Journal of
Peace Research

Journal of Peace Research
2017, Vol. 54(2) 298–312

DOI: 10.1177/0022343316684009
journals.sagepub.com/home/jpr
SAGE

# Predicting local violence: Evidence from a panel survey in Liberia

**Robert A Blair**

*Department of Political Science & Watson Institute for International and Public Affairs, Brown University*

**Christopher Blattman**

*Harris School of Public Policy, University of Chicago*

**Alexandra Hartman**

*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 counter-intuitive 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.

### Keywords

Africa, forecasting, machine learning, surveys, violence

## 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 (Fearon & Hoeffler, 2014). Local violence can also shape, and be shaped by, national conflict dynamics (Autesserre, 2010). 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 – for example, civil war, political instability (Goldstone et al., 2010) or 'irregular regime change'

**Corresponding author:**
robert_blair@brown.edu

(Beger, Dorff & Ward, 2016).[1] Recently, however, the most active frontier of conflict research has focused on local-level incidents, including murders (Blakeslee & Fishman, 2015), riots (Wischnath & Buhaug, 2014), domestic violence (Sekhri & Storeygard, 2013), 'low-intensity' sectarian clashes (Balcells, Daniels & Escribà-Folch, 2016), conflicts between states and underrepresented groups (Cederman, Wimmer & Min, 2010), and killings of suspected witches (Miguel, 2005). Like its cross-national counterpart, most subnational 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 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 five 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. Goldstone et al., 2010.)

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 (Blakeslee & Fishman, 2015), witch killings (Miguel, 2005), and domestic violence (Sekhri & Storeygard, 2013), among other outcomes (see Hsiang, Burke & Miguel, 2013 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 (Ward, Greenhill & Bakke, 2010), 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 (Blattman, Hartman & Blair, 2014). 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. Counter-intuitively, 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. Rothchild & Roeder, 2005). We discuss possible interpretations and implications of this finding for future research, and include additional results and robustness checks in the Online appendix.

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 Hirose, Imai & Lyall, 2017). Most researchers use either structural or event data instead (or, increasingly, some combination of the two, e.g. Beger, Dorff & Ward,

[1] For a literature review see Ward, Greenhill & Bakke (2010).

2016; Blair & Sambanis, 2016; Chiba & Gleditsch, 2017). 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 than 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 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 15 indigenous ethnic groups (or 'tribes') 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 June of 2016, 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 violence.

## Data and measurement

We collected survey data from residents and leaders in 242 communities across three of Liberia's 15 counties: Lofa, Nimba, and Grand Gedeh. Our unit of analysis is the smallest administrative unit in the country, ranging 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 a randomized controlled trial evaluating a UNHCR-sponsored alternative dispute resolution program (Blattman, Hartman & Blair, 2014), 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 (Vinck, Pham & Kreutzer, 2011). 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

[2] Though innovation in passive measurement should make data collection at this level increasingly reliable, even cross-nationally; see e.g. Weidmann & Schutte (2017) in this issue.

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 (Isser, Lubkemann & N'Tow, 2009), 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 alleged 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 violence 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 the 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 in the Online appendix, which also 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 allow us to capture a wide variety of theoretically and contextually relevant 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

[3] O'Brien (2013) and Schrodt & Van Brackle (2013) 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 the advantages and disadvantages of our dataset in detail in the Online appendix.

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 – for example, sub-Saharan Africa – or that drop or separately model countries believed to be 'immune' from conflict (Beger, Dorff & Ward, 2016). Furthermore, comparison to a nationally representative survey suggests that our towns were not much more conflicted than the average Liberian community (Vinck, Pham & Kreutzer, 2011). 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 sixth largest city in Liberia) and Zwedru (the eighth 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. Brass, 1997), 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 does not 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 nonviolent; 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 (Spear, 2006). Conflict between 'locals' and perceived 'foreigners' (including returned IDPs and refugees) is common across sub-Saharan Africa (Onoma, 2013), 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 (Putnam, Leonardi & Nanetti, 1994) and collective (especially ethnic) violence (Varshney, 2002).

**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 Wilkinson, 2009 for a review); extrajudicial violence may also be more common where citizens distrust the police and courts (Isser, Lubkemann & N'Tow, 2009).

**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 (Ross, 2004), though it is unclear whether the mechanisms underlying this relationship manifest at the subnational level as well. Even if they do not, proximity to natural resources may attract transient and potentially disruptive populations – ex-combatants, for example (Blattman & Annan, 2016) – 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. Horowitz, 2001) and national violence (e.g. Posen, 1993), especially in the absence of strong state institutions. Cross-national studies have similarly found a correlation between ethnic heterogeneity and violent crime (Hansmann & Quigley, 1982).

**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 Chiricos, 1987 and Wilkinson, 2009 for reviews), as well as with persecution of accused witches (Miguel, 2005), 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 (Boone, 2014), and in Liberia specifically (Blattman, Hartman & Blair, 2014).

**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 (Blakeslee & Fishman, 2015), domestic violence (Sekhri & Storeygard, 2013), and witch killings (Miguel, 2005), among other outcomes (see Hsiang, Burke & Miguel, 2013 for a review). More recent research has focused on testing the predictive power of climate change more generally (e.g. Witmer et al., 2017 in this issue).

**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 counter-intuitive salutary effects on cooperation, altruism, and civic life (see Bauer et al., 2016 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 (Tibshirani, 1996). 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 overfitting – 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 (Breiman, 2001). 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 all trees, 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. Beck, King & Zeng, 2000). They make no parametric assumptions, instead using an iterative algorithm to approximate the underlying structure of the data. The algorithm begins by constructing different non-linear 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 five 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 – that is, 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 reported incidents of local violence in the past is a poor predictor of new incidents in the future, and that including a 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 (Hastie, Tibshirani & Friedman, 2009). 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 fitted 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 (Blair, Blattman & Hartman, 2012).[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 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.

---

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

[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.

*Evaluating model performance*
Forecasting involves a trade-off between sensitivity (true positives) and specificity (true negatives). For continuous predictions, this trade-off is regulated by the 'discrimination threshold' above which we predict our outcome (violence) 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 all 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, for example by combining multiple events into a single indicator for 'political instability' (Goldstone et al., 2010) or 'adverse regime change' (Beger, Dorff & Ward, 2016). 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.

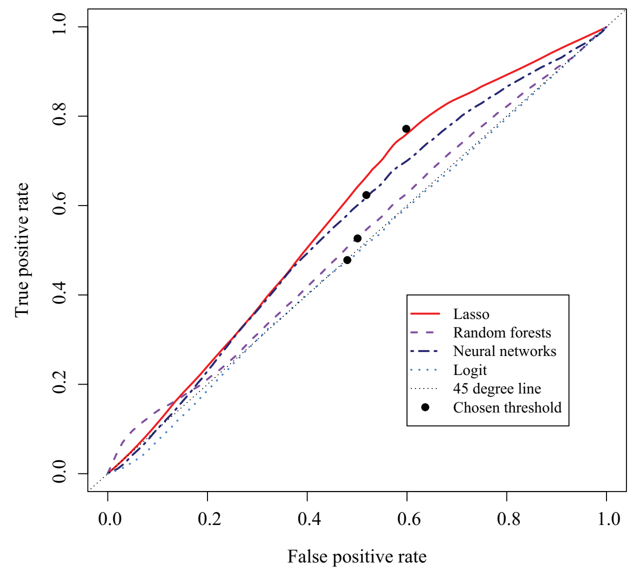


Figure 1. ROC curves for cross-validated forecasts of 2010 local violence using 2008 risk factors

*Cross-validated forecasts*
Figure 1 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 preferred threshold would have required increasingly large penalties in terms of specificity and accuracy.) In addition to the AUC and Brier score, Table I reports five

[8] By way of comparison, Goldstone et al. (2010) 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.

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) | |

performance metrics at the thresholds indicated by the solid circles in Figure 1. 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 later, 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 fewer 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 overpredicting 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 ten 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 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 2 plots ROC curves for our 'true forecasts' of 2012 local violence using 2010 risk factors. Table II

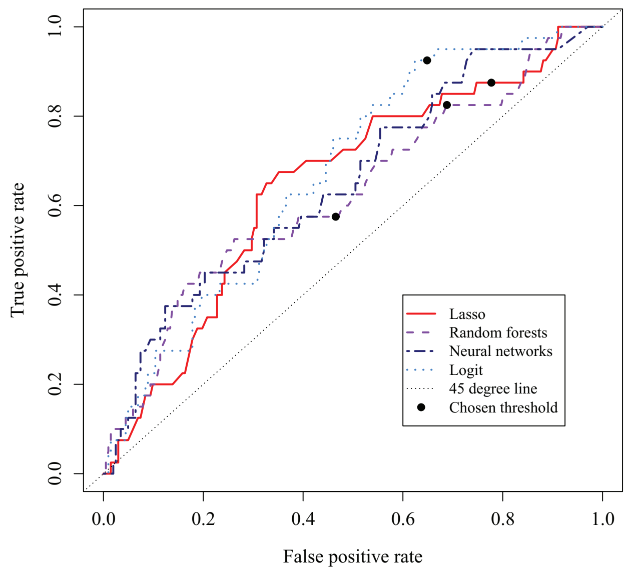


Figure 2. ROC curves for true forecasts of 2012 violence using 2010 risk factors

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 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 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 nonviolent incidents. Our qualitative follow-up revealed seven such incidents, but the performance of our models is virtually unchanged if we recode these as zeros.

### *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 overfitting[9] and reducing the stability of the results, which should thus be interpreted with caution.

### *Averaging across models*

Finally, recent research suggests that aggregating forecasts across multiple models can yield more accurate results (e.g. Beger, Dorff & Ward, 2016). 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.

[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 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<br>(0.05) | 0.63<br>(0.05) | 0.62<br>(0.05) | 0.67<br>(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 |

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 |

*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 III 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 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 ten 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 Hsiang, Burke & Miguel, 2013 for a review). Also unexpected is the absence of the lagged dependent variable from the top ten 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 III 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 right-hand-side variable. (Power-sharing appears among the top ten random forests predictors as well, but is lower ranked.) Yet, counter-intuitively, 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 analyses in Table III 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, 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 – for example, 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. Beger, Dorff & Ward, 2016; Blair & Sambanis, 2016; or, in this special issue, Chiba & Gleditsch, 2017). 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 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 (Cederman, Wimmer & Min, 2010), and that power-sharing helps mitigate the risk of violence (e.g. in this special issue, Cederman, Gleditsch & Wucherpfennig, 2017). Others, however, have

[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.

warned of the dangers of power-sharing (Rothchild & Roeder, 2005) or of cohabitation between equally dominant ethnic or religious groups (Balcells, Daniels & Escribà-Folch, 2016), and have suggested that violence often accompanies minority access to political power (Dancygier, 2010). 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 (Lake & Rothchild, 1996), and that past conflicts continue to predict future ones. A third is that the correlation is spurious – an artifact of overfitting. 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 under-explored frontier as one of the discipline's most promising.

## Replication data

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

## Acknowledgements

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.

## Funding

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 Bank's Italian Children and Youth Trust Fund.

## References

Autesserre, Severine (2010) *The Trouble with the Congo: Local Violence and the Failure of International Peacebuilding*. Cambridge: Cambridge University Press.

Balcells, Laia; Lesley-Ann Daniels & Abel Escribà-Folch (2016) The determinants of low-intensity intergroup violence: The case of Northern Ireland. *Journal of Peace Research* 53(1): 33–48.

Bauer, Michal; Christopher Blattman, Julie Chytilová, Joseph Henrich, Edward Miguel & Tamar Mitts (2016) Can war foster cooperation? *Journal of Economic Perspectives* 30(3): 249–274.

Beck, Nathaniel; Gary King & Langche Zeng (2000) Improving quantitative studies of international conflict: A conjecture. *American Political Science Review* 94(1): 21–35.

Beger, Andreas; Cassy L Dorff & Michael D Ward (2016) Irregular leadership changes in 2014: Forecasts using ensemble, split-population duration models. *International Journal of Forecasting* 32(1): 98–111.

Blair, Robert & Nicholas Sambanis (2016) Forecasting civil wars. Working paper.

Blair, Robert; Christopher Blattman & Alexandra Hartman (2012) Predicting local level violence. Working paper (https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2856479).

Blakeslee, David S & Ram Fishman (2015) Weather shocks, agriculture, and crime: Evidence from India. Working paper (https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2428249).

Blattman, Christopher & Jeannie Annan (2016) Can employment reduce lawlessness and rebellion? A field experiment with high-risk men in a fragile state. *American Political Science Review* 110(1): 1–17.

Blattman, Christopher; Alexandra C Hartman & Robert A Blair (2014) How to promote order and property rights under weak rule of law? An experiment in changing dispute resolution behavior through community education. *American Political Science Review* 108(1): 100–120.

Boone, Catherine (2014) *Property and Political Order in Africa: Land Rights and the Structure of Politics*. New York: Cambridge University Press.

Brass, Paul R (1997) *Theft of an Idol: Text and Context in the Representation of Collective Violence*. Princeton, NJ: Princeton University Press.

Breiman, Leo (2001) Random forests. *Machine Learning* 45(1): 5–32.

Cederman, Lars-Erik; Kristian Skrede Gleditsch & Julian Wucherpfennig (2017) Predicting the decline of ethnic civil war: Was Gurr right and for the right reasons? *Journal of Peace Research* 54(2): XX–XX.

Cederman, Lars-Erik; Andreas Wimmer & Brian Min (2010) Why do ethnic groups rebel? New data and analysis. *World Politics* 62(1): 87–119.

Chiba, Daina & Kristian Skrede Gleditsch (2017) The shape of things to come? Expanding the inequality and grievance model for civil war forecasts with event data. *Journal of Peace Research* 54(2): XX–XX.

Chiricos, Theodore G (1987) Rates of crime and unemployment: An analysis of aggregate research evidence. *Social Problems* 2(1): 187–212.

Dancygier, Rafaela M (2010) *Immigration and Conflict in Europe*. Cambridge: Cambridge University Press.

Fearon, James & Anke Hoeffler (2014) *Benefits and Costs of the Conflict and Violence Targets for the Post-2015 Development Agenda*. Copenhagen: Copenhagen Consensus Center (http://www.copenhagenconsensus.com/sites/default/files/conflict_assessment_-_hoeffler_and_fearon_0.pdf).

Goldstone, Jack A; Robert H Bates, David L Epstein, Ted Robert Gurr, Michael Lustik, Monty G Marshall, Jay Ulfelder & Mark Woodward (2010) A global forecasting model of political instability. *American Journal of Political Science* 54(1): 190–208.

Hansmann, Henry B & John M Quigley (1982) Population heterogeneity and the sociogenesis of homicide. *Social Forces* 61(1): 206–224.

Hastie, Trevor; Robert Tibshirani & Jerome Friedman (2009) *The Elements of Statistical Learning*. New York: Springer.

Hirose, Kentaro; Kosuke Imai & Jason Lyall (2017) Can civilian attitudes predict civil war violence? Ideology and insurgent tactical choice in civil war. *Journal of Peace Research* 54(1): 47–63.

Horowitz, Donald L (2001) *The Deadly Ethnic Riot*. Berkeley, CA: University of California Press.

Hsiang, Solomon M; Marshall Burke & Edward Miguel (2013) Quantifying the influence of climate on human conflict. *Science* 341(1235367).

Isser, Deborah H; Stephen C Lubkemann & Saah N'Tow (2009) *Looking for Justice: Liberian Experiences with and Perceptions of Local Justice Options*. Washington, DC: United States Institute for Peace.

Lake, David A & Donald Rothchild (1996) Containing fear: The origins and management of ethnic conflict. *International Security* 21(2): 41–75.

Miguel, Edward (2005) Poverty and witch killing. *Review of Economic Studies* 72(4): 1153– 1172.

O'Brien, Sean P (2013) A multi-method approach for near real time conflict and crisis early warning. In: VS Subrahmanian (ed.) *Handbook of Computational Approaches to Counterterrorism*. New York: Springer, 401–418.

Onoma, Ato Kwamena (2013) *Anti-Refugee Violence and African Politics*. New York: Cambridge University Press.

Posen, Barry (1993) The security dilemma and ethnic conflict. In: Michael E Brown (ed.) *Ethnic Conflict and International Security*. Princeton, NJ: Princeton University Press, 103–125.

Putnam, Robert D; Robert Leonardi & Raffaella Y Nanetti (1994) *Making Democracy Work: Civic Traditions in Modern Italy*. Princeton, NJ: Princeton University Press.

Ross, Michael L. (2004) What do we know about natural resources and civil war? *Journal of Peace Research* 41(3): 337–356.

Rothchild, Donald & Philip G Roeder (2005) Power sharing as an impediment to peace and democracy. In: Philip G Roeder & Donald Rothchild (eds) *Sustainable Peace: Power and Democracy after Civil Wars*. Ithaca, NY: Cornell University Press, 29–50.

Schrodt, Philip A & David Van Brackle (2013) Automated coding of political event data. In: VS Subrahmanian (ed.) *Handbook of Computational Approaches to Counterterrorism*. New York: Springer, 23–49.

Sekhri, Sheetal & Adam Storeygard (2013) Dowry deaths: Consumption smoothing in response to climate variability in India. Working paper (http://people.virginia.edu/∼ss5mj/womencrimes_2013.pdf).

Spear, Joanna (2006) Disarmament, demobilization, reinsertion and reintegration in Africa. In: Oliver Furley & Roy May (eds) *Ending Africa's Wars: Progressing to Peace*. London: Ashgate, 63–80.

Tibshirani, Robert (1996) Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)* 58(1): 267–288.

Varshney, Ashutosh (2002) *Ethnic Conflict and Civic Life: Hindus and Muslims in India*. New Haven, CT: Yale University Press.

Vinck, Patrick; Phuong Pham & Tino Kreutzer (2011) *Talking Peace: A Population-Based Survey on Attitudes about Security, Dispute Resolution, and Post-Conflict Reconstruction in Liberia*. Berkeley, CA: UC Berkeley Human Rights Center.

Ward, Michael D; Brian D Greenhill & Kristin M Bakke (2010) The perils of policy by p-value: Predicting civil conflicts. *Journal of Peace Research* 47(4): 363–375.

Weidmann, Nils B & Sebastian Schutte (2017) Using night light emissions for the prediction of local wealth. *Journal of Peace Research* 54(2): XX–XX.

Wilkinson, Steven I (2009) Riots. *Annual Review of Political Science* 12(1): 329–343.

Wischnath, Gerdis & Halvard Buhaug (2014) Rice or riots: On food production and conflict severity across India. *Political Geography* 43: 6–15.

Witmer, Frank DW; Andrew M Linke, John O'Loughlin, Andrew Gettelman & Arlene Laing (2017) Subnational violent conflict forecasts for sub-Saharan Africa, 2015–65, using climate-sensitive models. *Journal of Peace Research* 54(2): XX–XX.

ROBERT A BLAIR, b. 1982, 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, b. 1974, PhD in Economics (University of California, Berkeley, 2007); Ramalee E Pearson Professor of Global Conflict Studies, University of Chicago, Harris School of Public Policy.

ALEXANDRA HARTMAN, b. 1983, PhD in Political Science (Yale University, 2015); Lecturer in Qualitative Research Methods, University College London.