2008 and 2010, with individual survey responses aggregated to the town level. In some cases we have data from both residents and leaders and so have two measures of the same predictor. Table **??** reports full summary statistics. Figure **??** displays density plots for all non-binary variables. In the density plots for continuous variables, the outliers (towns above the 95 percentile) are excluded from the graphs.

Table A.4: Risk factors

Covariate	2008	2010	Covariate	2008	2010
Town population	2,032	3,117	% Muslim (leaders)	9%	5%
# of households	238	337	% Muslim (residents)	12%	12%
% male	56%	48%	Indicator for mosque in town	45%	21%
% under 30 yrs. old	20%	27%	% accepting inter-religious marriage	66%	73%
% non-native/strangers (leaders)	2%	3%	% saying Muslims shouldn't be leaders	27%	55%
% non-native/strangers (residents)	13%	29%	% believing other tribes violent	30%	65%
% ex-combatants (leaders)	2%	2%	% believing other tribes dirty	15%	56%
% ex-combatants (residents)	9%	8%	Minority tribe in town leadership	59%	84%
% returned from internal displacement	58%	36%	% reporting burglary or robbery	13%	19%
Mean educational attainment (years)	5.17	5.54	% reporting assault	19%	8%
% with no education	45%	42%	% reporting any land conflict	25%	21%
% receiving any peace education	28%	34%	Any violent event (lagged DV)	37%	17%
Group participation (0–9)	3.74	3.69	% of town landless (leaders)	1%	0%
Collective public goods index (0–11)	1.79	1.61	% of town landless (residents)	17%	12%
% who contribute to public facilities	86%	89%	% of town farmers	18%	55%
% saying town is safe at night	52%	43%	Unemployment rate	4%	7%
% saying neighbors helpful	70%	50%	Wealth index	-0.02	-0.02
% rely on NGOs for public goods	53%	53%	S.D. of wealth index in town	0.69	0.76
% rely on gov't for public goods	14%	17%	Exposure to war violence (0–13)	4.28	5.19
% describing police/courts as corrupt	33%	44%	Participation in war violence (0–3)	0.31	0.45
Perceived equity in institutions (0–3)	2.60	2.37	% reporting loss of land during war	10%	9%
# of tribes in town	2.63	2.66	% displaced or refugee during war	80%	84%
% in largest tribe	82%	87%	Social services in town (0–14)	5.61	6.81

Table A.5: Detailed summary statistics for predictors

Year	2008	2010	2008	2010	2008	2010	2008	2010	2008	2010
	Town population		# of households		# of tribes		Mosque in town (leaders)		% Muslims (leaders)	
Mean	2032	3117	238	337	2.63	2.66	0.45	0.21	0.09	0.05
Median	1811	2920	150	272	2	2	0.00	0.00	0.00	0.00
Minimum	20	30	4	7	1	1	0.00	0.00	0.00	0.00
10th %ile	439	894	36	42	1	1	0.00	0.00	0.00	0.00
90th %ile	4045	5400	568	740	5	4	1.00	1.00	0.43	0.19
Maximum	5000	7250	950	1063	7	8	1.00	1.00	1.00	0.67
Standard deviation	1332	1692	225	265	1.58	1.49	0.50	0.41	0.24	0.11
Skewness	0.55	0.27	1.45	0.72	0.88	1.30	0.18	1.39	2.68	2.98
	% Muslims (residents)		% non-native (residents)		% non-native (leaders)		% in dominant group		% ex-combatants (residents)	
Mean	0.12	0.12	0.13	0.29	0.02	0.03	0.82	0.87	0.09	0.08
Median	0.00	0.00	0.10	0.25	0.01	0.02	0.90	0.90	0.05	0.05
Minimum	0.00	0.00	0.00	0.00	0.00	0.00	0.25	0.35	0.00	0.00
10th %ile	0.00	0.00	0.00	0.10	0.00	0.00	0.50	0.67	0.00	0.00
90th %ile	0.60	0.50	0.26	0.50	0.04	0.07	1.00	1.00	0.20	0.18
Maximum	1.00	1.00	0.90	1.00	1.00	0.39	1.00	1.00	0.56	0.35
Standard deviation	0.28	0.28	0.13	0.18	0.07	0.05	0.21	0.14	0.09	0.07
Skewness	2.43	2.47	2.29	1.70	13.31	4.21	-0.93	-1.55	1.45	0.80
	% ex-combatants (leaders)		% returned from internal displacement		% under 30		% male		Mean educational attainment	
Mean	0.02	0.02	0.58	0.36	0.20	0.27	0.56	0.48	5.17	5.54
Median	0.01	0.01	0.70	0.31	0.20	0.25	0.55	0.48	5.05	5.63
Minimum	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.10	0.00	0.90
10th %ile	0.00	0.00	0.02	0.01	0.05	0.10	0.35	0.32	1.74	2.58
90th %ile	0.04	0.04	0.98	0.83	0.40	0.45	0.79	0.65	8.55	8.20
Maximum	1.00	0.15	1.00	1.00	0.65	0.65	1.00	0.80	11.25	11.60
Standard deviation	0.07	0.02	0.37	0.31	0.15	0.13	0.17	0.13	2.57	2.06
Skewness	14.06	3.28	-0.47	0.53	0.72	0.31	-0.19	-0.29	0.13	-0.08
	% with no education		% receiving any peace education		Group participation		% who contribute to public facilities		% saying town is safe at night	
Mean	0.45	0.42	0.28	0.34	3.74	3.69	0.86	0.89	0.52	0.43
Median	0.45	0.40	0.25	0.30	3.69	3.70	0.90	0.90	0.53	0.45
Minimum	0.00	0.00	0.00	0.00	0.00	1.91	0.13	0.56	0.00	0.00
10th %ile	0.15	0.18	0.05	0.15	2.15	2.70	0.65	0.76	0.17	0.15
90th %ile	0.80	0.70	0.55	0.60	5.20	4.70	1.00	1.00	0.80	0.68
Maximum	1.00	0.89	1.00	0.80	7.25	5.75	1.00	1.00	1.00	0.95
Standard deviation	0.23	0.19	0.19	0.18	1.22	0.78	0.16	0.09	0.24	0.20
Skewness	0.21	0.40	0.70	0.51	0.06	0.14	-1.71	-0.90	-0.28	0.08
	% saying neighbors are helpful		Capacity for collective action		% who rely on NGOs		% who rely on government		Perceived equity in institutions	
Mean	0.70	0.50	1.79	1.61	0.53	0.53	0.14	0.17	2.60	2.37
Median	0.72	0.50	1.00	1.00	0.53	0.52	0.10	0.15	2.65	2.40
Minimum	0.00	0.06	0.00	0.00	0.05	0.15	0.00	0.00	1.61	1.06

Table A.5: Detailed summary statistics for predictors

Year	2008	2010	2008	2010	2008	2010	2008	2010	2008	2010
10th %ile	0.50	0.30	0.00	0.00	0.25	0.30	0.00	0.05	2.20	1.95
90th %ile	0.90	0.70	4.00	3.00	0.80	0.76	0.35	0.33	2.95	2.71
Maximum	1.00	0.90	7.00	6.00	1.00	0.95	0.50	0.47	3.00	2.95
Standard deviation	0.17	0.16	1.60	1.28	0.20	0.17	0.13	0.11	0.29	0.32
Skewness	-0.70	0.02	0.80	0.76	-0.07	0.08	0.84	0.67	-0.80	-0.82

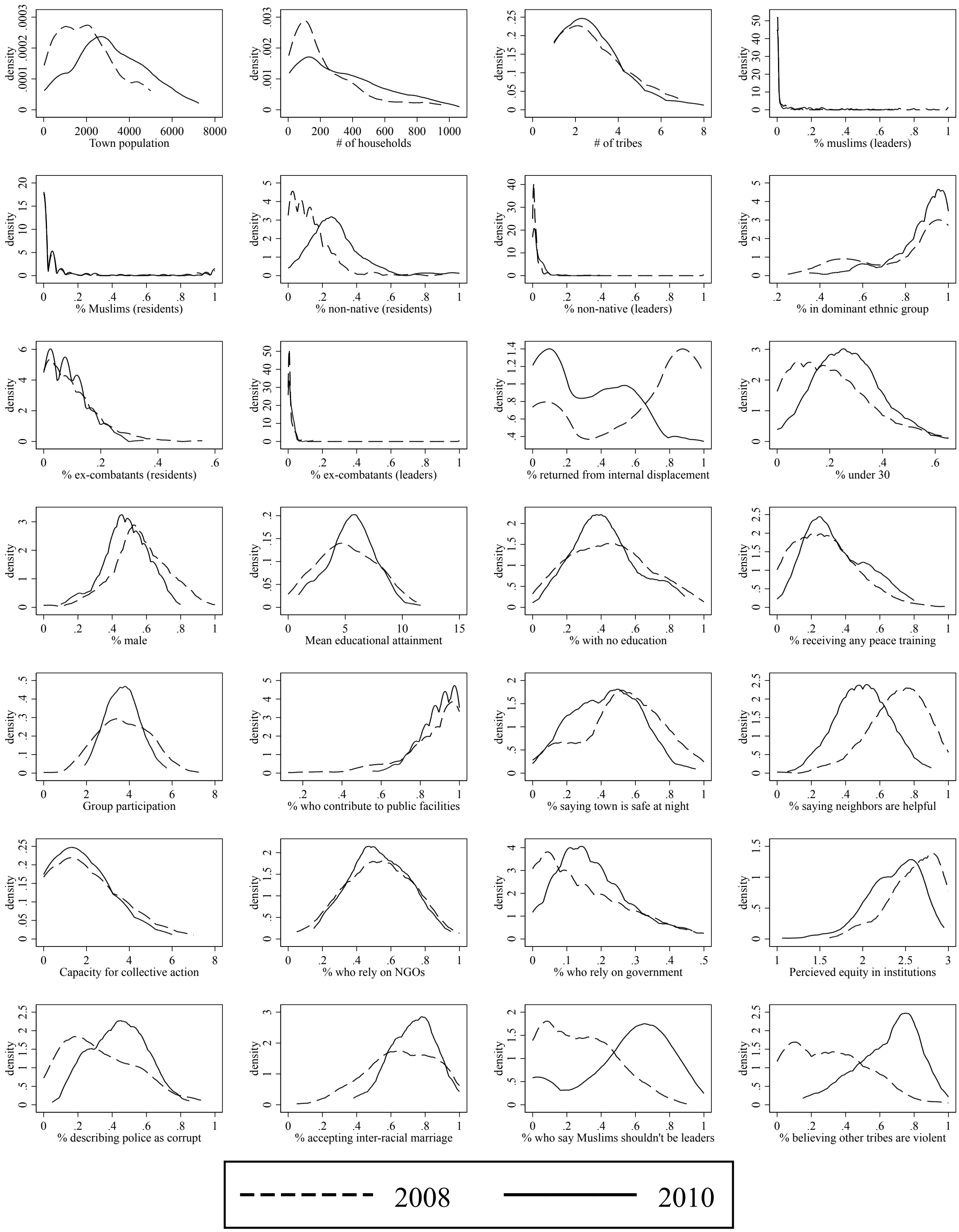
Table A.5: Detailed summary statistics for predictors

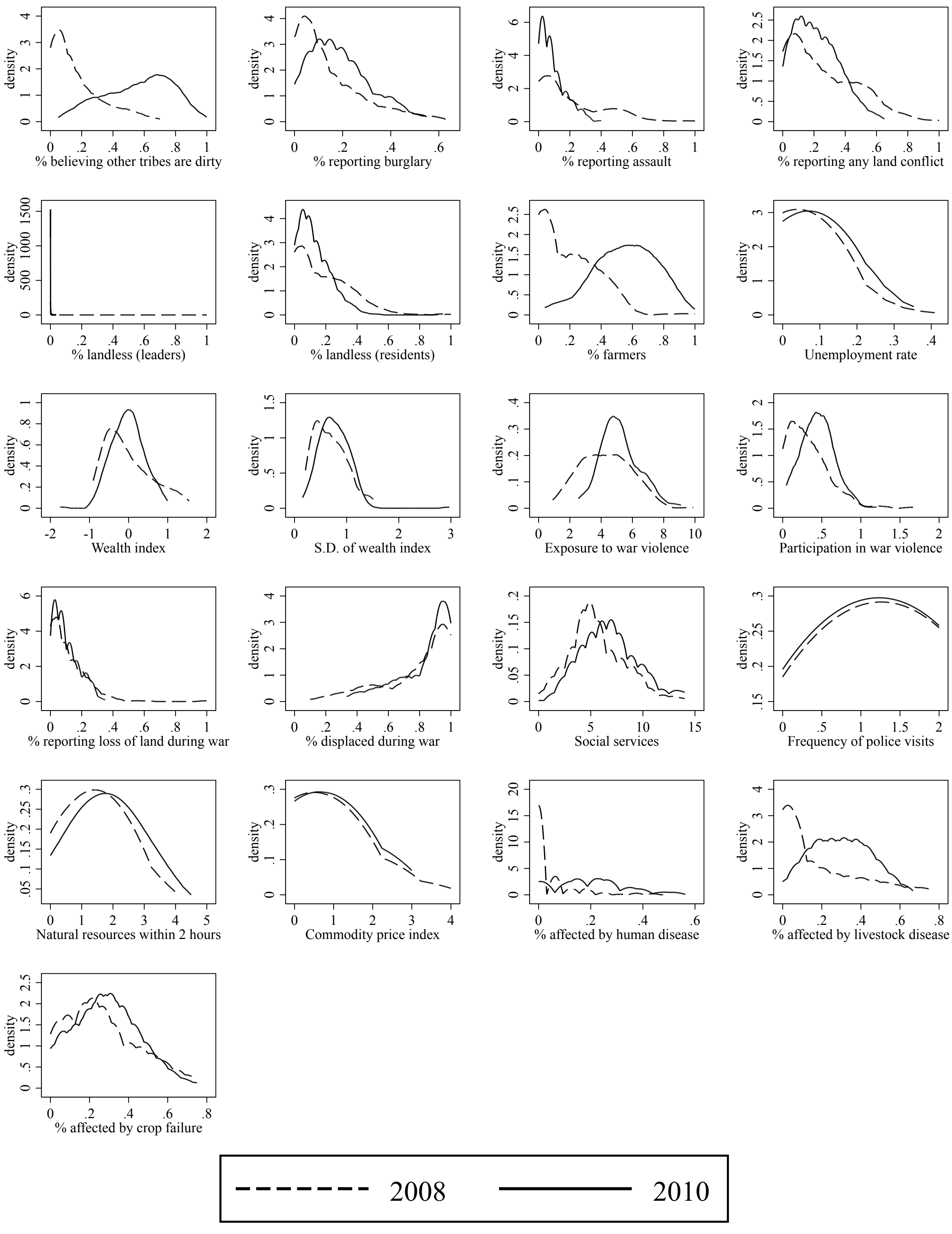
Year	2008	2010	2008	2010	2008	2010	2008	2010	2008	2010
	% describing police as corrupt		% accepting inter-racial marriage		% who say Muslims shouldn't be leaders		% believing other tribes are violent		% believing other tribes are dirty	
Mean	0.33	0.44	0.66	0.73	0.27	0.55	0.30	0.65	0.15	0.56
Median	0.30	0.45	0.67	0.75	0.25	0.60	0.30	0.70	0.10	0.60
Minimum	0.00	0.05	0.05	0.38	0.00	0.00	0.00	0.15	0.00	0.05
10th %ile	0.07	0.20	0.39	0.55	0.00	0.05	0.05	0.38	0.00	0.25
90th %ile	0.63	0.63	0.90	0.89	0.56	0.84	0.60	0.85	0.40	0.81
Maximum	0.94	0.85	1.00	1.00	0.90	1.00	1.00	1.00	0.70	1.00
Standard deviation	0.22	0.16	0.20	0.13	0.21	0.27	0.22	0.18	0.16	0.22
Skewness	0.57	0.05	-0.34	-0.27	0.48	-0.72	0.51	-0.62	1.26	-0.33
	1 if a (any) minority tribe has representation in leadership		% reporting burglary		% reporting assault		% reporting any land conflict		Any violent event (lagged DV)	
Mean	0.59	0.84	0.13	0.19	0.19	0.08	0.25	0.21	0.37	0.17
Median	1.00	1.00	0.10	0.16	0.10	0.05	0.20	0.20	0.00	0.00
Minimum	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
10th %ile	0.00	0.00	0.00	0.05	0.00	0.00	0.00	0.05	0.00	0.00
90th %ile	1.00	1.00	0.35	0.40	0.50	0.20	0.55	0.40	1.00	1.00
Maximum	1.00	1.00	0.65	0.55	1.00	0.40	1.00	0.65	1.00	1.00
Standard deviation	0.49	0.36	0.14	0.13	0.21	0.08	0.22	0.14	0.48	0.38
Skewness	-0.35	-1.89	1.35	0.68	1.26	1.12	0.78	0.53	0.53	1.72
	% landless (leaders)		% landless (residents)		% farmers		Unemployment rate		Wealth index	
Mean	0.01	0.00	0.17	0.12	0.18	0.55	0.04	0.07	-0.02	-0.02
Median	0.00	0.00	0.11	0.10	0.15	0.55	0.00	0.05	-0.18	-0.02
Minimum	0.00	0.00	0.00	0.00	0.00	0.04	0.00	0.00	-0.91	-1.75
10th %ile	0.00	0.00	0.00	0.00	0.00	0.30	0.00	0.00	-0.68	-0.54
90th %ile	0.02	0.00	0.40	0.25	0.45	0.80	0.15	0.16	0.93	0.57
Maximum	1.00	0.03	1.00	0.95	1.00	1.00	0.42	0.35	1.55	0.99
Standard deviation	0.07	0.00	0.17	0.11	0.18	0.20	0.07	0.07	0.60	0.42
Skewness	12.96	13.28	1.11	2.16	0.93	-0.28	2.11	1.01	0.76	-0.11
	S.D. of wealth index		Exposure to war violence		Participation in war violence		% reporting loss of land during war		% displaced during war	
Mean	0.69	0.76	4.28	5.19	0.31	0.45	0.10	0.09	0.80	0.84
Median	0.64	0.74	4.30	5.05	0.25	0.45	0.05	0.05	0.89	0.90
Minimum	0.21	0.16	0.90	2.55	0.00	0.05	0.00	0.00	0.10	0.33
10th %ile	0.32	0.42	2.20	3.85	0.00	0.16	0.00	0.00	0.45	0.55
90th %ile	1.11	1.11	6.50	6.95	0.63	0.70	0.25	0.21	1.00	1.00
Maximum	1.51	2.97	9.89	9.10	1.67	1.20	1.00	0.35	1.00	1.00
Standard deviation	0.31	0.30	1.64	1.25	0.27	0.21	0.12	0.08	0.23	0.17
Skewness	0.58	1.60	0.18	0.53	1.31	0.28	2.50	0.79	-1.17	-1.21

Table A.5: Detailed summary statistics for predictors

Year	2008	2010	2008	2010	2008	2010	2008	2010	2008	2010
	Social services		Magistrate or police station in town		Frequency of police visits		Over 1 hr. from nearest road (leaders)		Cell coverage in town	
Mean	5.61	6.81	0.19	0.18	1.26	1.23	0.14	0.03	0.58	0.74
Median	5.00	7.00	0.00	0.00	1.00	1.00	0.00	0.00	1.00	1.00
Minimum	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
10th %ile	2.00	3.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
90th %ile	9.00	10.00	1.00	1.00	2.00	2.00	1.00	0.00	1.00	1.00
Maximum	14.00	14.00	1.00	1.00	2.00	2.00	1.00	1.00	1.00	1.00
Standard deviation	2.64	2.77	0.39	0.38	0.81	0.75	0.35	0.18	0.49	0.44
Skewness	0.53	0.43	1.61	1.69	-0.51	-0.41	2.02	5.22	-0.32	-1.07
	1 or fewer radio stations in town		Natural resources within 2 hours		Commodity price index		% affected by human disease		% affected by livestock disease	
	0.87	0.04	1.44	1.77	0.60	0.69	0.04	0.19	0.16	0.30
Median	1.00	0.00	1.00	1.75	0.00	0.50	0.00	0.19	0.06	0.31
Minimum	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
10th %ile	0.00	0.00	0.75	0.75	0.00	0.00	0.00	0.00	0.00	0.10
90th %ile	1.00	0.00	2.50	2.75	2.00	2.00	0.13	0.38	0.50	0.50
Maximum	1.00	1.00	4.00	4.50	4.00	3.00	0.50	0.56	0.75	0.67
Standard deviation	0.34	0.20	0.75	0.83	0.88	0.81	0.09	0.14	0.21	0.15
Skewness	-2.17	4.61	0.92	0.17	1.52	0.89	2.81	0.36	1.29	0.09
	% affected by crop failure									
	0.26	0.29								
Median	0.25	0.28								
Minimum	0.00	0.00								
10th %ile	0.00	0.06								
90th %ile	0.56	0.50								
Maximum	0.75	0.75								
Standard deviation	0.20	0.17								
Skewness	0.56	0.28								

Figure A.2: Predictor density plots




density
% believing other tribes are dirty
% reporting burglary
% reporting assault
% reporting any land conflict
% landless (leaders)
% landless (residents)
% farmers
Unemployment rate
Wealth index
S.D. of wealth index
Exposure to war violence
Participation in war violence
% reporting loss of land during war
% displaced during war
Social services
Frequency of police visits
Natural resources within 2 hours
Commodity price index
% affected by human disease
% affected by livestock disease
% affected by crop failure
2008
2010

Sampling error in risk factors

In the paper we report a ranking of predictors in our models. A majority of these predictors come from a survey of 20 residents per town, and variables with especially high sampling error might be disproportionately penalized, especially in the lasso model. We examine relative sampling error by variable in Table **??**. We bootstrapped standard errors for each variable for each town, repeatedly resampling the 20 observations, with replacement. We then report the average rescaled bootstrapped standard error. We repeat this process for standardized versions of the predictors (which we use in the forecasting models) and non-standardized versions as well. In Table **??** the three resident survey variables selected by the main lasso model are in bold.

We see that the lasso does tend to favor variables with lower sampling variability. The lasso does not, however, select any of the low-variability economic variables (such as wealth and landlessness), which is consistent with our interpretation in the paper and suggests that economic risk factors are relatively less important in these towns. Moreover, some of the variables with the lowest sampling variability are not highly ranked in any model (e.g. percent Muslim), which suggests that sampling error alone is not driving our results.

Models

This section describes in more detail the estimation methods for each prediction model, summarizing the generic method and highlighting the specific modeling choices we made in each case.[20]

Lasso

Given some dataset $(\mathbf{x_i}, y_i)$ where $\mathbf{x_i}$ denotes a set of j standardized predictor variables and y_i a vector of responses, the lasso coefficients β are given by:

$$l\left(\beta\right) = -\sum_{i=1}^{N} \ln\left(1 + \exp\left(-\beta^{\mathbf{T}}\mathbf{x_i}y_i\right)\right) + \lambda\sum_{j=1}^{k}|\beta_j|$$

where the first expression on the right hand side is a standard maximum likelihood estimator, and the second is the penalty function specific to lasso. $\lambda \geq 0$ is a tuning parameter that controls the degree of coefficient shrinkage; the coefficients on poor performers are forced to 0 and thus dropped from the model. A ridge regression looks similar, except that the penalty is $\lambda\sum_{j=1}^{k}\beta_j{}^2$. The key difference between lasso and ridge regression is that the latter assigns non-zero coefficients to all predictors, though the coefficients on poorly performing indicators can be very small. Ridge regression thus performs coefficient shrinkage only, while lasso performs both coefficient shrinkage and variable selection, and so generally produces more parsimonious models. We opt for an elastic net, which uses a penalty that is a weighted sum of the lasso and ridge penalty:

[20] There are several online resources that provide introductions to machine learning for newcomers, including the Hopkins Practical Machine Learning online course, `https://www.coursera.org/course/predmachlearn`, and the StackExchange question and answer site, `http://stats.stackexchange.com/`.

Table A.6: Relative sampling variability of predictors (resident survey variables only)

	Standardized		Not Standardized	
Covariate	Rank	Mean Std. Err.	Rank	Mean Std. Err.
% male	1	0.63	9	0.11
Participation in war violence	2	0.58	5	0.15
% saying neighbors are helpful	3	0.55	15	0.09
% ex-combatants (residents)	4	0.54	32	0.05
% under 30	5	0.53	19	0.08
% who rely on NGOs	6	0.51	10	0.10
% who rely on government	7	0.49	21	0.06
% accepting inter-racial marriage	8	0.48	14	0.10
% describing police as corrupt	9	0.47	11	0.10
% non-native (residents)	10	0.46	26	0.06
% receiving any peace education	11	0.46	16	0.09
% with no education	12	0.43	12	0.10
Perceived equity in institutions	13	0.42	8	0.12
Unemployment rate	14	0.42	33	0.03
% saying town is safe at night	15	0.42	13	0.10
% reporting loss of land during war	16	0.41	31	0.05
% reporting burglary	17	0.40	28	0.06
S.D. of wealth index	18	0.40	7	0.12
% who say Muslims shouldn't be leaders	19	0.39	18	0.08
Exposure to war violence	20	0.39	2	0.63
% believing other tribes are violent	**21**	**0.38**	**17**	**0.09**
Mean educational attainment	22	0.38	1	0.98
% believing other tribes are dirty	23	0.37	25	0.06
% landless (residents)	24	0.36	24	0.06
% who contribute to public facilities	**25**	**0.36**	**29**	**0.06**
% farmers	26	0.34	23	0.06
% reporting any land conflict	27	0.33	20	0.07
Group participation	28	0.30	4	0.37
% reporting assault	29	0.28	27	0.06
% displaced during war	30	0.28	22	0.06
% in dominant group	**31**	**0.27**	**30**	**0.06**
Wealth index	32	0.25	6	0.15
# of tribes	33	0.24	3	0.38
% Muslims (residents)	34	0.08	34	0.02

Bolded rows represent individual-level variables selected by the lasso model in the paper.

$$l(\beta) = -\sum_{i=1}^{N} \ln\left(1+\exp\left(-\beta^{\mathbf{T}}\mathbf{x_i}y_i\right)\right)+\lambda\left(\alpha\sum_{j=1}^{k}|\beta_j| + (1-\alpha)\sum_{j=1}^{k}\beta_j^2\right)$$

We use a variation known as "elastic net optimization." that involves a scalar α, which regulates the weight given to lasso ($\alpha = 1$) versus ridge ($\alpha = 0$) optimization. In our preferred model we set $\alpha = .95$, thus weighting the lasso penalty much more strongly than the ridge. We use a modification of lasso analogous to logit in order to accommodate our binary dependent variable. Thus, for a given observation, our model generates a predicted probability of violence between 0 and 1. We then classify each observation as 0 or 1 (violence or no violence) according to a discrimination threshold that is chosen by cross-validation to maximize sensitivity, keeping accuracy above 50%.

Our training procedure is as follows. First, we split the sample into five subsets, or folds. We then train a lasso model on four of the five folds. This is the initial training set. The lasso is fit over a sequence of 80 lambdas in the training data, in effect producing 80 lasso models, each with a different lambda (and thus a different vector of coefficients). The lambda that maximizes sensitivity while maintaining accuracy above 50% in the training set is then applied to the test set. We iterate this process over the five possible combinations of folds into training and test sets. This is one cross-validation. We then repeat this process 200 times and calculate the average optimal lambda across these 200 cross-validationsâĂŤ80,000 regressions in total. Finally, we repeat the cross-validation procedure 200 additional times, this time applying the average optimal lambda to every model. We calculate performance metrics within each of these 200 trials, then report the average of each metric.

Random forests

Given some dataset $(\mathbf{x_i}, y_i)$, a regression tree sorts observations into leaves and makes a prediction, $\hat{y}$, for each leaf. Trees are constructed stepwise. Initially, all observations are on the same leaf. The observations are then divided into two leaves based on values of one of the k predictors, so that the sum of squared deviations from the mean in each leaf is minimized. More formally, we minimize:

$$MSE = \sum_{l=1}^{2}\left(\sum_{i=1}^{N_l}(y_i - \bar{y}_j)^2\right)$$

where $\bar{y}_j$ is the average outcome in leaf j and N_j is the number of observations in leaf j. In the next step, each of the these leaves is split again based on the predictor that most reduces the sum of squared deviations from the mean in the leaf (this could be the same predictor that was chosen in the first step). In principle, this process could continue until all leaves contain only observations of the same value (and $MSE = 0$). However, researchers typically employ some sort of stopping criteria before that happens. In our case, we set the maximum number of nodes to be 5. Because we use regression trees rather than classification trees, each observation is assigned a predicted probability rather than a binary (0/1) prediction.

Random forests are comprised of many trees fit to random subsets of the data with random subsets of predictors available for splitting. Each tree generates a distinct predicted probability for each observation, and the prediction for the entire random forest model is just the average of the predictions of each tree in the forest.

For our random forests model, we grow 1,000 trees with a maximum of 5 terminal nodes each and $\sqrt{56}$ variables sampled (without replacement) at each node.

Neural networks

Neural networks are layered systems of weighted sums of predictor variables with a final weighted sum mapped into the prediction space. In order to control model complexity, practitioners specify the number of layers and the number of weighted sums (called nodes) that comprise each layer. Our model has one layer and 5 nodes, and a weight decay of 0.1 with randomly selected near-zero starting values. Using five different sets of weights, the 56 predictors plus a constant are mapped onto each of the 5 nodes. Then, these five nodes plus another constant are mapped, by some linear combination of weights, to a scalar. Finally, this scalar is mapped by a logistic function to the interval $[0, 1]$, our prediction space. Hence, our network is defined by 287 weights $(56 \times 5 + 5 + 2)$. they are initially chosen at random, and then tuned iteratively to minimize the mean-squared error in the prediction space. The net is trained via back-propagation.
More specifically, a neural network is a two-stage regression or classification model, typically represented as a âĂIJnetwork diagramâĂİ with K units at the top; the kth unit models the probability of class k. In our classification model $k = 1$ and the response $Y_{k=1}$ is simply a binary variable.
Neural networks capture interactivity by generating âĂIJderived features,âĂİ denoted Z_m, from linear combinations of the predictors, then modeling the response as a function of linear combinations of the derived features:

$$Z_m = \sigma(\alpha_{0m} + \alpha_m^T X), m = 1, ..., M$$

$$T_k = \beta_{0k} + \beta_k^T Z, k = 1, ..., K$$

$$f_k(X) = g_k(T), k = 1, ..., K$$

where $Z = \{Z_1, ..., Z_m\}, T = \{T_1, ..., T_k\}, \sigma(\nu)$ is an initial non-linear transformation of the predictors, and $g_k(T)$ is a final transformation of the output vector T. For the special case where $\sigma = 1$, the network collapses to a linear model. For a more thorough explanation of neural networks and their analogies to maximum likelihood, see **?**.

Supplemental tables

Reconciliation of original to current models

The original lasso model results, calculated before the 2012 data collection, are presented in Column 1 of Table **??**. A small number of relatively minor technical changes were made after the 2012 data were collected, not with an eye to improving model performance but to correct small errors or adhere to expert recommendations. Table **??** details each change in turn, including the cumulative effect on predictions and risk factors.
First, we switched from the *lars* package to the *glmnet* package in R (Column 2). Second, we estimated cross-validated forecast errors by applying a single set of optimal parameters

across 200 cross-validated trials, rather than estimating the error using a varied set of optimal parameters identified within each trial (Column 2). Finally, we standardized dummy variables, which we had not been doing in previous models. These changes had little material effect on model performance or risk factor rankings in the cross-validated forecasts.

Robustness to alternate model parameters

Lasso

Table **??** reports various robustness checks for the cross-validated forecasts (Panel A), true forecasts (Panel B), and corresponding risk factors and rankings (Panel C), limiting to the latter to the top five factors only. The columns are as follows:

1. *Main specification:* From Tables 3 and 4 in main paper.
2. *New seed:* We specify an alternate randomization seed for the selection of folds which are used to estimate model parameters and error rates using 2010 outcomes.
3. *Dummies not standardized:* We keep binary predictors on a (0,1) scale rather than standardizing them to have mean 0 and standard deviation 1.
4. *10-fold cross validation:* We identify optimal parameters and estimate forecast error rates using 10-fold cross validation rather than 5-fold cross validation.
5. $\alpha = 1$: α is the weight placed on the lasso penalty (sum of coefficient magnitudes) relative to the ridge penalty (sum of squared coefficients). In the paper we use $\alpha = .95$. When $\alpha = 1$, we have a pure lasso penalty.
6. $\alpha = .5$: sets the penalty to be half-way between a lasso and ridge penalty.
7. *Subset (30) from OLS:* We first fit an OLS model to the 2008/2010 data to determine the 30 coefficients of greatest magnitude. We then use only those 30 predictors for the model.
8. *Nonlinear transformations of skewed variables:* Several covariates have a high skew, and it is possible that nonlinear transformations could improve performance or change the relevant risk factors. We take all variables with a skew greater than 1 and use their natural logarithm (or the logarithm of one plus the variable if the range includes zero).

Random forests

Table **??** reports various robustness checks for the cross-validated forecasts (Panel A), true forecasts (Panel B), and corresponding risk factors and rankings (Panel C), limiting the latter to the top five factors only. The columns are as follows:

Table A.7: Reconciliation of original to current lasso predictions

Performance metric	Original Model (1)	Change in cross validation (2)	Standardized indicators (3)
AUC		0.56	0.58
True positives (sensitivity)	69%	67%	77%
True negatives (specificity)	49%	44%	41%
Overall accuracy	52%	48%	47%
Ratio of false + to true +	3.98	4.00	3.68
Ratio of false - to true +	0.52	0.50	0.31

(a) Cross-validated forecast (2010)

Performance metric	Original Model (1)	Change in cross validation (2)	Standardized indicators (3)
AUC		0.66	0.65
True positives (sensitivity)		85%	88%
True negatives (specificity)		35%	23%
Overall accuracy		43%	33%
Ratio of false + to true +		3.88	4.46
Ratio of false - to true +		0.18	0.14

(b) 2012 Forecasts

Performance metric	Original Model (1)	Change in cross validation (2)	Standardized indicators (3)
Minority tribe in leadership	1	1	1
Town population	2	9	2
% in largest tribe	3	6	4
% Muslim	4	10	
% reporting armed robbery or burglary	5		
% contributing to public facilities	6	4	5
# of tribes	7		
% reporting loss of land during war	8		
Number of resources available	9	8	
Wealth index	10	14	
% farmers		2	
% believing other tribes are violent		3	3
Participation in war violence		5	
Frequency of police visits		7	

(c) Risk Factor Rankings

Table A.8: Lasso model robustness checks

(a) Cross-validated forecast (2010)

Performance metric	Main Specification	New seed	Dummies not standardized	10-Fold Cross-Validation	$\alpha = 1$	$\alpha = .5$	Subset (30) from OLS	Nonlinear trans.
AUC	0.58	0.58	0.56	0.58	0.58	0.57	0.61	.57
Brier score	0.14	0.14	0.16	0.14	0.14	0.14	0.14	0.14
True positives (sensitivity)	77%	75%	68%	80%	76%	86%	83%	78%
True negatives (specificity)	40%	40%	43%	42%	40%	24%	30%	36%
Overall accuracy	47%	46%	47%	49%	46%	35%	39%	44%
Ratio of false + to true +	3.71	3.84	4.06	3.45	3.82	4.23	4.03	3.89
Ratio of false - to true +	0.30	0.33	0.49	0.25	0.33	0.17	0.20	0.29

Performance metric	Main Specification	New seed	Dummies not standardized	10-Fold Cross-Validation	$\alpha = 1$	$\alpha = .5$	Subset (30) from OLS	Nonlinear trans.
AUC	0.65	0.66	0.66	0.61	0.65	0.67	0.68	0.66
Brier score	0.13	0.13	0.13	0.14	0.13	0.13	0.13	0.13
True positives (sensitivity)	88%	88%	88%	88%	88%	90%	88%	88%
True negatives (specificity)	22%	25%	34%	17%	25%	27%	24%	24%
Overall accuracy	33%	36%	43%	29%	35%	37%	35%	34%
Ratio of false + to true +	4.49	4.31	3.83	4.77	4.34	4.11	4.37	4.40
Ratio of false - to true +	0.14	0.14	0.14	0.14	0.14	0.11	0.14	0.14

(b) 2012 Forecasts

Performance metric	Main Specification	New seed	Dummies not standardized	10-Fold Cross-Validation	$\alpha = 1$	$\alpha = .5$	Subset (30) from OLS	Nonlinear trans.
Minority tribe in town leadership	1	1	1	1	1	1	1	1
Town population	2	2	9	2	2	2		2
% believing other tribes are violent	3	3	3	3	3	3	2	3
% in dominant group	4	4	6	4	4	5	4	4
% who contribute to public facilities	5	5	4		5	4	3	6
% of town farmers			2				7	
Participation in war violence			5				11	

(c) Risk factor rankings

1. *Main specification:* From Tables 3 and 4 in main paper.

2. *New seed:* We specify an alternate randomization seed for the selection of folds which are used to estimate parameters and forecast error rates using 2010 outcomes.

3. *Classification:* Observations on a given leaf are classified by majority vote (with a weight given to positive votes that is chosen by cross-validation). For all other specifications, observations on a given leaf are assigned a predicted probability of violence and are then classified based on a discrimination threshold.

4. *10-fold cross validation:* We identify optimal parameters and estimate forecast error rates using 10-fold cross validation rather than 5-fold cross validation.

5. *10 nodes:* We limit each tree in the forest to have no more than 10 nodes. For all other models, we limit trees to 5 nodes.

6. *Trees fit to larger sample:* We fit each tree to a random sample of 36 observations rather than 24, the sample size for all other models.

7. *10,000 tree forests:* We compose the random forest from 10,000 trees rather than 1,000 trees as we do for all other models.

8. *Subset (30) from OLS:* We first fit an OLS model to the 2008/2010 data to determine the 30 coefficients of greatest magnitude. We then use only those 30 predictors for the model.

9. *Subset of 5 lasso variables:* We use only the 5 risk factors selected by the main lasso model, listed in the main paper.

Neural networks

Table **??** reports various robustness checks for the cross-validated forecasts (Panel A) and true forecast results (Panel B). The columns are as follows:

1. *Main specification:* From Tables 3 and 4.

2. *New seed:* We change the randomization seed to get different cross-validation runs and fit our models using different (randomly selected) initial weights.

3. *10-Fold Cross-Validation:* We choose a threshold and estimate forecast error using 10-fold cross-validation rather than 5-fold, as in our preferred model.

4. *Size = 10:* We use 10 nodes in our hidden layer rather than 5, as in our preferred model.

5. *Low decay:* We force our weights to decay at a rate of 0.01 rather than 0.1 as in our preferred model.

Table A.9: Random forests model robustness checks

(a) Cross-validated forecasts (2010)

Performance metric	Main specification	New seed	Classification	10-fold cross-validation	10 nodes	Trees fit to larger sample	10,000 tree forests	Subset (30) from OLS	Subset (5) from lasso
AUC	0.52	0.52		0.52	0.52	0.52	0.52	0.51	0.67
Brier score	0.15	0.15		0.15	0.15	0.15	0.15	0.15	0.14
True positives (sensitivity)	53%	54%	57%	53%	52%	54%	52%	49%	77%
True negatives (specificity)	50%	50%	48%	50%	50%	50%	50%	51%	45%
Overall accuracy	50%	50%	50%	50%	50%	50%	50%	50%	51%
Ratio of false + to true +	4.57	4.51	4.35	4.56	4.59	4.50	4.62	4.85	3.43
Ratio of false - to true +	0.92	0.88	0.77	0.91	0.93	0.88	0.94	1.06	0.31

Performance metric	Main specification	New seed	Classification	10-fold cross-validation	10 nodes	Trees fit to larger sample	10,000 tree forests	Subset (30) from OLS	Subset (5) from lasso
AUC	0.63	0.62		0.59	0.59	0.60	0.58	0.57	0.64
Brier score	0.14	0.12		0.12	0.12	0.12	0.12	0.13	0.12
True positives (sensitivity)	83%	82%	80%	85%	73%	88%	79%	82%	88%
True negatives (specificity)	31%	29%	37%	25%	28%	24%	30%	22%	29%
Overall accuracy	40%	36%	44%	33%	34%	33%	36%	30%	37%
Ratio of false + to true +	4.21	5.52	3.97	5.61	6.29	5.45	5.65	6.04	5.14
Ratio of false - to true +	0.21	0.22	0.25	0.18	0.38	0.14	0.27	0.22	0.14

(b) 2012 Forecasts

	Main specification	New seed	Classification	10-fold cross-validation	10 nodes	Trees fit to larger sample	10,000 tree forests	Subset (30) from OLS	Subset (5) from lasso
Town population	1	1	1	1	1	1	1		2
# of households	2	3	9	14	3	9	5		
Mean educational attainment	3	6	6	4	9	7	6	8	
# of tribes	4	2	2	5	7	5	3		
% reporting loss of land during war	5	4	5	3	2	2	2		
% in dominant group	6	5	3	2	4	3	4	1	5
S.D. of wealth index in town	7	10	13	22	18	26	12	9	
% believing other tribes are violent	8	9	39	30	12	16	15	14	4
Minority tribe in town leadership	9	8	8	10	15	25	8	7	1

(c) Risk factor rankings

6. *High decay:* We force our weights to decay at a rate of 0.5 rather than 0.1 as in our preferred model.
7. *Subset (30) from OLS:* We first fit an OLS model to the 2008/2010 data to determine the 30 coefficients of greatest magnitude. We then use only those 30 predictors for the model.
8. *Subset of 5 lasso variables:* We use only the 5 risk factors selected by the main lasso model, listed in the main paper.
9. *Nonlinear transformations of skewed variables:* Several covariates have a high skew, and it is possible that nonlinear transformations could improve performance or change the relevant risk factors. We take all variables with a skew greater than 1 and use their natural logarithm (or the logarithm of one plus the variable if the range includes zero).

Logit

While logit is advantageous in its simplicity and familiarity, it has a number of disadvantages for our purposes. In particular, while lasso, random forests and neural networks can accommodate many highly collinear predictors simultaneously, logit generally cannot. Because we have a large number of predictors relative to the number of observations, and because some of those predictors are highly collinear, our logit coefficients are unstable and therefore unreliable.

By way of illustration, Table **??** reports the variance inflation factor (VIF) for all logit coefficients in the cross-validated forecasts. The VIF measures the extent to which multicollinearity increases the variance of a given coefficient. When a regressor is completely orthogonal to the other regressors, the VIF is equal to one.

Column (1) shows that multicollinearity is indeed an issue in the logit model. The logit VIF is larger than five on 14 of our risk factors, larger than 10 on four of our risk factors, and as large as 39.86 on our measure of participation in wartime violence.

Lasso is designed to address precisely this issue. Column (2) reports the VIF for the five predictors assigned non-zero coefficients in the cross-validated lasso models. The largest lasso VIF is 1.13, which is still smaller than the smallest logit VIF (1.73). Indeed, the lasso VIFs are remarkably close to 1, suggesting that the algorithm successfully chooses predictors

Table A.10: Neural networks model robustness checks

(a) Cross-validated forecasts (2010)

Performance metric	Main specification	New seed	10-fold cross-validation	Size = 10	Low decay (.01)	High decay (.5)	Subset (30) from OLS	Subset (5) from lasso	Nonlinear trans.
AUC	0.56	0.56	0.56	0.55	0.55	0.59	0.64	0.68	0.55
Brier score	0.20	0.20	0.20	0.20	0.22	0.16	0.16	0.14	0.20
True positives (sensitivity)	62%	64%	63%	60%	59%	71%	72%	79%	61%
True negatives (specificity)	48%	48%	48%	49%	49%	46%	46%	44%	48%
Overall accuracy	51%	51%	50%	51%	51%	50%	50%	50%	50%
Ratio of false + to true +	3.99	3.91	3.96	4.11	4.18	3.65	3.58	3.37	4.08
Ratio of false - to true +	0.62	0.57	0.59	0.68	0.71	0.42	0.39	0.27	0.66

Performance metric	Main specification	New seed	10-fold cross-validation	Size = 10	Low decay (.01)	High decay (.5)	Subset (30) from OLS	Subset (5) from lasso	Nonlinear trans.
AUC	0.62	0.68	0.68	0.67	0.65	0.65	0.62	0.65	0.66
Brier score	0.17	0.16	0.16	0.17	0.19	0.14	0.16	0.13	0.16
True positives (sensitivity)	58%	70%	70%	65%	65%	68%	63%	70%	60%
True negatives (specificity)	53%	56%	57%	53%	61%	51%	50%	43%	51%
Overall accuracy	54%	58%	59%	55%	62%	54%	52%	48%	53%
Ratio of false + to true +	4.09	3.18	3.11	3.65	3.04	3.63	4.04	4.11	4.08
Ratio of false - to true +	0.74	0.43	0.43	0.54	0.54	0.48	0.60	0.43	0.67

(b) 2012 forecasts

that are relatively uncorrelated with one another. The result is a parsimonious model in which each predictor contains as much unique information as possible relative to the others.

Table A.11: Variance inflation factors for logit coefficients

	Logit	Lasso
	(1)	(2)
Town population	4.01	1.02
# of households	3.45	
# of tribes	5.71	
% Muslims (leaders)	9.88	
Has mosque	2.06	
% Muslims (residents)	9.10	
% non-native (residents)	2.65	
% non-native (leaders)	7.59	
% in dominant group	8.52	1.13
% ex-combatants (residents)	35.10	
% ex-combatants (leaders)	9.02	
% returned from internal displacement	2.55	
% under 30	2.33	
% male	2.99	
Mean educational attainment	19.76	
% with no education	16.53	
% receiving any peace education	1.88	
Group participation (0-9)	5.09	
% who contribute to public facilities	2.76	1.06
% saying town is safe at night	2.69	
% saying neighbors are helpful	2.45	
Collective public goods	3.55	
% who rely on NGOs	2.30	
% who rely on government	1.94	
Perceived equity in institutions	3.76	
% describing police/courts as corrupt	3.64	
% accepting inter-racial marriage	2.33	
% who say Muslims shouldn't be leaders	2.70	

% believing other tribes are violent	3.86	1.07
% believing other tribes are dirty	3.34	
Minority tribe in town leadership	1.73	1.11
% reporting burglary or robbery	2.96	
% reporting assault	4.33	
% reporting any land conflict	6.05	
Any violent event (lagged DV)	2.08	
% of town landless (leaders)	8.05	
% of town landless (residents)	2.78	
% of town farmers	4.18	
Unemployment rate	1.95	
Wealth index	7.30	
S.D. of wealth index in town	4.14	
Exposure to war violence	3.97	
Participation in war violence	39.86	
% reporting loss of land during war	3.45	
% displaced during war	2.45	
Social services in town	3.18	
Police or magistrate in town	1.91	
Frequency of police/NGO visits	2.45	
Town >1 hour from road	2.16	
Mobile phone coverage	1.75	
Less than 2 radio stations	1.74	
Natural resources in 2 hours	2.01	
Commodity price index	2.65	
% affected by human disease	2.00	
% affected by livestock disease	2.27	
% affected by crop failure	2.35	

Robustness to alternate coding rules

Recoding to reduce potential under-reporting

For most of our analysis, we code the dependent variable according to the following procedure:

1. Ask each of four local leaders about each of seven types of crime and violence over the past year.
2. For each type of violence, take the modal response across the four leaders.
3. If at least two leaders report that a given type of violence occurred over the past year, code the indicator for that type as 1.
4. Aggregate the seven indicators into three categories (capital crimes, collective violence and extrajudicial punishment).
5. Aggregate the seven indicators into a single indicator for any major destabilizing incident of crime or violence over the past year.

We tested the robustness of our results to two variations on this procedure. First, we aggregated the seven indicators into three categories before taking the modal response, then took the mode and aggregated the resulting three categories into a single indicator. Second, we aggregated the seven indicators into a single indicator before taking the modal response—i.e. if at least two leaders reported any kind of violence, then a violent event is coded for that community. Both of these variations increased the prevalence of violence in the sample. In 2010, for example, the first and second alternatives increased the prevalence of violence from 17% to 20% and 27%, respectively.

As we see in Table **??** and Figures **??** and **??** below, all of our models perform better under these alternative coding rules relative to the original. Our AUCs increase by 5 to 10 percentage points, and our false positive to true positive ratios decrease to below 3:1.

There are at least two possible explanations for this improvement. First, since rare events are generally more difficult to predict, we may have improved performance by making the dependent variable less rare. Second, the more restrictive coding rule systematically underestimated the prevalence of violence in ambiguous cases. For example, if in a given community one leader reported that a murder occurred and another reported that a rape occurred, our original coding rule would have assigned that community a 0 on the aggregate indicator, but the alternatives would both assign it a 1. If these more ambiguous cases are, for whatever reason, easier to predict than less ambiguous ones, then we may have improved performance by including them. The first explanation strikes us as more likely than the second, but regardless, the results in Table **??** suggest that, if anything, the coding rule we report in the paper may underestimate the predictive power of our models (at least relative to these two alternatives).

Recoding to reduce potential over-reporting

Table **??** reproduces Table 4 from the main paper, omitting the 7 ambiguously violent and malicious events. The AUC declines by a percentage point or two but the results are generally similar.

Table A.12: Alternative coding rule performances

(a) Cross-validated forecasts (2010)

	Lasso			Random forests			Neural networks			Logit		
Performance metric	Main coding rule	Modal category	Mode after aggre-gating	Main coding rule	Modal category	Mode after aggre-gating	Main coding rule	Modal category	Mode after aggre-gating	Main coding rule	Modal category	Mode after aggre-gating
AUC	0.58	0.61	0.67	0.52	0.56	0.62	0.56	0.59	0.58	0.52	0.55	0.58
Brier score	0.13	0.13	0.15	0.14	0.14	0.15	0.17	0.18	0.21	0.26	0.26	0.28
True positives (sensitivity)	77%	80%	48%	53%	62%	78%	62%	70%	71%	48%	59%	69%
True negatives (specificity)	40%	38%	89%	50%	47%	40%	48%	45%	43%	52%	49%	45%
Overall accuracy	47%	46%	33%	50%	50%	50%	51%	50%	50%	51%	51%	51%
Ratio of false + to true +	3.71	3.07	2.05	4.57	3.35	2.10	3.99	3.10	2.20	4.88	3.46	2.20
Ratio of false - to true +	0.30	0.25	0.13	0.92	0.61	0.28	0.62	0.44	0.41	1.15	0.74	0.46

	Lasso			Random forests			Neural networks			Logit		
Performance metric	Main coding rule	Modal category	Mode after aggre-gating	Main coding rule	Modal category	Mode after aggre-gating	Main coding rule	Modal category	Mode after aggre-gating	Main coding rule	Modal category	Mode after aggre-gating
AUC	0.65	0.73	0.73	0.63	0.69	0.74	0.62	0.67	0.67	0.67	0.67	0.66
Brier score	0.13	0.13	0.15	0.14	0.14	0.15	0.17	0.18	0.21	0.15	0.17	0.20
True positives (sensitivity)	88%	91%	96%	83%	98%	96%	58%	70%	75%	93%	95%	92%
True negatives (specificity)	22%	29%	32%	31%	24%	28%	53%	54%	56%	35%	31%	31%
Overall accuracy	33%	40%	45%	40%	37%	42%	54%	57%	60%	45%	43%	43%
Ratio of false + to true +	4.49	3.62	2.85	4.21	3.62	3.02	4.09	3.03	2.39	3.54	3.34	3.05
Ratio of false - to true +	0.14	0.10	0.04	0.21	0.02	0.04	0.74	0.43	0.33	0.08	0.05	0.09

(b) 2012 forecasts

Table A.13: Top predictors under alternative coding rules

(a) Lasso risk factors

	Main coding rule		Modal category		Mode after aggregating	
Risk factor	Ranking	Coef.	Ranking	Coef.	Ranking	Coef.
Minority tribe in town leadership	1	0.31	1	0.40	4	0.20
Town population	2	0.16	2	0.17	1	0.26
% believing other tribes are violent	3	0.08	4	0.12		
% in dominant group	4	-0.05	11	-0.01		
% who contribute to public facilities	5	0.01	9	0.03		
% reporting burglary or robbery			3	0.16	2	0.22
% saying town is safe at night			5	-0.10	3	-0.21
% Muslims (residents)			6	-0.09	7	-0.08
% Reporting Any Major Destabilizing Event			7	0.05	8	0.03
% with no education			8	-0.05	5	-0.16

	Main coding rule		Modal category		Mode after aggregating	
Risk factor	Ranking	Importance	Ranking	Importance	Ranking	Importance
Town population	1	0.0021	1	0.0018	1	0.0046
# of households	2	0.0008	17	0.0003	8	0.0014
% in dominant group	3	0.0007	19	0.0002	12	0.0006
% reporting loss of land during war	4	0.0007	16	0.0003	5	0.0019
Wealth index	5	0.0006	2	0.0011	6	0.0017
# of tribes	6	0.0005	14	0.0004	22	0.0002
% believing other tribes are dirty	7	0.0004	12	0.0005	13	0.0005
Group participation (0-9)	8	0.0003	9	0.0006	19	0.0003
% reporting assault	9	0.0003	15	0.0004	26	0.0000
Minority tribe in town leadership	10	0.0003	5	0.0007	14	0.0004

(b) Random forest risk factors

Figure A.3: ROC curves for simulated forecasts of 2010 violence using a less conservative coding rule

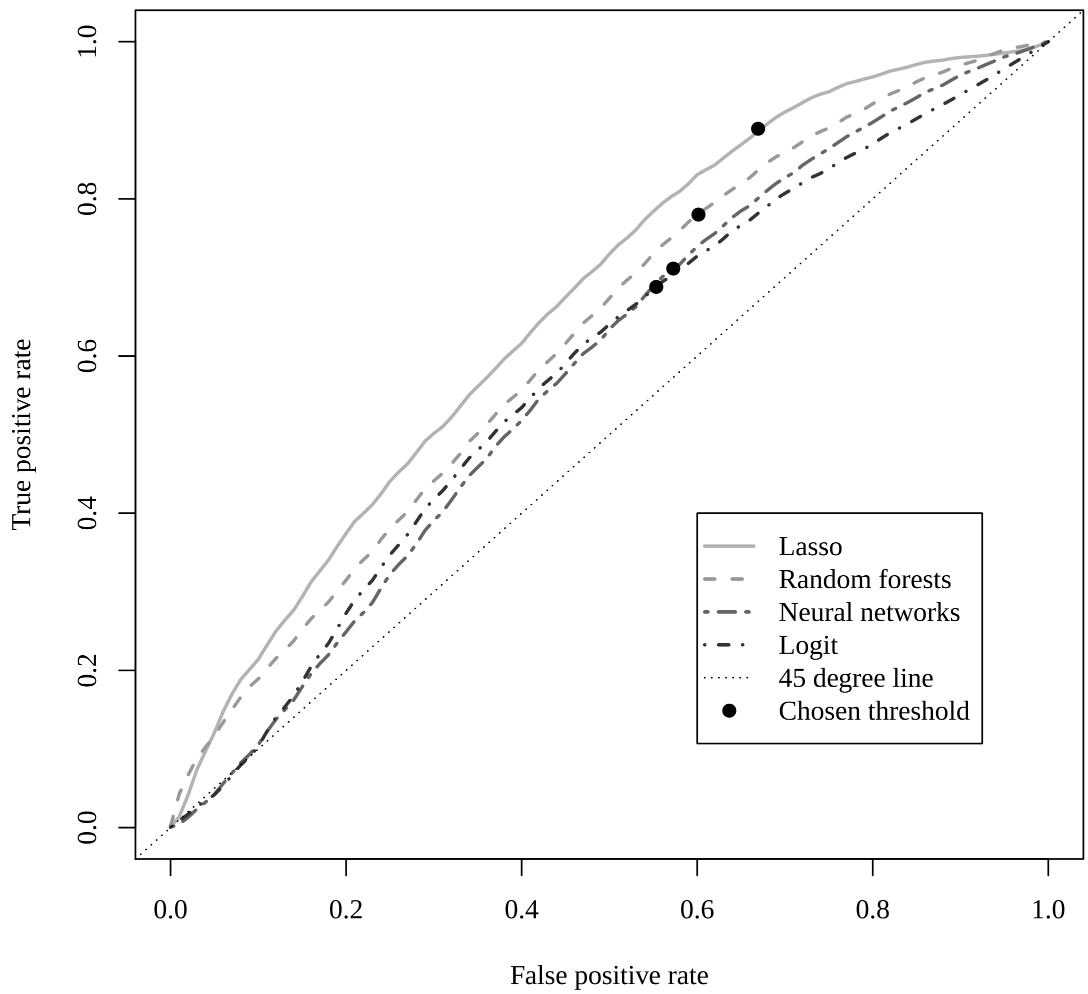


Figure A.4: ROC curves for true forecasts of 2012 violence using a less conservative coding rule

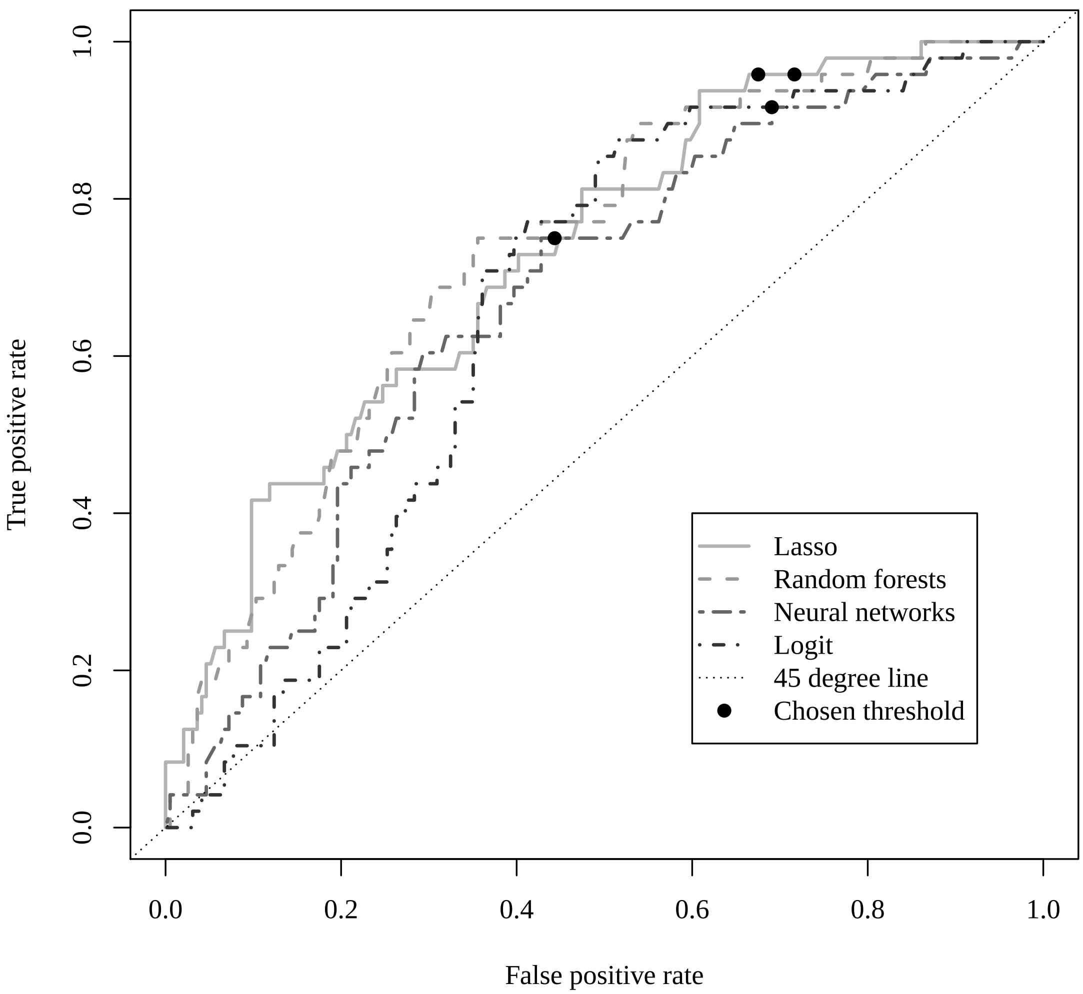


Table A.14: Performance of 2008–10 models in predicting 2012 aggregate violence, omitting non-violent or ambiguous events

Performance metric	Dependent variable: aggregate violence			
	Logit	Lasso	Random forests	Neural networks
AUC	0.66	0.65	0.60	0.64
	(0.05)	(0.05)	(0.06)	(0.05)
Brier score	0.14	0.12	0.12	0.15
True positives (sensitivity)	94%	88%	79%	64%
True negatives (specificity)	34%	22%	30%	52%
Overall accuracy	43%	31%	37%	54%
Ratio of false + to true +	4.42	5.62	5.62	4.76
Ratio of false - to true +	0.06	0.14	0.27	0.57

Disaggregating incidents by category

Table **??** reports the AUC and Brier score for each of our models when separately trained and tested on each of our three categories of local violence (collective, interpersonal and extrajudicial). The table also reports five performance metrics at our preferred predicted probability threshold, selected to maximize sensitivity while maintaining accuracy at or above 50%. The results are unstable when the dependent variable is disaggregated in this way, however, and should be interpreted with caution.

Model performance varies across time periods and dependent variables, and no model unambiguously dominates the others. Moreover, in some cases the model that performs best in terms of AUC performs worst in terms of Brier score (e.g. the cross-validated neural networks model for interpersonal violence). Nonetheless, overall the results suggest that we may be able to forecast even disaggregated violence with accuracy better than chance.

In some cases the performance of these models meets or exceeds that of our models designed to forecast aggregate violence. For example, while the neural networks model achieves an AUC of just 0.62 when forecasting aggregate violence in 2012, it achieves an AUC of 0.71 when forecasting interpersonal violence. Again, however, these disaggregated results are unstable, and caution is warranted when comparing them to the aggregated ones.

Table A.15: Cross-validated and true out-of-sample forecasts for disaggregated violence

	Collective violence (9)			Extrajudicial violence (8)			Interpersonal violence (32)		
Performance metric	Lasso	RF	NN	Lasso	RF	NN	Lasso	RF	NN
AUC	0.55	0.65	0.59	0.57	0.54	0.62	0.65	0.57	0.67
	(0.06)	(0.06)	(0.06)	(0.06)	(0.05)	(0.06)	(0.03)	(0.03)	(0.03)
Brier score	0.05	0.04	0.05	0.04	0.03	0.05	0.14	0.11	0.15
	(0.003)	(0)	(0.003)	(0.002)	(0)	(0.002)	(0.008)	(0.001)	(0.009)
True positives (sensitivity)	47%	73%	60%	44%	43%	29%	85%	62%	81%
	(0.14)	(0.12)	(0.13)	(0.16)	(0.16)	(0.11)	(0.06)	(0.05)	(0.06)
True negatives (specificity)	49%	50%	50%	46%	51%	51%	40%	49%	46%
	(0.03)	(0.02)	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Overall accuracy	49%	51%	51%	46%	51%	51%	46%	50%	51%
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Ratio of false + to true +	31.49	18.17	22.68	43.08	40.34	60.03	4.65	5.46	4.39
	(15.78)	(3.58)	(7.13)	(22.79)	(23.14)	(29.57)	(0.33)	(0.5)	(0.37)
Ratio of false - to true +	1.42	0.41	0.77	1.72	1.83	3.22	0.19	0.62	0.24
	(1.28)	(0.28)	(0.57)	(1.46)	(1.64)	(2.08)	(0.09)	(0.13)	(0.09)

(a) Cross-validated forecasts (2010)

	Collective violence (7)			Extrajudicial violence (21)			Interpersonal violence (16)		
Performance metric	Lasso	RF	NN	Lasso	RF	NN	Lasso	RF	NN
AUC	0.49	0.48	0.65	0.70	0.54	0.61	0.61	0.61	0.71
Brier score	0.03	0.03	0.04	0.06	0.06	0.07	0.11	0.08	0.11
True positives (sensitivity)	57%	86%	71%	88%	81%	69%	76%	71%	81%
True negatives (specificity)	45%	24%	56%	45%	23%	56%	34%	34%	47%
Overall accuracy	45%	26%	56%	48%	26%	57%	38%	38%	50%
Ratio of false + to true +	32.50	29.67	20.80	8.93	13.46	9.00	9.06	9.67	6.88
Ratio of false - to true +	0.75	0.17	0.40	0.14	0.23	0.45	0.31	0.40	0.24

(b) True forecasts (2012)

RF is random forests and NN is neural networks.

Table A.16: Cross-validated forecasts of 2010 violence using 2008 risk factors plus distance to nearest community with violence

	Aggregate dependent variable		
Performance metric	Lasso	Random forests	Neural networks
AUC	0.59	0.53	0.55
	(0.02)	(0.03)	(0.03)
Brier score	0.15	0.15	0.21
	(0.002)	(0.002)	(0.012)
True positive rate (sensitivity)	77%	55%	61%
	(0.05)	(0.05)	(0.06)
True negative rate (specificity)	41%	49%	48%
	(0.02)	(0.02)	(0.03)
Accuracy	47%	50%	50%
	(0.02)	(0.02)	(0.03)
Ratio of false + to true +	3.52	4.25	3.94
	(0.24)	(0.44)	(0.44)
Ratio of false - to true +	0.31	0.83	0.66
	(0.09)	(0.16)	(0.17)

Robustness to adjustments for spatial autocorrelation

Given the evidence of clustering above, we might expect that including a spatial lag of violence in nearby towns would improve predictive performance. We explore this possibility in two ways. First, we add a measure of distance to the nearest town with violence to our set of risk factors. Table **??** reports performance metrics for cross-validated forecasts including the spatial lag, and Table **??** reports performance metrics for the true forecasts. Table **??** provides a risk factor ranking. The lasso assigns the spatial lag a coefficient of 0, and the model's performance is unchanged. While the spatial lag is ranked 5th most important in the random forests model, the model's performance is again unchanged. Including the spatial lag does improve the true forecast AUC for the neural networks model, but does not change the Brier score.

As we discuss above, however, distance to the nearest town with violence incorporates

Table A.17: Forecasts of 2012 violence using 2010 risk factors plus distance to nearest community with violence

Performance metric	Aggregate dependent variable		
	Lasso	Random forests	Neural networks
AUC	0.64	0.63	0.69
	(0.05)	(0.05)	(0.05)
Brier score	0.13	0.13	0.16
True positive rate (sensitivity)	87%	79%	74%
True negative rate (specificity)	20%	32%	49%
Accuracy	31%	39%	53%
Ratio of false + to true +	4.73	4.47	3.54
Ratio of false - to true +	0.15	0.27	0.36

Table A.18: Rankings of risk factors by model plus distance to nearest community with violence

Risk factor	Lasso		Random forests	
	Rank	Coeff.	Rank	Importance
Minority tribe in town leadership	1	0.298	16	0.00018
Town population	2	0.141	1	0.00285
% believing other tribes are violent	3	0.054	10	0.00033
% in dominant group	4	-0.049	4	0.00070
% who contribute to public facilities	5	0.002	48	-0.00013
Mean educational attainment			2	0.00073
% reporting loss of land during war			3	0.00072
Distance to nearest community with violence			5	0.00058
% reporting any land conflict			6	0.00055
# of households			7	0.00050
% with no education			8	0.00048
# of tribes			9	0.00044

information about distance to the nearest town in general, and this latter distance may be an important correlate of violence in and of itself. To adjust for this possibility, we construct an alternate measure of nearby violence that does not change with a community's distance from other communities in general. For each town i, we construct the measure:

$$spillover_i = \frac{\sum_{j \in J} I(violence_j)/\delta_{i,j}}{\sum_{j \in J} 1/\delta_{i,j}}$$

where J is the set of all towns in the sample excluding i, $I(violence_j)$ is an indicator for past violence in community j, and $\delta_{i,j}$ is the distance between town i and town j.

Tables **??** and **??** report performance metrics for the cross-validated and true forecasts, respectively, including this alternate spatial lag. Table **??** provides the risk factor ranking. While lasso assigns a non-zero (positive) coefficient to this alternate measure, performance again remains largely unchanged across models. We conclude that incorporating spatial spillover may improve our models' performance in some cases, but probably not by much.

Model averaging

Rather than adjudicate between models, here we consider "ensemble" methods instead. One promising approach is Ensemble Bayesian Model Averaging (BMA). This method does not seem feasible in our case, however, as we do not have enough cross-sections of data to both train and calibrate our models. Also, it is not clear that BMA is as relevant to our problem, where we care more about sensitivity than accuracy. BMA weights are functions of log likelihoods, which are themselves functions of accuracy, not sensitivity.

A simpler "majority vote" method takes the binary predictions from each of the four models (logit, lasso, random forests, and neural networks) and generates a single prediction according to what the majority predicts. Since we have an even number of models, we code ties as a prediction of violence. This is consistent with our overall approach of erring on the side of sensitivity over specificity. This approach performs similarly to our best models.

An alternative ensemble method is to take predicted probabilities from the four models and generate a single predicted probability using a logistic regression model. That probability is then translated into a binary prediction using a discrimination threshold chosen by cross validation. This approach has the downside, like BMA, of giving equal weight to false negatives and false positives. This logistic "stack" performs surprisingly poorly.

Table A.19: Cross-validated forecasts of 2010 violence using 2008 risk factors plus a spatial lag

	Aggregate dependent variable		
Performance metric	Lasso	Random forests	Neural networks
AUC	0.59	0.53	0.56
	(0.03)	(0.02)	(0.03)
Brier score	0.15	0.15	0.20
	(0.002)	(0.002)	(0.011)
True positive rate (sensitivity)	77%	55%	64%
	(0.05)	(0.05)	(0.06)
True negative rate (specificity)	38%	49%	48%
	(0.02)	(0.03)	(0.03)
Accuracy	45%	50%	50%
	(0.02)	(0.02)	(0.03)
Ratio of false + to true +	3.70	4.27	3.81
	(0.24)	(0.44)	(0.41)
Ratio of false - to true +	0.31	0.84	0.59
	(0.08)	(0.16)	(0.16)

Table A.20: Forecasts of 2012 violence using 2010 risk factors plus spatial lag

	Aggregate dependent variable		
Performance metric	Lasso	Random forests	Neural networks
AUC	0.66	0.66	0.63
	(0.05)	(0.05)	(0.05)
Brier score	0.13	0.13	0.17
True positive rate (sensitivity)	87%	82%	61%
True negative rate (specificity)	20%	36%	56%
Accuracy	31%	43%	56%
Ratio of false + to true +	4.73	4.06	3.78
Ratio of false - to true +	0.15	0.23	0.65

Table A.21: Rankings of risk factors by model plus spatial lag

	Lasso		Random forests	
Risk factor	Rank	Coeff.	Rank	Importance
Minority tribe in town leadership	1	0.321	8	0.00031
Town population	2	0.134	1	0.00244
Nearby violence rate	3	0.080	42	-0.00015
% believing other tribes are violent	4	0.066	9	0.00027
% in dominant group	5	-0.063	4	0.00049
% who contribute to public facilities	6	0.030	48	-0.00019
# of tribes	7	0.000	14	0.00019
# of households			2	0.00096
% believing other tribes are dirty			3	0.00050
Wealth index			5	0.00045
% reporting loss of land during war			6	0.00042
Mean educational attainment			7	0.00032
% reporting neighbors are helpful			10	0.00025

Table A.22: Alternate ensemble methods

(a) Cross-validated forecast (2010)

	Logistic stack	Model average	Majority vote
Performance metric	(1)	(2)	(3)
AUC	0.50	0.55	
Brier score	0.29	0.16	
True positives (sensitivity)	44%	59%	72%
True negatives (specificity)	54%	49%	38%
Overall accuracy	52%	50%	44%
Ratio of false + to true +	5.08	4.16	4.17
Ratio of false - to true +	1.39	0.70	0.40

	Logistic stack	Model average	Majority vote
Performance metric	(1)	(2)	(3)
AUC	0.54	0.63	
Brier score	0.27	0.14	
True positives (sensitivity)	65%	78%	90%
True negatives (specificity)	41%	40%	20%
Overall accuracy	45%	46%	32%
Ratio of false + to true +	4.58	3.90	4.47
Ratio of false - to true +	0.54	0.29	0.11

(b) 2012 Forecasts

Visualization of trade-off between sensitivity and specificity at our preferred threshold

We provide another way to visualize the trade-off between true and false positives for the lasso model in the bar chart in Figure **??**. Each bar represents the predicted probability of violence in one town in 2012. The discrimination threshold is the probability above which we predict violence will occur—the optimal threshold identified through our cross-validated forecasts above.

Two features of the bar chart are noteworthy. First, while the number of false positives is relatively high, the number of false negatives is very low. This is by design: in training the lasso model to maximize true positives, we also train it to minimize false negatives, subject only to the constraint that overall accuracy remain at or above 50%. Second, many of these false positives have relatively high predicted probabilities of violence; indeed, of the 242 towns in our sample, the two with the highest predicted probabilities are in fact false positives. This pattern does not necessarily imply that the model is inherently flawed, or that the risk of violence in these towns is in fact lower than it appears. Estimates of risk are by nature probabilistic: todayâĂŹs false positive may prove to be tomorrowâĂŹs true positive, though without a longer panel we cannot test this proposition directly.

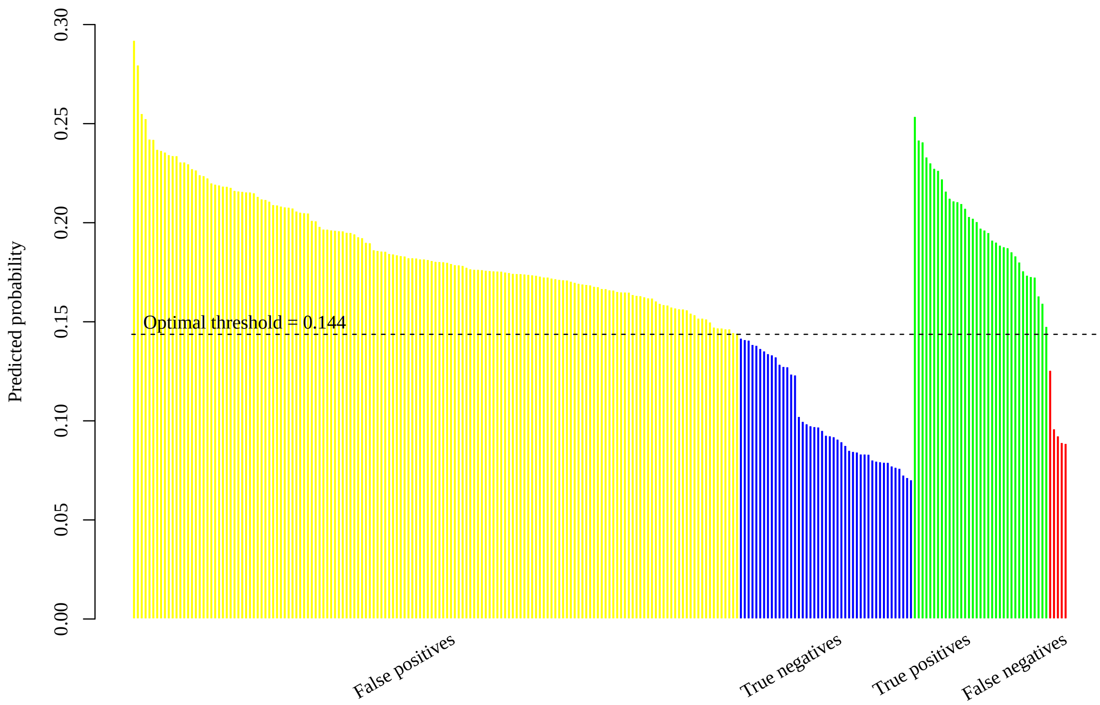
Visualization of distinction between lasso and logit

Figure **??** offers a visualization of the relationship between lasso and logit. λ is the penalty placed on the sum of the magnitudes of coefficients for included variables. As we move along the x-axis and λ decreases, the model becomes more flexible, and the number of predictors included in the model (fitted with non-zero coefficients) increases. When $\lambda = 0$, our lasso objective function is just the logit objective function. The figure below shows that our optimally-chosen λ is relatively restrictive. Cross-validation reveals that most of our available predictors add more noise than signal, and, consequently, we heavily penalize model complexity.

Risk factor rankings

Table **??** reports full risk factor rankings for the lasso, random forests and logit models described in the main paper. Tables **??**, **??** and **??** report the same rankings sorted by the magnitude of the lasso coefficients (in absolute value), random forests importance scores and logit coefficients (in absolute value), respectively.

Figure A.5: Predicted probabilities of 2012 violence, lasso model, by prediction accuracy



We apply the parameters from the lasso model estimated in Table 3 to 2010 data calculate the predicted probabilities of violence in 2012. The dotted line is the optimal threshold above which we predict violence, estimated in the same exercise (via 200 5-fold cross-validation trials).

Figure A.6: Lasso coefficients as penalty for additional variables is relaxed

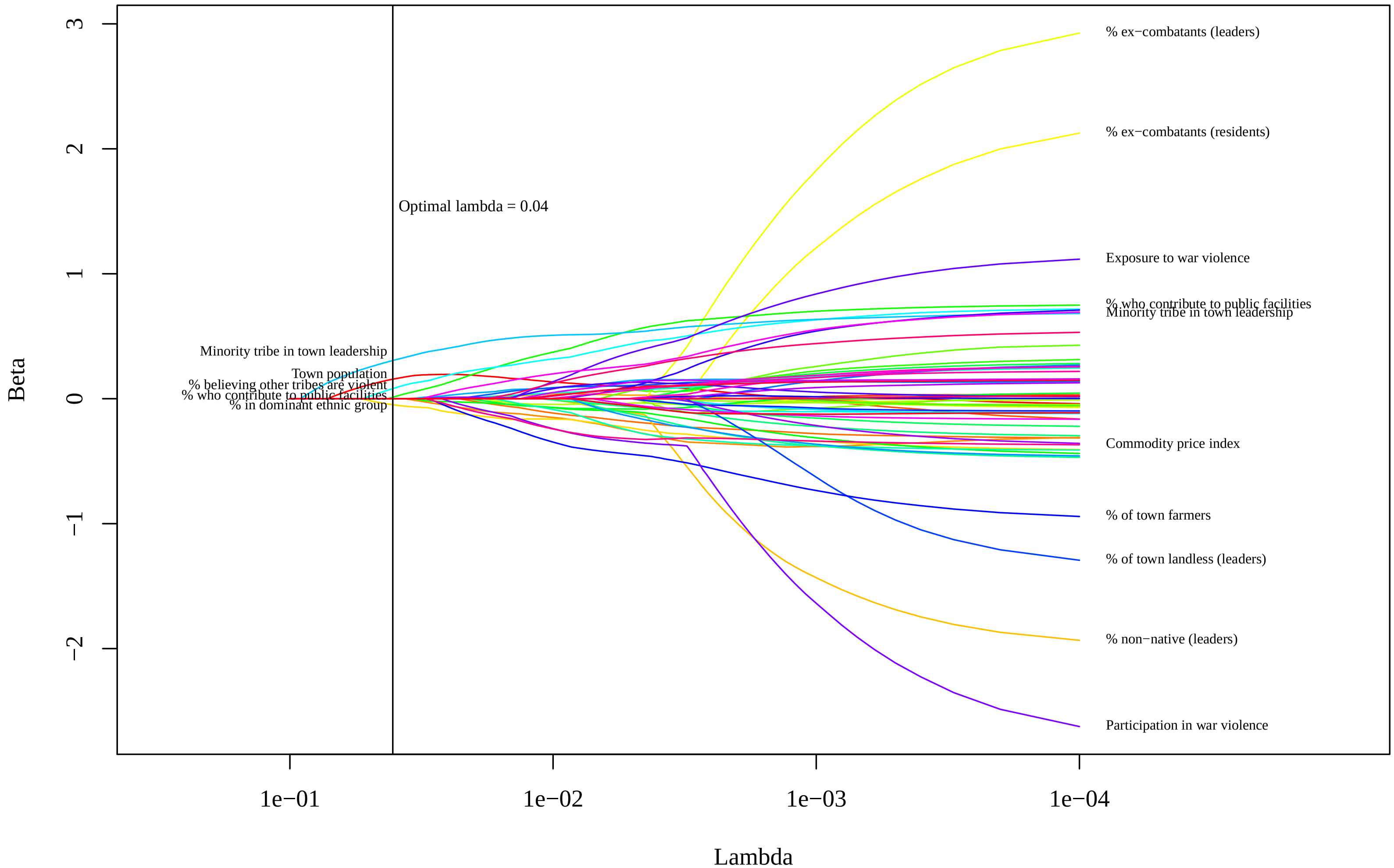


Table A.23: Full risk factor rankings

Risk factor	Lasso		Random forests		Logit		
	Rank	Coeff.	Rank	Importance	Rank	Coeff.	s.e.
Town population	2	0.156	1	0.002	47	-0.058	(0.416)
# of households			5	0.001	50	-0.046	(0.352)
# of tribes			4	0.001	52	0.029	(0.461)
% Muslims (leaders)			52	0.000	33	-0.198	(1.034)
Has mosque			30	0.000	23	-0.319	(0.294)
% Muslims (residents)			45	0.000	27	-0.289	(0.983)
% non-native (residents)			10	0.000	45	-0.074	(0.283)
% non-native (leaders)			53	0.000	4	-2.011	(1.209)
% in dominant group	4	-0.048	2	0.001	18	-0.419	(0.555)
% ex-combatants (residents)			39	0.000	3	2.292	(1.253)
% ex-combatants (leaders)			51	0.000	1	3.088	(1.344)
% returned from internal displacement			54	0.000	44	-0.078	(0.319)
% under 30			50	0.000	55	-0.006	(0.312)
% male			48	0.000	49	-0.048	(0.343)
Mean educational attainment			3	0.001	51	0.036	(1.010)
% with no education			14	0.000	15	0.475	(0.950)
% receiving any peace education			55	0.000	46	-0.060	(0.286)
Group participation (0-9)			12	0.000	22	0.329	(0.485)
% who contribute to public facilities	5	0.012	47	0.000	8	0.756	(0.455)
% saying town is safe at night			25	0.000	17	-0.459	(0.318)
% saying neighbors are helpful			24	0.000	48	0.057	(0.314)
Collective public goods			27	0.000	25	0.292	(0.353)
% who rely on NGOs			46	0.000	31	-0.230	(0.317)
% who rely on government			33	0.000	24	-0.308	(0.299)
Perceived equity in institutions			49	0.000	14	-0.485	(0.393)
% describing police/courts as corrupt			22	0.000	19	-0.418	(0.395)
% accepting inter-racial marriage			37	0.000	30	0.258	(0.338)
% who say Muslims shouldn't be leaders			42	0.000	41	-0.112	(0.362)
% believing other tribes are violent	3	0.081	23	0.000	10	0.730	(0.395)
% believing other tribes are dirty			28	0.000	40	-0.121	(0.382)
Minority tribe in town leadership	1	0.306	7	0.000	12	0.691	(0.319)
% reporting burglary or robbery			9	0.000	39	0.129	(0.316)
% reporting assault			16	0.000	16	-0.470	(0.392)
% reporting any land conflict			11	0.000	26	0.291	(0.483)
Any violent conflict			19	0.000	37	0.143	(0.287)
% of town landless (leaders)			35	0.000	5	-1.384	(1.476)
% of town landless (residents)			20	0.000	43	-0.100	(0.330)
% of town farmers			15	0.000	7	-0.979	(0.443)
Unemployment rate			41	0.000	56	0.000	(0.331)
Wealth index			13	0.000	9	0.733	(0.566)
S.D. of wealth index in town			6	0.000	54	0.009	(0.428)
Exposure to war violence			56	-0.001	6	1.162	(0.437)
Participation in war violence			36	0.000	2	-2.802	(1.315)
% reporting loss of land during war			8	0.000	20	-0.380	(0.328)
% displaced during war			26	0.000	38	0.133	(0.310)
Social services in town			31	0.000	35	0.161	(0.366)
Police or magistrate in town			17	0.000	34	-0.167	(0.265)
Frequency of police/NGO visits			29	0.000	11	0.717	(0.345)
Town $>$1 hour from road			40	0.000	28	0.274	(0.323)
Mobile phone coverage			32	0.000	29	0.265	(0.269)
Less than 2 radio stations			34	0.000	32	0.225	(0.294)
Natural resources in 2 hours			38	0.000	21	-0.373	(0.295)
Commodity price index			21	0.000	13	0.547	(0.298)
% affected by human disease			44	0.000	53	0.022	(0.285)
% affected by livestock disease			18	0.000	36	0.150	(0.311)
% affected by crop failure			43	0.000	42	-0.107	(0.324)

Table A.24: Full risk factor rankings - sorted on lasso

	Lasso		Random forests		Logit		
Risk factor	Rank	Coeff.	Rank	Importance	Rank	Coeff.	s.e.
Minority tribe in town leadership	1	0.306	7	0.0004	12	0.691	(0.319)
Town population	2	0.156	1	0.0024	47	-0.058	(0.416)
% believing other tribes are violent	3	0.081	23	0.0000	10	0.730	(0.395)
% in dominant group	4	-0.048	2	0.0008	18	-0.419	(0.555)
% who contribute to public facilities	5	0.012	47	-0.0001	8	0.756	(0.455)
Mean educational attainment			3	0.0007	51	0.036	(1.010)
# of tribes			4	0.0006	52	0.029	(0.461)
# of households			5	0.0005	50	-0.046	(0.352)
S.D. of wealth index in town			6	0.0004	54	0.009	(0.428)
% reporting loss of land during war			8	0.0004	20	-0.380	(0.328)
% reporting burglary or robbery			9	0.0003	39	0.129	(0.316)
% non-native (residents)			10	0.0003	45	-0.074	(0.283)
% reporting any land conflict			11	0.0002	26	0.291	(0.483)
Group participation (0-9)			12	0.0002	22	0.329	(0.485)
Wealth index			13	0.0002	9	0.733	(0.566)
% with no education			14	0.0002	15	0.475	(0.950)
% of town farmers			15	0.0001	7	-0.979	(0.443)
% reporting assault			16	0.0001	16	-0.470	(0.392)
Police or magistrate in town			17	0.0001	34	-0.167	(0.265)
% affected by livestock disease			18	0.0001	36	0.150	(0.311)
Any violent event (lagged DV)			19	0.0001	37	0.143	(0.287)
% of town landless (residents)			20	0.0001	43	-0.100	(0.330)
Commodity price index			21	0.0001	13	0.547	(0.298)
% describing police/courts as corrupt			22	0.0001	19	-0.418	(0.395)
% saying neighbors are helpful			24	0.0000	48	0.057	(0.314)
% saying town is safe at night			25	0.0000	17	-0.459	(0.318)
% displaced during war			26	0.0000	38	0.133	(0.310)
Collective public goods			27	0.0000	25	0.292	(0.353)
% believing other tribes are dirty			28	0.0000	40	-0.121	(0.382)
Frequency of police/NGO visits			29	0.0000	11	0.717	(0.345)
Has mosque			30	0.0000	23	-0.319	(0.294)
Social services in town			31	0.0000	35	0.161	(0.366)
Mobile phone coverage			32	0.0000	29	0.265	(0.269)
% who rely on government			33	0.0000	24	-0.308	(0.299)
Less than 2 radio stations			34	-0.0001	32	0.225	(0.294)
% of town landless (leaders)			35	-0.0001	5	-1.384	(1.476)
Participation in war violence			36	-0.0001	2	-2.802	(1.315)
% accepting inter-racial marriage			37	-0.0001	30	0.258	(0.338)
Natural resources in 2 hours			38	-0.0001	21	-0.373	(0.295)
% ex-combatants (residents)			39	-0.0001	3	2.292	(1.253)
Town >1 hour from road			40	-0.0001	28	0.274	(0.323)
Unemployment rate			41	-0.0001	56	0.000	(0.331)
% who say Muslims shouldn't be leaders			42	-0.0001	41	-0.112	(0.362)
% affected by crop failure			43	-0.0001	42	-0.107	(0.324)
% affected by human disease			44	-0.0001	53	0.022	(0.285)
% Muslims (residents)			45	-0.0001	27	-0.289	(0.983)
% who rely on NGOs			46	-0.0001	31	-0.230	(0.317)
% male			48	-0.0002	49	-0.048	(0.343)
Perceived equity in institutions			49	-0.0002	14	-0.485	(0.393)
% under 30			50	-0.0002	55	-0.006	(0.312)
% ex-combatants (leaders)			51	-0.0002	1	3.088	(1.344)
% Muslims (leaders)			52	-0.0003	33	-0.198	(1.034)
% non-native (leaders)			53	-0.0003	4	-2.011	(1.209)
% returned from internal displacement			54	-0.0003	44	-0.078	(0.319)
% receiving any peace education			55	-0.0003	46	-0.060	(0.286)
Exposure to war violence			56	-0.0007	6	1.162	(0.437)

Table A.25: Full risk factor rankings - sorted on random forests

	Lasso		Random forests		Logit		
Risk factor	Rank	Coeff.	Rank	Importance	Rank	Coeff.	s.e.
Town population	2	0.156	1	0.0024	47	-0.058	(0.416)
% in dominant group	4	-0.048	2	0.0008	18	-0.419	(0.555)
Mean educational attainment			3	0.0007	51	0.036	(1.010)
# of tribes			4	0.0006	52	0.029	(0.461)
# of households			5	0.0005	50	-0.046	(0.352)
S.D. of wealth index in town			6	0.0004	54	0.009	(0.428)
Minority tribe in town leadership	1	0.306	7	0.0004	12	0.691	(0.319)
% reporting loss of land during war			8	0.0004	20	-0.380	(0.328)
% reporting burglary or robbery			9	0.0003	39	0.129	(0.316)
% non-native (residents)			10	0.0003	45	-0.074	(0.283)
% reporting any land conflict			11	0.0002	26	0.291	(0.483)
Group participation (0-9)			12	0.0002	22	0.329	(0.485)
Wealth index			13	0.0002	9	0.733	(0.566)
% with no education			14	0.0002	15	0.475	(0.950)
% of town farmers			15	0.0001	7	-0.979	(0.443)
% reporting assault			16	0.0001	16	-0.470	(0.392)
Police or magistrate in town			17	0.0001	34	-0.167	(0.265)
% affected by livestock disease			18	0.0001	36	0.150	(0.311)
Any violent event (lagged DV)			19	0.0001	37	0.143	(0.287)
% of town landless (residents)			20	0.0001	43	-0.100	(0.330)
Commodity price index			21	0.0001	13	0.547	(0.298)
% describing police/courts as corrupt			22	0.0001	19	-0.418	(0.395)
% believing other tribes are violent	3	0.081	23	0.0000	10	0.730	(0.395)
% saying neighbors are helpful			24	0.0000	48	0.057	(0.314)
% saying town is safe at night			25	0.0000	17	-0.459	(0.318)
% displaced during war			26	0.0000	38	0.133	(0.310)
Collective public goods			27	0.0000	25	0.292	(0.353)
% believing other tribes are dirty			28	0.0000	40	-0.121	(0.382)
Frequency of police/NGO visits			29	0.0000	11	0.717	(0.345)
Has mosque			30	0.0000	23	-0.319	(0.294)
Social services in town			31	0.0000	35	0.161	(0.366)
Mobile phone coverage			32	0.0000	29	0.265	(0.269)
% who rely on government			33	0.0000	24	-0.308	(0.299)
Less than 2 radio stations			34	-0.0001	32	0.225	(0.294)
% of town landless (leaders)			35	-0.0001	5	-1.384	(1.476)
Participation in war violence			36	-0.0001	2	-2.802	(1.315)
% accepting inter-racial marriage			37	-0.0001	30	0.258	(0.338)
Natural resources in 2 hours			38	-0.0001	21	-0.373	(0.295)
% ex-combatants (residents)			39	-0.0001	3	2.292	(1.253)
Town >1 hour from road			40	-0.0001	28	0.274	(0.323)
Unemployment rate			41	-0.0001	56	0.000	(0.331)
% who say Muslims shouldn't be leaders			42	-0.0001	41	-0.112	(0.362)
% affected by crop failure			43	-0.0001	42	-0.107	(0.324)
% affected by human disease			44	-0.0001	53	0.022	(0.285)
% Muslims (residents)			45	-0.0001	27	-0.289	(0.983)
% who rely on NGOs			46	-0.0001	31	-0.230	(0.317)
% who contribute to public facilities	5	0.012	47	-0.0001	8	0.756	(0.455)
% male			48	-0.0002	49	-0.048	(0.343)
Perceived equity in institutions			49	-0.0002	14	-0.485	(0.393)
% under 30			50	-0.0002	55	-0.006	(0.312)
% ex-combatants (leaders)			51	-0.0002	1	3.088	(1.344)
% Muslims (leaders)			52	-0.0003	33	-0.198	(1.034)
% non-native (leaders)			53	-0.0003	4	-2.011	(1.209)
% returned from internal displacement			54	-0.0003	44	-0.078	(0.319)
% receiving any peace education			55	-0.0003	46	-0.060	(0.286)
Exposure to war violence			56	-0.0007	6	1.162	(0.437)

Table A.26: Full risk factor rankings - sorted on logit

Risk factor	Lasso		Random forests		Logit		
	Rank	Coeff.	Rank	Importance	Rank	Coeff.	s.e.
% ex-combatants (leaders)			51	-0.0002	1	3.088	(1.344)
Participation in war violence			36	-0.0001	2	-2.802	(1.315)
% ex-combatants (residents)			39	-0.0001	3	2.292	(1.253)
% non-native (leaders)			53	-0.0003	4	-2.011	(1.209)
% of town landless (leaders)			35	-0.0001	5	-1.384	(1.476)
Exposure to war violence			56	-0.0007	6	1.162	(0.437)
% of town farmers			15	0.0001	7	-0.979	(0.443)
% who contribute to public facilities	5	0.012	47	-0.0001	8	0.756	(0.455)
Wealth index			13	0.0002	9	0.733	(0.566)
% believing other tribes are violent	3	0.081	23	0.0000	10	0.730	(0.395)
Frequency of police/NGO visits			29	0.0000	11	0.717	(0.345)
Minority tribe in town leadership	1	0.306	7	0.0004	12	0.691	(0.319)
Commodity price index			21	0.0001	13	0.547	(0.298)
Perceived equity in institutions			49	-0.0002	14	-0.485	(0.393)
% with no education			14	0.0002	15	0.475	(0.950)
% reporting assault			16	0.0001	16	-0.470	(0.392)
% saying town is safe at night			25	0.0000	17	-0.459	(0.318)
% in dominant group	4	-0.048	2	0.0008	18	-0.419	(0.555)
% describing police/courts as corrupt			22	0.0001	19	-0.418	(0.395)
% reporting loss of land during war			8	0.0004	20	-0.380	(0.328)
Natural resources in 2 hours			38	-0.0001	21	-0.373	(0.295)
Group participation (0-9)			12	0.0002	22	0.329	(0.485)
Has mosque			30	0.0000	23	-0.319	(0.294)
% who rely on government			33	0.0000	24	-0.308	(0.299)
Collective public goods			27	0.0000	25	0.292	(0.353)
% reporting any land conflict			11	0.0002	26	0.291	(0.483)
% Muslims (residents)			45	-0.0001	27	-0.289	(0.983)
Town >1 hour from road			40	-0.0001	28	0.274	(0.323)
Mobile phone coverage			32	0.0000	29	0.265	(0.269)
% accepting inter-racial marriage			37	-0.0001	30	0.258	(0.338)
% who rely on NGOs			46	-0.0001	31	-0.230	(0.317)
Less than 2 radio stations			34	-0.0001	32	0.225	(0.294)
% Muslims (leaders)			52	-0.0003	33	-0.198	(1.034)
Police or magistrate in town			17	0.0001	34	-0.167	(0.265)
Social services in town			31	0.0000	35	0.161	(0.366)
% affected by livestock disease			18	0.0001	36	0.150	(0.311)
Any violent event (lagged DV)			19	0.0001	37	0.143	(0.287)
% displaced during war			26	0.0000	38	0.133	(0.310)
% reporting burglary or robbery			9	0.0003	39	0.129	(0.316)
% believing other tribes are dirty			28	0.0000	40	-0.121	(0.382)
% who say Muslims shouldn't be leaders			42	-0.0001	41	-0.112	(0.362)
% affected by crop failure			43	-0.0001	42	-0.107	(0.324)
% of town landless (residents)			20	0.0001	43	-0.100	(0.330)
% returned from internal displacement			54	-0.0003	44	-0.078	(0.319)
% non-native (residents)			10	0.0003	45	-0.074	(0.283)
% receiving any peace education			55	-0.0003	46	-0.060	(0.286)
Town population	2	0.156	1	0.0024	47	-0.058	(0.416)
% saying neighbors are helpful			24	0.0000	48	0.057	(0.314)
% male			48	-0.0002	49	-0.048	(0.343)
# of households			5	0.0005	50	-0.046	(0.352)
Mean educational attainment			3	0.0007	51	0.036	(1.010)
# of tribes			4	0.0006	52	0.029	(0.461)
% affected by human disease			44	-0.0001	53	0.022	(0.285)
S.D. of wealth index in town			6	0.0004	54	0.009	(0.428)
% under 30			50	-0.0002	55	-0.006	(0.312)
Unemployment rate			41	-0.0001	56	0.000	(0.331)

Risk factor rankings after disaggregating incidents by category

Tables **??**, **??** and **??** report risk factor rankings for models trained and tested on each of our three categories of violence–collective, interpersonal and extrajudicial, respectively. In general, different categories of violence appear to have different predictors. For example, proxies for ethnic heterogeneity, polarization and fractionalization tend to predict interpersonal violence more accurately than they predict collective or extrajudicial violence, while estimates for the proportion of residents that fought in the civil war—and, relatedly, the proportion of residents that self-identify as ex-combatants—predict extrajudicial violence more accurately than they predict collective or interpersonal violence. Our indicator for power-sharing is the highest ranked predictor of interpersonal violence in the lasso model, but does not appear among the top 10 predictors of either collective or extrajudicial violence. Exposure to disease—an adverse economic shock—is highly ranked in the lasso model for interpersonal violence as well, but not for collective or extrajudicial violence. Interestingly, the coefficient on exposure to disease is negative, suggesting that, if anything, shocks of this sort are associated with less rather than more interpersonal violence.

Some risk factors do, however, recur across models. For example, more agricultural towns tend to be less susceptible to both collective and extrajudicial violence. Curiously, towns whose residents perceive the police, courts and local leaders as just are *more* rather than less prone to collective violence, and towns whose residents perceive the police and courts as corrupt tend to be less rather than more prone to extrajudicial violence. This may be because perceptions of malfeasance are most severe in places where the police and courts are most active (assuming, of course, that police presence mitigates the risk of violence, which it may not).

Table A.27: Full risk factor rankings for predicting collective violence

	Lasso		Random forests	
Risk factor	Rank	Coeff.	Rank	Importance
Town population	15	0.227	3	0.00030

Risk factor	Lasso Rank	Lasso Coeff.	Random forests Rank	Random forests Importance
# of households			43	0.00000
# of tribes			44	0.00000
% Muslims (leaders)	22	-0.166	10	0.00010
Has mosque	13	-0.299	30	0.00000
% Muslims (residents)			13	0.00010
% non-native (residents)	9	0.434	4	0.00010
% non-native (leaders)			39	0.00000
% in dominant group			49	0.00000
% ex-combatants (residents)			31	0.00000
% ex-combatants (leaders)			53	-0.00010
% returned from internal displacement			55	-0.00010
% under 30	24	-0.126	38	0.00000
% male			42	0.00000
Mean educational attainment			21	0.00000
% with no education			29	0.00000
% receiving any peace education			56	-0.00010
Group participation (0-9)	11	0.349	26	0.00000
% who contribute to public facilities			27	0.00000
% saying town is safe at night	17	-0.181	1	0.00050
% saying neighbors are helpful			51	-0.00010
Collective public goods	29	-0.045	19	0.00000
% who rely on NGOs	6	0.472	6	0.00010
% who rely on government	19	0.173	35	0.00000
Perceived equity in institutions	4	0.578	7	0.00010
% describing police/courts as corrupt			52	-0.00010
% accepting inter-racial marriage	8	0.443	25	0.00000
% who say Muslims shouldn't be leaders	20	-0.169	16	0.00000
% believing other tribes are violent	14	0.288	47	0.00000
% believing other tribes are dirty			45	0.00000
Minority tribe in town leadership	26	0.103	32	0.00000
% reporting burglary or robbery	2	0.632	5	0.00010
% reporting assault	3	-0.583	24	0.00000
% reporting any land conflict			11	0.00010
Any violent event (lagged DV)	18	-0.181	33	0.00000
% of town landless (leaders)			41	0.00000
% of town landless (residents)	23	0.154	17	0.00000
% of town farmers	1	-0.717	8	0.00010
Unemployment rate	16	0.221	9	0.00010
Wealth index	12	-0.323	46	0.00000
S.D. of wealth index in town	31	-0.03	14	0.00010
Exposure to war violence			36	0.00000
Participation in war violence	5	-0.55	28	0.00000
% reporting loss of land during war	28	0.085	22	0.00000
% displaced during war	10	-0.399	2	0.00050
Social services in town			12	0.00010
Police or magistrate in town	25	0.113	48	0.00000
Frequency of police/NGO visits	7	0.454	50	-0.00010
Town >1 hour from road	21	0.167	20	0.00000
Mobile phone coverage			15	0.00000
Less than 2 radio stations	30	-0.04	37	0.00000
Natural resources in 2 hours			54	-0.00010
Commodity price index			34	0.00000
% affected by human disease	27	0.09	18	0.00000
% affected by livestock disease			23	0.00000
% affected by crop failure			40	0.00000

Table A.28: Full risk factor rankings for predicting interpersonal violence

Risk factor	Lasso Rank	Lasso Coeff.	Random forests Rank	Random forests Importance
Town population	38	0.031	1	0.00250
# of households			3	0.00090
# of tribes	17	0.234	2	0.00090
% Muslims (leaders)			37	0.00000

Risk factor	Lasso Rank	Lasso Coeff.	Random forests Rank	Random forests Importance
Has mosque	7	-0.443	30	0.00000
% Muslims (residents)	31	-0.14	46	-0.00010
% non-native (residents)	39	0.017	10	0.00030
% non-native (leaders)			29	0.00000
% in dominant group			12	0.00020
% ex-combatants (residents)			23	0.00010
% ex-combatants (leaders)			40	0.00000
% returned from internal displacement	33	0.08	51	-0.00010
% under 30	19	-0.223	54	-0.00020
% male			38	0.00000
Mean educational attainment	22	0.194	5	0.00070
% with no education			8	0.00040
% receiving any peace education	21	0.195	36	0.00000
Group participation (0-9)			53	-0.00010
% who contribute to public facilities	10	0.381	35	0.00000
% saying town is safe at night	20	-0.2	16	0.00010
% saying neighbors are helpful	23	-0.187	9	0.00040
Collective public goods	30	0.14	32	0.00000
% who rely on NGOs	9	-0.404	14	0.00020
% who rely on government	12	-0.365	55	-0.00020
Perceived equity in institutions	34	-0.077	27	0.00000
% describing police/courts as corrupt	28	-0.146	48	-0.00010
% accepting inter-racial marriage			28	0.00000
% who say Muslims shouldn't be leaders	36	-0.034	31	0.00000
% believing other tribes are violent	27	0.146	15	0.00020
% believing other tribes are dirty			6	0.00060
Minority tribe in town leadership	1	0.682	22	0.00010
% reporting burglary or robbery	25	0.171	11	0.00030
% reporting assault			26	0.00000
% reporting any land conflict			18	0.00010
Any violent event (lagged DV)	6	0.469	21	0.00010
% of town landless (leaders)			52	-0.00010
% of town landless (residents)			44	0.00000
% of town farmers	13	-0.33	17	0.00010
Unemployment rate	29	0.14	43	0.00000
Wealth index	18	0.23	13	0.00020
S.D. of wealth index in town	24	0.174	24	0.00010
Exposure to war violence	4	0.507	56	-0.00040
Participation in war violence	14	-0.294	47	-0.00010
% reporting loss of land during war	40	0.002	4	0.00070
% displaced during war	11	0.378	25	0.00010
Social services in town	26	0.171	7	0.00050
Police or magistrate in town	37	-0.034	20	0.00010
Frequency of police/NGO visits	5	0.498	34	0.00000
Town >1 hour from road	16	-0.246	39	0.00000
Mobile phone coverage			33	0.00000
Less than 2 radio stations	15	0.278	41	0.00000
Natural resources in 2 hours	3	-0.509	45	-0.00010
Commodity price index	8	0.407	19	0.00010
% affected by human disease	2	-0.635	42	0.00000
% affected by livestock disease	32	0.113	50	-0.00010
% affected by crop failure	35	-0.076	49	-0.00010

Table A.29: Full risk factor rankings for predicting ritual violence

Risk factor	Lasso Rank	Lasso Coeff.	Random forests Rank	Random forests Importance
Town population			32	0.00000
# of households			53	-0.00010
# of tribes			14	0.00000
% Muslims (leaders)			21	0.00000
Has mosque			22	0.00000
% Muslims (residents)			23	0.00000
% non-native (residents)			7	0.00000

	Lasso		Random forests	
Risk factor	Rank	Coeff.	Rank	Importance
% non-native (leaders)			54	-0.00010
% in dominant group	9	-0.276	41	0.00000
% ex-combatants (residents)	14	0.136	2	0.00010
% ex-combatants (leaders)			29	0.00000
% returned from internal displacement	4	-0.607	46	0.00000
% under 30			3	0.00010
% male			37	0.00000
Mean educational attainment			9	0.00000
% with no education			10	0.00000
% receiving any peace education	6	-0.48	55	-0.00010
Group participation (0-9)			43	0.00000
% who contribute to public facilities	5	0.524	34	0.00000
% saying town is safe at night	20	-0.002	40	0.00000
% saying neighbors are helpful			15	0.00000
Collective public goods			4	0.00000
% who rely on NGOs			56	-0.00010
% who rely on government	12	0.179	42	0.00000
Perceived equity in institutions	17	0.084	6	0.00000
% describing police/courts as corrupt	3	-0.628	12	0.00000
% accepting inter-racial marriage	18	0.078	17	0.00000
% who say Muslims shouldn't be leaders	15	0.118	44	0.00000
% believing other tribes are violent			39	0.00000
% believing other tribes are dirty			28	0.00000
Minority tribe in town leadership			24	0.00000
% reporting burglary or robbery			36	0.00000
% reporting assault			45	0.00000
% reporting any land conflict			50	0.00000
Any violent event (lagged DV)			25	0.00000
% of town landless (leaders)			47	0.00000
% of town landless (residents)			30	0.00000
% of town farmers	2	-0.654	49	0.00000
Unemployment rate	1	-0.841	26	0.00000
Wealth index	16	0.116	35	0.00000
S.D. of wealth index in town	10	0.27	51	0.00000
Exposure to war violence			5	0.00000
Participation in war violence	13	0.163	1	0.00040
% reporting loss of land during war	19	-0.01	20	0.00000
% displaced during war			31	0.00000
Social services in town			48	0.00000
Police or magistrate in town			33	0.00000
Frequency of police/NGO visits			19	0.00000
Town >1 hour from road	7	0.464	13	0.00000
Mobile phone coverage			27	0.00000
Less than 2 radio stations			38	0.00000
Natural resources in 2 hours			11	0.00000
Commodity price index	8	0.443	16	0.00000
% affected by human disease	11	0.238	18	0.00000
% affected by livestock disease			8	0.00000
% affected by crop failure			52	-0.00010

Using predictors of one category of violence to predict the others

??If different categories of violence have similar predictors, then it should be possible to forecast one category reasonably accurately using the same risk factors and model parameters used to forecast the others. We consider two approaches to testing this proposition: first, using models trained on one category of violence to predict each of the others, and second, using models trained on two categories of violence to predict the third.

Tables **??**, **??**, and **??** report performance metrics for the first approach. Not surprisingly,

Table A.30: Performance metrics for models trained on collective violence and tested on other categories

(a) Cross-validated forecasts for models trained on collective violence

	Interpersonal			Extrajudicial		
Performance metric	Lasso	RF	NN	Lasso	RF	NN
AUC	0.55	0.53	0.53	0.56	0.54	0.55
Brier score	0.13	0.12	0.13	0.04	0.03	0.05
True positives (sensitivity)	57%	55%	54%	41%	46%	42%
True negatives (specificity)	51%	50%	51%	49%	49%	50%
Overall accuracy	51%	51%	51%	49%	49%	49%
Ratio of false + to true +	5.79	6.06	6.13	40.56	34.85	39.89
Ratio of false - to true +	0.79	0.84	0.89	1.74	1.34	1.71

(b) True forecasts for models trained on collective violence

	Interpersonal			Extrajudicial		
Performance metric	Lasso	RF	NN	Lasso	RF	NN
AUC	0.65	0.51	0.53	0.67	0.48	0.60
Brier score	0.08	0.08	0.09	0.06	0.06	0.07
True positives (sensitivity)	81%	81%	52%	75%	81%	56%
True negatives (specificity)	47%	24%	56%	46%	24%	56%
Overall accuracy	50%	29%	55%	48%	28%	56%
Ratio of false + to true +	6.88	9.82	8.91	10.17	13.15	11.11
Ratio of false - to true +	0.24	0.24	0.91	0.33	0.23	0.78

the models generally (though not universally) perform better when they are trained and tested on the same dependent variable than when they are trained on one and tested on the others. These results are unstable, however, and should be interpreted with caution.

Table **??** reports performance metrics for the second approach. Again, the models generally (though not universally) perform better when they are trained and tested on the same dependent variable than when they are trained on two and tested on the third. More surprisingly, however, the models are reasonably accurate—and almost uniformly better than chance—even when predicting across categories, and the differences between within- and cross-category performance are in many cases small. This is especially true when we train our models on two categories of violence and test them on the third. Outside the realm of forecasting, many studies implicitly assume that different categories of violence are distinct, and restrict their analyses to one at a time.

Table A.31: Performance metrics for models trained on one category of local violence and tested on each of the others

(a) Cross-validated forecasts for models trained on interpersonal violence

	Collective			Extrajudicial		
Performance metric	Lasso	RF	NN	Lasso	RF	NN
AUC	0.54	0.56	0.55	0.54	0.52	0.53
Brier score	0.08	0.05	0.10	0.09	0.05	0.11
True positives (sensitivity)	64%	61%	64%	59%	52%	52%
True negatives (specificity)	37%	47%	43%	37%	47%	42%
Overall accuracy	38%	48%	43%	37%	47%	42%
Ratio of false + to true +	25.53	22.73	23.40	32.47	31.92	33.76
Ratio of false - to true +	0.56	0.67	0.58	0.76	1.06	1.00

(b) True forecasts for models trained on interpersonal violence

	Collective			Extrajudicial		
Performance metric	Lasso	RF	NN	Lasso	RF	NN
AUC	0.46	0.71	0.63	0.39	0.54	0.59
Brier score	0.08	0.05	0.10	0.11	0.07	0.13
True positives (sensitivity)	71%	86%	86%	44%	44%	38%
True negatives (specificity)	34%	34%	46%	32%	32%	43%
Overall accuracy	35%	36%	47%	33%	33%	43%
Ratio of false + to true +	31.20	25.67	21.33	22.00	21.86	21.33
Ratio of false - to true +	0.40	0.17	0.17	1.29	1.29	1.67

Table A.32: Performance metrics for models trained on extrajudicial violence and tested on other categories

(a) Cross-validated forecasts for models trained on extrajudicial violence

	Collective			Interpersonal		
Performance metric	Lasso	RF	NN	Lasso	RF	NN
AUC	0.56	0.54	0.53	0.52	0.51	0.52
Brier score	0.04	0.04	0.05	0.13	0.12	0.13
True positives (sensitivity)	47%	53%	46%	55%	50%	48%
True negatives (specificity)	46%	52%	52%	46%	52%	52%
Overall accuracy	46%	52%	52%	47%	51%	52%
Ratio of false + to true +	31.91	25.40	29.17	6.53	6.46	6.60
Ratio of false - to true +	1.28	1.03	1.34	0.85	1.04	1.10

(b) True forecasts for models trained on extrajudicial violence

	Collective			Interpersonal		
Performance metric	Lasso	RF	NN	Lasso	RF	NN
AUC	0.54	0.77	0.59	0.53	0.59	0.54
Brier score	0.03	0.03	0.04	0.09	0.08	0.09
True positives (sensitivity)	71%	57%	29%	71%	67%	38%
True negatives (specificity)	43%	22%	54%	44%	21%	54%
Overall accuracy	44%	23%	53%	46%	25%	52%
Ratio of false + to true +	26.80	46.00	54.00	8.27	12.43	12.75
Ratio of false - to true +	0.40	0.75	2.50	0.40	0.50	1.63

Our results, while ambiguous, suggest at the very least that these distinctions should be treated not as an assumption, but rather as a hypothesis to be tested.

Table A.33: Performance metrics for models trained on two categories of local violence and tested on the third

(a) Cross-validated forecasts for models trained and tested on two categories of local violence

Trained and tested on:	Interpersonal + extrajudicial			Interpersonal + collective			Collective + extrajudicial		
Performance metric	Lasso	RF	NN	Lasso	RF	NN	Lasso	RF	NN
AUC	0.57	0.54	0.58	0.57	0.53	0.58	0.53	0.53	0.55
Brier score	0.13	0.13	0.18	0.14	0.13	0.18	0.10	0.06	0.09
True positives (sensitivity)	0.79	58%	50%	78%	53%	51%	53%	49%	51%
True negatives (specificity)	0.33	49%	65%	32%	50%	66%	49%	51%	43%
Overall accuracy	0.41	50%	48%	39%	50%	48%	49%	51%	51%
Ratio of false + to true +	4.52	4.78	4.37	4.81	5.23	4.41	14.26	14.67	17.01
Ratio of false - to true +	0.26	0.75	0.56	0.28	0.89	0.53	0.98	1.11	1.47

(b) Cross-validated forecasts for models trained on two categories of local violence and tested on the third

Trained on:	Interpersonal + extrajudicial			Interpersonal + collective			Collective + extrajudicial		
Tested on:	Collective			Extrajudicial			Interpersonal		
Performance metric	Lasso	RF	NN	Lasso	RF	NN	Lasso	RF	NN
AUC	0.60	0.54	0.53	0.53	0.53	0.54	0.52	0.54	0.52
Brier score	0.05	0.06	0.12	0.06	0.06	0.12	0.16	0.12	0.14
True positives (sensitivity)	78%	58%	46%	57%	44%	46%	51%	56%	51%
True negatives (specificity)	32%	48%	59%	30%	49%	51%	49%	52%	49%
Overall accuracy	33%	49%	46%	31%	49%	46%	49%	52%	52%
Ratio of false + to true +	22.90	23.57	24.55	36.19	36.99	33.08	6.61	5.70	6.53
Ratio of false - to true +	0.29	0.76	0.76	0.77	1.49	1.08	0.98	0.80	1.06

(c) True forecasts for models trained on two categories of local violence and tested on the third

Trained on:	Interpersonal + extrajudicial			Interpersonal + collective			Collective + extrajudicial		
Tested on:	Collective			Extrajudicial			Interpersonal		
Performance metric	Lasso	RF	NN	Lasso	RF	NN	Lasso	RF	NN
AUC	0.70	0.66	0.58	0.58	0.52	0.52	0.58	0.49	0.54
Brier score	0.05	0.06	0.12	0.07	0.08	0.14	0.11	0.08	0.10
True positives (sensitivity)	86%	86%	71%	88%	63%	56%	67%	81%	38%
True negatives (specificity)	21%	37%	52%	22%	38%	46%	39%	17%	53%
Overall accuracy	23%	39%	53%	26%	39%	46%	42%	23%	52%
Ratio of false + to true +	31.00	24.50	22.40	12.64	14.10	13.67	9.57	10.76	13.00
Ratio of false - to true +	0.17	0.17	0.40	0.14	0.60	0.78	0.50	0.24	1.63