The Pearson Institute Discussion Paper

The Promise and Pitfalls of Conflict Prediction: Evidence from Colombia and Indonesia

Samuel Bazzi

Robert A. Blair

Christopher Blattman

Oeindrila Dube

Matthew Gudgeon

Richard Peck

2020-2

THE PEARSON INSTITUTE
FOR THE STUDY AND RESOLUTION OF GLOBAL CONFLICTS

The Promise and Pitfalls of Conflict Prediction: Evidence from Colombia and Indonesia

Samuel Bazzi
Boston University

Robert A. Blair
Brown University

Christopher Blattman
University of Chicago

Oeindrila Dube
University of Chicago

Matthew Gudgeon
U.S. Military Academy

Richard Peck
Northwestern University *

Abstract

How feasible is violence early-warning prediction? Columbia and Indonesia have unusually fine-grained data. We assemble two decades of local violent events alongside hundreds of annual risk factors. We attempt to predict violence one year ahead with a range of machine learning techniques. Our models reliably identify persistent, high-violence hot spots. Violence is not simply autoregressive, as detailed histories of disaggregated violence perform best, but socioeconomic data substitute well for these histories. Even with unusually rich data, however, our models poorly predict new outbreaks or escalations of violence. These "best case" scenarios with annual data fall short of workable early-warning systems.

---

*Bazzi: sbazzi@bu.edu; Blair: robert_blair@brown.edu; Blattman: blattman@uchicago.edu; Dube: dube@uchicago.edu; Gudgeon: matthew.gudgeon@westpoint.edu; Peck: richardpeck2021@u.northwestern.edu. For comments, we thank Austin Wright and seminar participants at ESOC, MWIEDC, ISF, NBER Economics of National Security, and NEUDC. Miguel Morales-Mosquera and Sebastian Hernandez provided excellent research assistance.

1 Introduction

Advances in data and computing techniques have kindled hopes that civil society, police, or peacekeepers will be able to predict costly violence ahead of time. Such early warning systems could be used to target scarce security personnel and resources, and prevent violence from occurring or escalating.

Until recently, prediction focused on large-scale, country-level events, including coups, civil wars, and terror attacks.[1] These macro-level efforts have informed policy, the science of prediction, and our understanding of violence. But such high-level predictions are not easy to act on. Scholars such as Cederman and Weidmann (2017) argue that country-level conflict predictions are unlikely to improve much in the future: there is simply too much complexity and randomness, they argue, to develop reliable forecasts over such wide time and space.

Subnational predictions could prove more fruitful. The past decade has seen the study of conflict push down to the micro-level causes, processes, and consequences of violence, and we can avail ourselves of these data to investigate prediction. Policy options to prevent an ethnic riot or local unrest are likely better than policy options to prevent a civil war. The feasibility of these early warning systems is unknown, however. Now is a good moment to take stock of what existing methods and the richest available micro data can deliver.

This paper takes advantage of high-quality, extensive annualized data in two countries, Colombia and Indonesia. Both countries have been ravaged by violence for decades—a

[1]The Political Instability Task Force's prediction efforts are likely the most well known (Goldstone et al., 2010). For other examples of cross-national prediction studies, see Beck et al. (2000); Blair and Sambanis (2020); Brandt et al. (2011); Celiku and Kraay (2017); Gleditsch and Ward (2013); Gurr and Lichbach (1986); Harff (2003); Hegre et al. (2013, 2016); Perry (2013); Ward et al. (2013). For an early exception see Schrodt (2006) who studies violence in the Balkans.

situation that typically does not bode well for data availability. But both countries are also wealthy enough (and have strong enough states and research communities) to produce some of the highest-quality, local-level panel data in the developing world. This includes a trove of information on annual socioeconomic conditions and other characteristics, plus more than a decade of subdistrict- or municipal-level data on annual local violence.

We chose these two cases because they are among the current "best case" scenarios in terms of data availability on both violence and potential predictors of violence, in annual panel form. Such annual data are, to date, the most common kind of data available across countries (especially for predictor variables). If conflict prediction proves fruitful in these two instances, they could be models for other prediction efforts. If not, then we must ask where or with what other forms of data we can expect early warning systems to bear fruit (if any). Finally, both countries have suffered recurring episodes of local violence during transitions to national peace. Anticipating and preventing these episodes is of both substantive and practical importance.

We identified, collected, and merged dozens of subnational datasets in each country. This gives us an unusually rich array of hundreds of covariates per locality, including covariates that the empirical and theoretical literatures commonly associate with conflict onset and escalation (Blattman and Miguel, 2010). Each country also has high quality violence data from which we obtain our outcomes. Using data from 1998–2014 in Indonesia, we study conflict related to interethnic and religious tensions, as well as electoral, governance, and resource disputes, among others. In Colombia, our data span more than a quarter century, 1988–2014. We predict clashes between state, guerrilla, and paramilitary forces during a period of protracted civil conflict. These granular violence data also allow us to leverage detailed violence histories, adding to an already rich set of predictors.

We then use several machine learning methods to generate predictions of local violent incidents at the annual level. In our main year-ahead predictions, we train the algorithms on four to thirteen years of data, and forecast local conflict during the following year. We

generate and assess predictions for whether there will be any violence next year as well as whether there will be a large number of events or an escalation. We further examine predictive power across space as well as time.

Our results illustrate both the promise and pitfalls of annual local violence forecasting. An ensemble of machine learning models effectively identifies locations at risk of having a violent incident. We are particularly effective at identifying "hot spots" with high concentrations of violence, defined as five or more incidents in a single year. Indeed, our ensemble model, which combines the best new methods, performs better than previous subnational attempts (Blair et al., 2017; Colaresi et al., 2016; Weidmann and Ward, 2010; Witmer et al., 2017).[2] We view these results as especially important given that such local hot spots can pose a serious risk of regional or national escalation, and some of these locations will not be known to policymakers, especially in large, diverse countries.

Local violence is not merely autoregressive: a model consisting of the lagged dependent variable alone performs consistently poorly. Nonetheless, more nuanced histories of violence tend to be very strong predictors of violence. In other words, to predict future violence, it is not enough to know where violence occurred in the past. But detailed and disaggregated histories of violence—including the severity of particular incidents (e.g., number of deaths, property damage) and the identity of the actors involved—perform very well.

Even without such detailed and disaggregated histories, however, our covariates also predict hot spots well. This suggests that much of the information contained in these violence

[2]Our improved performance has multiple explanations. We exploit a much longer panel and a much wider variety of data sources than Blair et al. (2017). We also test a more diverse set of prediction algorithms, and ensemble routines have been shown to surpass the performance of any given model (Montgomery et al., 2012). Additionally, our event-based data rely on local media and are more comprehensive and less prone to misreporting than the widely-used ACLED data. Finally, we draw on a much wider set of subnational predictors that go beyond lagged violence and country-level predictors.

histories is representative of observable characteristics of the units in our two samples. The most predictive risk factors tend to be slow-moving or time-invariant. In Colombia, for example, one of the most reliable predictors is terrain ruggedness. In Indonesia, robust predictors include remoteness as well as sectoral shares of the local economy. Time-invariant predictors alone do just as well detailed violence histories.

Surprisingly, predictive accuracy improves little when we add time-varying factors, including economic output, government finance, communication infrastructure, natural disasters, elections, and fluctuations in rainfall, temperature, commodity prices, drug production, and U.S. military activity. This stands in contrast to a large causal literature on conflict, where an array of findings associate economic and political shocks with intensified violence (Bazzi and Blattman, 2014; Berman and Couttenier, 2013; Blattman and Miguel, 2010; Burke et al., 2015; Dube and Vargas, 2013; Miguel et al., 2004).[3]

Our algorithms' strong performance is mainly driven by forecasts of where, but not when, violence is likely to occur. Our models perform poorly when predicting annual deviations from average levels of violence over the study period. Even nuanced histories of violence, a rich set of covariates, and the most widely studied economic and political shocks do not help us identify what hot spots are likely to get hotter in the coming year.

In contrast, our models perform well when we forecast conflict across space. We use training data from all years in one set of locations to predict conflict in another set of locations. In these cross-location predictions, time-varying shocks typically improve performance. The models leverage the longest time series currently available at the local level. This leads us to believe that a lack of common support in the training and testing periods may explain some of the limited predictive performance of time-varying shocks when we attempt to forecast

[3]We find similar associations of commodity price and weather shocks with conflict in Colombia and Indonesia using our datasets (see Appendix C.4). While these shocks may cause conflict and help us forecast conflict hotspots, they add little to our ability to forecast conflict over time.

conflict over time. Thus, early warning performance should improve over time, but this (by definition) means that better one-year-ahead predictions may be a long way off.

We see these patterns consistently across two different country cases, with very different forms of violence. Taken together, our results are both encouraging and disappointing. On the one hand, we are able to predict hot spots for local violence remarkably well, and much more accurately than previous exercises of this sort or simpler benchmark models. Anticipating where violence is most likely to occur is potentially highly valuable to resource-constrained governments in conflict-affected states.

On the other hand, early warning systems would ideally be able to predict not just the location but also the timing of new outbreaks of violence. Our inability to do so, with some of the richest and most systematic subnational annual panel data in the developing world, is important but disheartening news for conflict forecasters.

One interpretation of our results is that local conflict prediction is less fruitful than hoped, at least with the data most commonly at hand: human-coded violence data drawn from newspapers and observer reports, annual aggregates of violence counts, and annual predictor variables. This pessimistic conclusion would resonate with warnings that big data and machine learning may not deliver the precision that policymakers long for (Jasny and Stone, 2017, p. 469).[4] It also aligns with a view that conflict breaks out for largely idiosyncratic reasons, as reflected in the warning by Gartzke (1999) that "war is in the error term".

Another interpretation is that early warning systems are feasible, but require longer, more high-frequency data, or additional or different risk factors. Longer training samples could give algorithms more variation to train on, and in principle this could help them to identify more complex relationships with time-varying predictors. If true, this implies that there will be large gains to collecting longer conflict time series, for example, by delving into

[4]Of course, conflict may be more difficult to predict than other policy-relevant outcomes. For example, Guha and Ng (2019) strike a more hopeful note in the context of predicting retail sales in a high-frequency, subnational panel.

past historical archives. That said, our longest training samples are already decades long, and longer time series can introduce their own challenges, like potential structural breaks in the violence-generating process. There could also be gains from using higher-frequency data, or more data on new risk factors—data from mobile phones, social media, local price fluctuations, and so forth.

A case in point is work by Mueller and Rauh (2017), which successfully uses topics analyses from newspapers to forecast conflict within countries. Likewise, Berger et al. (2014) use cell phone call patterns to predict temporal variation in conflict. These data innovations, and our increasing ability to collect newer and wider forms of big data, may enhance our capacity to forecast conflict over time.

Our paper offers several results that future work should leverage to explore the promise of machine learning methods for modeling and predicting conflict. Overall, conflict can be forecasted well across space, but not over time with annualized panel data. While violence is not simply autoregressive, detailed conflict histories can substitute for a broader array of covariates, which are potentially more expensive to collect. But if such detailed histories of violence are unavailable (as is the case in most countries), a more limited set of common or easy-to-measure covariates can also predict hot spots remarkably well, at least in our two cases. Cross-sectional hot spot prediction systems are probably feasible in a wide range of countries, even if temporal early warnings system may not be.

2 Settings

2.1 Indonesia

Following the 1998 collapse of Suharto's authoritarian regime, Indonesia experienced large-scale collective violence.[5] Separatist movements in Aceh, East Timor (as it parted from

[5]For detailed accounts see Barron et al. (2014, 2016); Tadjoeddin (2014).

Indonesia), and Papua resulted in over 10,000 deaths. At the same time, religious and ethnic conflict reached new highs.

Collective violence abated by 2003, and the separatist conflict in Aceh ended in 2005. Post-2004, there were far fewer fatalities, and the composition of violence shifted as electoral and resource-related violence rose. The violence also had different consequences: after 2004 it was more likely to lead to injuries and property damage than deaths. Deadly conflict nevertheless remained prevalent across the archipelago, primarily concentrated in regions with histories of large-scale violence.

Scholars debate the drivers of today's conflict in Indonesia, highlighting fixed factors like ethnicity and religion, as well as local resources like forests, minerals, and plantation crops. Regional variation in violence has been linked to political and economic shocks associated with decentralization and electoral reforms (Bazzi and Gudgeon, forthcoming; Pierskalla and Sacks, 2017), with economic inequality (Indra et al., 2019), and with natural disasters, weather shocks, and commodity price fluctuations (Barron et al., 2009; Wright and Signoret, 2016). While the literature has identified a variety of proximate causes, it is not clear which of these factors are the best predictors of violence.

2.2 Colombia

Colombia's long-running civil war has resulted in 220,000 deaths and 25,000 disappearances, and the forced displacement of over five million civilians (Historical Memory Group, 2013). During our primary analysis period, 1988–2005, the conflict mainly involved left-wing guerrilla groups, the government military, and right-wing paramilitary groups. The insurgency was launched by communist guerrillas in the 1960s. Paramilitaries arose in the 1980s when landowners organized in response to extortion and violence perpetrated by the guerrillas. Paramilitaries and the government colluded extensively, though their relationship varied over time and space.

Low violence levels prevailed through the 1980s, but escalated in the 1990s when paramil-

itaries expanded and centralized authority. Intensity remained high until the paramilitaries demobilized, a process that began in 2003 and continued until 2006, when the main paramilitary organization officially disbanded. The conflict subsided further as the death of a number of guerrilla leaders weakened their respective groups. It drew to an official end in 2016, when the largest guerrilla group signed a peace deal with the government, though residual violence (including killings of civil society leaders) continues.

Scholars have linked regional variation in conflict in Colombia to a host of political and economic factors, including shocks to drug production (Angrist and Kugler, 2008), fluctuations in commodity prices (Dube and Vargas, 2013), revenue decentralization (Chacon, 2014), collusion between paramilitaries and politicians (Acemoglu et al., 2013), American military aid (Dube and Naidu, 2015), and military incentives in the targeting of civilians (Acemoglu et al., 2016). As in Indonesia, it is unknown whether these causal factors are also good predictors of violence.

3 Data

An important contribution of this study is the two datasets we assembled. In each country, we collected and stitched together dozens of local-level data sets, most of which had not been consolidated before. The result is a uniquely rich trove of data that can be used for purposes of prediction.

3.1 Indonesia

Our units of analysis are Indonesia's third-tier administrative divisions, known as subdistricts or *kecamatan*. The country had 7,094 subdistricts in 514 districts in 2014. These subdistricts had a median population of around 22,000.[6] While districts are the key au-

[6]To deal with Indonesia's pervasive administrative unit proliferation, we harmonize all observations to boundaries in 2000.

tonomous administrative units, responsible for providing major public goods, subdistricts are also important sites of political organization. They are also the most granular level at which violence can be systematically tracked over time.

Subdistrict-Level Violence Data. Our main measures of violence come from the Indonesian National Violence Monitoring System (known by its Indonesian acronym, SNPK). Coverage begins in 1998 for nine conflict-prone provinces and increases to 15 provinces plus parts of 3 provinces in greater Jakarta beginning in 2005. The data is not formally representative of Indonesia, but by 2005 it spans all major island groups and covers a majority of the population.

The SNPK is built from local media reports of violence. SNPK researchers collected all available print archives of 120 local newspapers, recording over 2 million images. Coders then used a standardized template to code each incident based on the underlying trigger, beginning with broad groupings: domestic violence, violent crime, violence during law enforcement, and conflict. Within conflict, the coders further sorted into identity, elections/appointments, governance, resource violence, popular justice, separatist, and other (could not be classified). Appendix Table C.1 defines each of these.

SNPK offers uniquely rich data on violence at the micro level. In 2014, the architects of SNPK wrote "as far as we know, the [SNPK] is the largest dataset of violence created for any single country" (Barron et al., 2014). Barron et al. provide additional detail on the multi-year process involved in creating this data set, including the source selection process and various quality control measures (Barron et al., 2014, 2016).[7]

We also draw on additional measures of violence from a triennial administrative census of villages known by its acronym, *Podes*. *Podes* asks local government officials about a host

[7]The data and supporting documentation can be accessed here: https://microdata.worldbank.org/index.php/catalog/2626.

of village characteristics, including recent violent events.[8]

Outcome Measurement. Our main outcome is an indicator for any "social conflict." This groups all of the various forms of violence, except domestic violence and crime, into one category. It guards against miscoding of conflict triggers. Predictive performance is similar when retaining domestic violence and crime.

In addition to indicators of any social conflict, which occur in around half of the subdistricts each year, we also predict an indicator for at least five social conflict incidents in a given year—a "hot spot". This is meant to capture higher intensity episodes. These episodes occur in around 10% of subdistricts each year. We predict indicators rather than counts in order to simplify the interpretation of performance: the models either correctly predict the incident or they do not. We predict counts in Appendix A.2; this exercise does not meaningfully change our conclusions.

Of course, levels of violence tend to be persistent, and we are often interested in predicting the onset and escalation of violence after a period of peace. There is no natural definition of "onset" here, as in the civil war literature, since there are no discontinuities in subnational event-level data.[9] Instead, we construct an indicator for a standard deviation increase in violence since the previous year, and seek to predict this escalation. A standard deviation increase is around 4.7 acts of violence in a year, and we observe an increase of this size in 3.3% of subdistrict–years.[10]

---

[8]To the extent that local leaders face strategic incentives to misreport violence, *Podes* measures may be more biased than those from external media reporting (for discussion, see Barron et al., 2014).

[9]In Appendix A.5, we consider prediction of future violent events in places that are not currently experiencing violence.

[10]In Appendix A.6, we consider alternative magnitudes and time horizons of escalations. We generally find the measure of a one standard deviation increase in incidents to be

Covariates. In addition to detailed violence histories from SNPK and *Podes*, we assemble a set of 482 subdistrict-level predictors from multiple data sources.[11] In order to assess how predictive performance differs across related covariate sets, we group covariates into the following predictor groups: (a) Population, (b) Religion, (c) Ethnicity, (d) Demographics (e.g., fraction male, fraction young), (e) Education (e.g., mean years of schooling, school presence), (f) Health (e.g., doctor and facility presence, rates of self-reported health problems), (g) Geography (e.g., soil quality, ruggedness), (h) Remoteness (e.g., distance to the capital, road presence, transportation terminals), (i) Sector Shares (e.g., population share in industry or agriculture), (j) Agricultural Features (e.g., irrigation access, share of households with agricultural land), (k) Public Goods (e.g., safe water, garbage disposal, electricity and gas sources), (l) Output (e.g., night light intensity, district Gross Domestic Product), (m) Distributional Measures (e.g., inequality and poverty), (n) Communication (e.g., access to telephones, cell signal), (o) Government Revenues and Expenditures, (p) Electoral Outcomes (e.g., vote share polarization), (q) Natural Disasters, (r) Weather Histories and Shocks, and (s) Commodity Shocks (e.g., food, cash crop, and mineral price shocks).

We use time-varying covariates whenever possible. Some, like geographic features, are time-invariant by nature. Others, like ethnic and religious shares from the 2000 Population Census, were only observed once at the beginning of the study period, and so are time-invariant in our panel. Covariates derived from *Podes* are time-varying but only observed triennially. Other variables, like education and health infrastructure, are slow moving by nature. The violence data, night lights, economic shocks, as well as district-level revenues, GDP and unemployment, and poverty and inequality, among others, vary annually. Our models also include two lags of all violence predictors as well as several lags of our other

most predictable, but the differences when predicting other definitions of escalation are not substantial.

[11] Unless specified otherwise in Appendix C.1, these measures are available at the subdistrict level or finer, and are aggregated to their subdistrict boundaries in 2000.

time-varying predictors subject to availability.[12] We assess how predictive performance varies when using only time-invariant or time-varying covariates.

Details on sources and variable construction can be found in Appendix C.1. Summary statistics for each covariate, organized by predictor group, can be found in Appendix C.3.

3.2 Colombia

Our unit of analysis for Colombia is the municipality, or *municipio*. The country had 1,023 municipalities in our study period, averaging 37,000 residents—about two-thirds larger than the average Indonesian subdistrict. These are the main administrative units in the country, and municipalities play a key role in the allocation and contestation of public resources.

Municipality-Level Violence Data. The Conflict Analysis Resource Center (CERAC) provides data on armed confrontations from 1988 to 2005. We also extend beyond 2005 by using an additional source of conflict data collected by Universidad del Rosario through 2014 (see Appendix A.8).

The CERAC dataset contains over 21,000 events from the Colombian civil war, drawn from periodicals published by two Colombian NGOs, the Center for Research and Popular Education/Peace Program (CINEP) and *Justicia y Paz*. These periodicals, in turn, are based on two underlying sources: reports of political violence and human rights abuses that appear in 25 printed media outlets with local and national coverage, and a broad network of priests from the Catholic church with representation in almost every Colombian municipality, who report conflict episodes to CINEP. The priests are seen as neutral actors, often serving as negotiators between the two sides. Thus, their accounts are viewed as both credible and indispensable for attaining a comprehensive picture of conflict events in rural areas. Reliance on a large number of media sources as well as the priests leads to the inclusion of every part

[12]For the *Podes* variables, we use the two most recent measures of each conflict variable. Appendix C.1 details the number of lags included for each time-varying predictor.

of the Colombian territory: violence events are reported in over 950 of 1,000 Colombian municipalities over this 18-year period.

All events in the CERAC data are hand-coded and checked extensively to ensure accuracy, with the precise coding procedures documented in Restrepo et al. (2004). For example, all large events associated with double-digit casualties and a random sample of smaller events are cross-checked against the archives of the leading Colombian newspaper *El Tiempo*, to ensure that the data have been entered accurately, without double counting. The events are also cross-checked against databases from the National Police, Human Rights Watch, and Amnesty International.

CERAC also screens out events unrelated to the civil war (e.g., incidents of domestic violence or crime). The dataset instead hones in on war-related actions carried out by politically motivated armed groups. Events are coded as either (a) bilateral clashes between sides, or (b) unilateral attacks by any one side against another. The data separately categorize violence by the military, various paramilitary organizations, and several guerrilla groups (the largest being the FARC). Clashes occur between all three kinds of actors, though government versus paramilitary clashes are rare, as are clashes between guerrillas or between paramilitaries. The CERAC dataset typically does not include kidnapping of individuals under event categories, since kidnapping is commonly used by criminal groups, and tends to be underreported, which makes it difficult to measure comprehensively with accuracy (Restrepo et al., 2004). The exception is when kidnapping results directly from war-related actions, for example alongside other actions by an armed group that result in the classification of the event as an attack.

We only use CERAC data after 1992 for the training sample since a consistent set of covariates is unavailable before then. We further examine the possibility of prediction in a longer panel by combining the CERAC series through 2005 with the Universidad del Rosario dataset from 2006 to 2014. We present this in the appendix rather than the main analysis in part because there are some coding differences between the two datasets, which

do not make them perfectly comparable. For example, the Universidad del Rosario dataset includes additional political events, like the kidnapping of political actors, under the attack classification. In addition, it categorizes attacks and clashes differently for complex events involving the military. Nevertheless, predictive performance with the combined series is similar to the baseline using only CERAC (see Appendix A.8).

Outcome Measurement. We construct indicators of any attack or clash, analogous to the indicators for Indonesia. This grouping combines attacks initiated by the government with attacks initiated by other armed actors. Results are similar when we remove government-initiated violence.

In Colombia, our indicator of any conflict occurs in about one-third of municipalities each year. "Hot spots" with five or more incidents occur about 8% of the time. A standard deviation change is 3.4 events, and we observe such a change in 4.4% of municipality-years.

Covariates. As in Indonesia, we assemble a broad set of predictors from multiple sources: over 310 in all. First, we include detailed violence histories capturing incident types, actors, and outcomes. Beyond these violence histories, we incorporate the following predictor groups of related covariates: (a) Population, (b) Geography (e.g., terrain ruggedness), (c) Remoteness (e.g., road presence), (d) Distributional Measures (e.g., poverty and inequality), (e) Historical Traits (e.g., colonial population and infrastructure, following Acemoglu et al., 2015), (f) Demilitarized Zone Proximity, (g) Municipality Revenues and Expenditure, (h) Electoral Outcomes, (i) U.S. Military Presence and Spending (Dube and Naidu, 2015), (j) Drug Production and Drug Price Shocks, (k) Weather Histories and Shocks, and (l) Commodity Production and Price Shocks (as in Dube and Vargas, 2013).

As in Indonesia, we strive to use time-varying data when available. Electoral outcomes, municipality revenues and spending, drug production, U.S. military involvement, and commodity and weather shocks all vary annually. Our models include two lags of all violence predictors as well as lags of other time-varying predictors subject to availability. Summary

statistics for each covariate, organized by predictor group, can be found in Appendix C.3, while details on sources and variable construction can be found in Appendix C.2.

3.3 Comments on Data Quality

It is worth discussing why we selected Colombia and Indonesia, and why scholars regard their data as unusually high quality, especially the violence data. First, as we have already mentioned, it is rare to have such a large number of covariates systematically available in so many subnational units for such a long period. We know of few close comparisons in countries with histories of violence.

Second, both countries offer state-of-the-art subnational violence data. The SNPK, for example, was an explicit attempt to address shortcomings of earlier violence data sets within Indonesia (Barron et al., 2014). Both sources are meticulously documented and subject to various quality control measures (Restrepo et al., 2004; Barron et al., 2016). Each has been effectively deployed in the academic literature (see, e.g., Bazzi and Gudgeon, forthcoming; Dube and Vargas, 2013).

Third, the data draw on a large number of high-quality, local-language newspapers. This is rare, as most news-coded datasets draw mainly from international news services. In many low-income countries, the number and quality of local newspapers is extremely poor, and there is little news coverage in conflicted places. Hence, many events are simply never reported.

In fact, both countries' conflict data offer significantly more comprehensive coverage than other popular event-based data. The widely-used Uppsala Conflict Data Program (UCDP) Georeferenced Event Data (GED) (Sundberg and Melander, 2013), which relies on international media, misses over 99% of events in the SNPK and 61% of events in CERAC, many of which involve substantial casualties. We detail these comparisons with UCDP-GED in Appendix C.1.3 for Indonesia and Appendix C.2.3 for Colombia. Another popular source, the Armed Conflict Location & Event Data Project (ACLED) data, is available for a more

limited range of years, beginning only in 2015 for Indonesia and 2019 for Colombia.

Nevertheless, we do not wish to overstate the quality of the data. The main violence datasets undoubtedly omit or misclassify some events, or get the timing or location wrong, thus impeding accurate prediction. Coding procedures also change over time. In Colombia, for instance, the local sources and coding methods for the 2005–2014 data differ slightly from the original 1988–2005 data, which is why we only include the later period in the Appendix. Our results are not particularly sensitive to the way in which we create a longer series by tying them together, but the issue is indicative of the challenges one can expect from quarter-century long conflicts and data collection.

Another issue is that the data are typically aggregated to the level of a year, or to a subdistrict or municipality. This introduces sources of measurement error in timing or location that could, in principle, reduce predictive power. It is possible that we need to reinvent the way that conflict scholars collect and code data in order to facilitate early warning. But that would be a huge undertaking. This illustrates the purpose of this paper: to evaluate how prediction works when applied to unusually rich existing data, before radically changing data approaches to data collection. We are careful to limit our conclusions to the specific kind of data that is currently available and one-year-ahead prediction exercises.

4 Prediction Methods

4.1 Training and Testing

For each year t we forecast violence in year $t + 1$. While the events are coded with specific dates, we aggregate to the annual level because few predictors are measured at sub-annual frequency, and because disaggregation would exacerbate a class imbalance problem (i.e., the fact that there are far more non-events than events). Our procedure is as follows:

1. For each model, we take predictors measured from t_0 to $t - 1$ as our training set, and violence measures up to and including period t are the training outcomes. That is,

violence in period t is matched to predictors in period $t-1$, period $t-1$ violence is matched to period $t-2$ predictors, etc., and we have $t-1$ observations per location.

2. We use 5-fold cross validation to choose optimal tuning parameters specific to each machine learning algorithm (see Section 4.2). We choose tuning parameters to maximize out-of-sample area under the receiver operating characteristic, or ROC, curve (AUC), a metric detailed in Section 4.3. 5-fold cross validation simulates out-of-sample prediction. First, the data is randomly partitioned into 5 equal-sized subsamples. A model is then fit to four subsamples and used to predict violence in the fifth. This is repeated for each of the five subsamples, so that there is an "out-of-sample" prediction of each observation. We replicate this exercise for each tuning parameter value in the parameter search space. The best performing parameter, in terms of AUC, is chosen.

3. We repeat step 2 ten times with different random partitions, in order to generate 10 "optimal" tuning parameters, and then we take the average over these 10 trials.[13]

4. Using the selected tuning parameter values, we fit the model to the entire training set.

5. With this fitted model, we use the predictors measured in year t and the estimated parameters to forecast violence in year $t+1$.

We generate out-of-sample predictions starting in 2008 (and ending in 2014) for Indonesia and in 1998 for Colombia (ending in 2005). We use all data up to the test year to generate out-of-sample forecasts. So, to predict violence in year $t+1$, we train each model on data through year t; to predict violence in $t+2$, we train each model on data through $t+1$; etc.

[13]In Appendix B.2, we assess the extent to which multiple cross-validation rounds add stability to the hyper-parameter choice. We conclude that, for the most part, one single cross-validation exercise is sufficient to choose a good hyper-parameter value. Nevertheless, this step is parallelizable and not costly in terms of computing time, so we use 10 cross-validation runs throughout the paper in order to be conservative.

For Indonesia, this procedure generates seven predictions per algorithm. The first is trained on four years of data to forecast conflict in 2008, and the last is trained on ten years of data to forecast conflict in 2014. For Colombia, we generate eight predictions per algorithm. The first is trained on six years of data to forecast conflict in 1998, and the last is trained on thirteen years of data to forecast conflict in 2005.

4.2 Machine Learning Algorithms

We apply several machine learning methods. Since each has its own strengths and drawbacks discussed below, we also take a weighted average of the four using an **Ensemble Bayesian Model Average** (Beger et al., 2016). Starting with the prior that each algorithm is equally appropriate, we use cross-validation to update our weights based on the accuracy of each model (Montgomery et al., 2012). Bayesian model averaging is especially important for our auxiliary analyses in which we explore different subsets of predictors and alternative prediction tasks. Since some procedures may be more or less suited to particular tasks, the model average considers the full potential of the algorithms as a whole.

1. **LASSO** (Tibshirani, 1994) is a logistic regression model that penalizes large coefficients and forces all but the most important to zero. This algorithm is the simplest of the four that we test, and arguably the least susceptible to overfitting. It is less suited to identifying complex relationships between covariates and outcomes. It is also most familiar to social scientists.

2. **Random Forests** comprise many independent decision trees. Each tree is a sequence of rules that splits the sample into subsets, called leaves, based on variable cutoffs. The prediction for each leaf is the mean outcome for the observations on that leaf, and trees are fit so as to minimize mean squared error. Each tree is constructed by sampling a random subset of the training data and a random subset of the predictors. Each of these trees generates a prediction, and the overall prediction of the Random Forest is

the average of the predictions from each tree. Random forests are very flexible, being able to model complicated interactions between variables. Random forests are also relatively straightforward in terms of tuning parameter selection, and have been used (albeit sparingly) in the conflict forecasting literature (Blair et al., 2017; Blair and Sambanis, 2020; Muchlinski et al., 2016).

3. **Gradient Boosted Machines** are a variant of Random Forests. Trees are fit neither randomly nor independently. Instead, each tree is fit sequentially to the full dataset, but observations are weighted by the error rates of previous trees in the forest, such that later trees are fit with a larger weight on observations that previous trees found difficult to predict. In this way, each new tree slightly improves the model (Freund and Schapire, 1999). Gradient boosted machines can improve upon random forests by fitting trees in a more targeted manner, but they also require more decisions about tuning parameters and are more susceptible to overfitting.

4. **Neural Networks** consist of systems of "nodes," which are each functions of predictors. The functions input a linear combination of predictors and output a value between zero and one. The outputs of these nodes are then further combined to produce a single output, with an organization evoking the structure of the human brain (Hastie et al., 2001). The optimization problem is to choose appropriate weights in each linear combination.[14] Neural networks are widely applied in industry and are best suited for the most complex classification tasks such as image and speech recognition.

[14]Because there is a separate set of weights for each node, the number of free parameters can grow very quickly. Since we do not have many observations relative to our predictor dimensionality, we must first cut down the number of predictors by taking principal components of the covariates. We use 30 principal components in Indonesia and 20 in Colombia. Predictive performance is not sensitive to these particular choices. In Appendix B.3 we explore alternative neural network architectures.

However, neural networks require relatively large datasets, carefully chosen network architecture, and computing resources to achieve high performance.

Appendix B reports further details about hyper-parameter choices and the mechanics of each algorithm.

While there is an enormous variety of additional algorithms we might have tested, we focus on these four because they are well established in the machine learning literature, because they have been used (albeit infrequently) for purposes of forecasting in economics and political science, and because they reflect much of the variation across the most prominent categories of machine learning models: selection and shrinkage techniques (LASSO), ensemble and tree-based techniques (Random Forests and Gradient Boosted Machines), and nonlinear adaptive weighting techniques (Neural Networks). Our goal is not to be exhaustive, but rather to evaluate the predictive power of well-established models applied to uniquely rich within-country data on conflict and its correlates.

4.3 Performance Metrics

To evaluate our models, we focus on the area under the ROC curve, known as the "area under the curve" or AUC. Other performance metrics such as the mean squared error, area under the precision-recall curve, and maximal accuracy and sensitivity are reported in Appendix A.1. ROC curves plot the tradeoff between true and false positives for a given model. The AUC captures the probability that a randomly chosen pair of observations is correctly ordered in terms of predicted risk of violence. A model that performs no better than chance would have an AUC of 0.5; a perfect model would have an AUC of 1.

An advantage of the AUC is that it does not require specifying a probability threshold above which we predict violence will occur. Selecting a specific threshold requires making a tradeoff between accuracy, sensitivity (the proportion of incidents correctly predicted), and specificity (the proportion of non-incidents correctly predicted). The threshold one chooses depends on one's relative tolerance for false positives and false negatives.

For example, a policymaker with ample resources might choose a low threshold, increasing sensitivity at the cost of specificity and accuracy, while a policymaker with scarce resources might choose a high threshold, increasing specificity at the cost of sensitivity and accuracy. We are more interested in overall performance than in performance at any given threshold, and so opt to focus on the AUC. But we recognize that the AUC has some limitations as well, especially in the presence of class-imbalanced data, and report alternative performance metrics in the appendix.[15]

5 Results

5.1 Next Year's Violence is Predictable

Table 1 shows that all of the machine learning methods we test have strong predictive performance. For the ensemble average (EBMA), the AUC is above 0.82 for predicting $\geq$1 event, above 0.91 for $\geq$5 events, and above 0.80 for escalations of $\geq$1 standard deviation. In general, AUCs of 0.8 and above are considered very good, and AUCs of 0.9 and above are considered excellent.

To fix ideas, given a random pair of Indonesian subdistricts in which one location experiences 5 or more incidents and the other does not, there is a 0.941 probability that the more violent subdistrict would have a higher predicted probability of violence. The lower AUC for escalations implies that changes are inherently more difficult to predict, and that increasing

[15] Appendix A.1 reports the MSE, the precision-recall-AUC, as well as accuracy, sensitivity, and specificity at two different thresholds, one that maximizes accuracy and one that maximizes sensitivity while keeping accuracy above 50%. Our results are qualitatively similar when we compare models using these alternative performance metrics rather than the AUC. Appendix section A.1 also reports alternative models that are trained to minimize MSE. Again, the results are qualitatively similar.

the number of true positives comes at the cost of more false positives.

These models exceed the performance of other recent subnational conflict forecasting exercises. By way of comparison, Blair et al. (2017) report a maximum out-of-sample AUC of 0.74 in a sample of 250 Liberian towns over three years, Weidmann and Ward (2010) achieve a maximum out-of-sample AUC of 0.78 in Bosnia 1992–95, and Witmer et al. (2017) find a maximum *in*-sample AUC of 0.85 across sub-Saharan Africa using 1 degree gridded monthly data, 1980–2012.[16] Gains in the range of 0.05 or 0.10 represent 10–20% of the difference between the worst and best possible prediction.

The models perform similarly well in Indonesia and Colombia, and performance is similar across algorithms. LASSO performs roughly as well as the more sophisticated algorithms, which is notable given its relative simplicity. The tree-based models also perform well, with gradient boosted machines edging out random forests in all instances. The neural networks are generally the worst performers. They may be ill-suited for this prediction task, which uses a relatively modest amount of data. They are also the most difficult to tune among our candidate algorithms. Finally, the ensemble (EBMA) is generally the best performer, giving us confidence in focusing on the ensemble as an indicator of the overall potential of these methods. That said, the gains from model averaging are not large relative to the individual top performers.

[TABLE 1 HERE]

5.1.1 Machine Learning Outperforms Simpler Benchmarks

Table 2 compares our machine learning approaches to simpler benchmarks. Column 1 reproduces the results from our main EBMA specification in column 5 of Table 1. Column 2 reports the performance of an OLS model using all predictor variables. Using our expansive

[16] *In-sample* performance refers to models that are trained and tested on the same data. *Out-of-sample* refers to models that are trained on one subset of data and tested on another.

dataset of conflict predictors, OLS alone does a good job of predicting conflict. However, our ensemble machine learning model outperforms simple OLS. The gains are moderate in the case of predicting any incident and more significant for the rarer outcomes. Linear regression models appear to over-fit the data: the generated predictions underperform those of the less flexible LASSO models in column 1 of Table 1. Moreover, the linear regression is not able to model interactions between variables, which may explain the outperformance of Random Forests and Gradient Boosted Machines.

Columns 3-5 consider alternative benchmarks that use fewer predictors. Column 3 reports the performance of a simple autoregressive model (AR1) in which positive cases in period t are predicted to remain positive cases in $t + 1$. Our prediction models outperform this simple AR1 model as well. Unsurprisingly, the lagged predictand model is particularly poor at predicting rapid increases in the number of conflicts in both Indonesia and Colombia.

Our models include enough time-invariant covariates, however, that in principle the machine learning performance could simply approximate fixed effects. Therefore, column 4 examines a simple OLS fixed effects model. In each year, we regress all previous years' outcomes on fixed effects for each location. Then we take that fixed effect as our prediction for the following year. We see that in all cases, prediction using our full set of covariates outperforms the fixed effects model. The improvements for any violent event are moderate, ranging from 0.025–0.05, which is roughly 5–10% of the difference between a random prediction and perfection. Improvements are larger when predicting hot spots and escalations. The fact that the relative outperformance of our baseline model is greatest in Indonesia and in cases where the dependent variable is rarer is intuitive from the perspective of estimator variance.[17] These fixed effects are noisily estimated, and our prediction algorithms are better able to estimate the relationship between fixed factors and conflict.

[17]The fixed effects model is required to estimate 1,023 parameters in Colombia and 2,009 parameters in Indonesia. As the variation in the dependent variable decreases, which happens as it becomes rarer, this estimation becomes increasingly difficult.

The specification in column 5 attempts to remedy this imprecision by estimating fewer fixed effects at a higher level of aggregation—the department in Colombia and the district in Indonesia.[18] In this case, performance for the rarest event, a 1 standard deviation increase in violence, improves, but still falls well short of our benchmark model.

Together, these results show that the full models perform better than several simpler alternatives. While performance varies across countries and outcomes, it is clear that there are gains to machine learning approaches.

[TABLE 2 HERE]

5.2 We Predict Time-Invariant Risk Over Space Rather than Time

In this section, we clarify the nature and sources of predictability, and conclude that our predictions mainly capture time-invariant risks of violence. Table 3 reports results. For purposes of comparison, column 1 reproduces the baseline EBMA results from column 5 of Table 1.

5.2.1 Violence Histories Alone are a Good Predictor of Future Conflict

First we examine the predictive power of violence histories alone. Importantly, these histories are not simply lagged dependent variables. Instead, they comprise the number and severity of incidents (number of deaths, destruction of property, etc.) and the actors involved, and, in Indonesia, distinguish between each of the ten different violence categories.[19]

Columns 2 and 3 of Table 3 consider models that use all available information about past violence. Column 3 adds measures of population and population density to reflect the fact

[18] There are 33 departments in our Colombia sample and 168 districts in our Indonesia sample. Therefore, the number of parameters drops considerably, as does the imprecision in their estimation.

[19] See Appendix C.3 for the full list of past violence predictors in each country.

that more populous places mechanically have more people who can engage in conflict with one another.

We find that prediction models using violence histories alone perform almost as well as our full model in column 1, suggesting that additional covariates yield, at most, modest improvements beyond these rich conflict histories. The addition of population in column 3 leads to little improvement. The model with violence histories also vastly outperforms a simple lagged dependent variable alone (column 3 of Table 2). Thus, performance is not simply driven by the autoregressive properties of conflict, but by the rich set of conflict measurements provided by our data sources. Appendix A.4 provides further evidence on the predictive returns to more detailed violence data.

5.2.2 Time-Invariant Predictors are Most Effective in Our Models

If detailed, past violence predicts future violence, do we need other predictors at all? We develop a number of tests for parsing the sources of predictability. Column 4 of Table 3 shows results for a model that only uses predictors that do not directly measure past violence. These include the hundreds of socioeconomic and demographic measures discussed above. Performance is comparable to the full model in column 1 and the violence-only model in column 2. This suggests that these socioeconomic and demographic variables contain more or less the same information as the detailed histories of violence, but add little value over them.

Of course, our models contain hundreds of variables, and it is possible that some contribute much less than others. In particular, our models include a number of predictors that change slowly or not at all. Some, such as topographical traits or colonial history, do not vary by definition. Others, such as ethnic and religious traits in Indonesia, do not vary over our sample because they are measured only once. Variables that do not change over our sample cannot, by their nature, predict the timing of violent conflict on their own. If remote areas are at risk for conflict in one year, then they continue to be at risk in the following

years, because they continue to be remote.

To examine the relative performance of time-varying and time-invariant predictors, we compare models composed entirely of one or the other. Column 5 uses only time-invariant traits to predict violence, and performance roughly matches or outperforms the model in column 4. Column 6 uses only time-varying predictors and performance diminishes.[20] Thus, most of our model's performance can be achieved by successfully predicting time-invariant (or at least highly persistent) violence risk.

[TABLE 3 HERE]

In Figure 1, we go one step further and examine the predictive performance of clusters of related predictors. We start with a baseline model that uses only population (level, growth rate, and density) to generate predictions. We then add subgroups of predictors to that baseline model and estimate the change in predictive performance. This approach estimates the predictive power of sets of predictors beyond their association with population.

Figure 1 plots the change in model performance from adding each subgroup. Each out-of-sample year is indicated with a small dot, to give a sense of the range of performance changes. The first out-of-sample year is marked with a triangle, and the last is marked with a square. While the first out-of-sample year generally has worse performance because the training sample is smallest, it is difficult to see a clear improvement in performance over successive out-of-sample predictions. The final year of data is not necessarily the best performing. In the discussion below, we focus on the average change—reflected by the larger open circle.

Consistent with the results above, time-invariant and slow-moving predictors appear to

[20] This is true even for the ≥ 1 SD increase outcome. While such increases naturally have a temporal element to them, it appears that our performance comes from leveraging *which* places are most at risk of experiencing escalations as opposed to *when* these escalations occur.

add the most to predictive performance.[21] This is generally the case even when looking at conflict escalation. The time-invariant predictors that add the most to predictive performance are also notably similar across countries and outcomes. Measures of remoteness, like distance from major cities and road access, and geographic traits like terrain ruggedness, are generally the best predictors. Time-invariant measures of economic structure such as sectoral shares, agricultural features, or mineral presence are also important predictors.

In contrast, the covariates with the most year-to-year variation seldom improve predictive power, even when predicting escalations. This is particularly evident for natural disasters, commodity price shocks, drug price shocks, and weather shocks. In some instances, adding these predictors seems to decrease performance. However, the direction in which they affect performance varies, and the range of performance contributions over out-of-sample years is large, which limits our ability to definitively conclude how these predictors affect model performance.[22]

---

[21] We classify variables as "slow-moving" in Indonesia if they vary triennially, originating from the *Podes* survey. These predictors measure characteristics that tend to vary more between locations than over time within location (see the summary statistics in Appendix C.3, which show the within over between variance). Indeed, a prior draft used a purely time-invariant version of these predictor groups from just the 1999 *Podes*, and their predictive performance was extremely similar (Bazzi et al., 2019).

[22] The inconclusive performance of these variables in our year-ahead prediction models stands in contrast to causal studies of conflict, where variables such as commodity prices and weather shocks have robust significant effects on conflict intensity and sometimes conflict onset. See, for example, Miguel et al. (2004), Bazzi and Blattman (2014), Berman and Couttenier (2013), Berman et al. (forthcoming), Burke et al. (2015) and Dube and Vargas (2013). This underscores the observation that the relationship between causation and prediction is complex (Shmueli, 2010), and that the objective of prediction differs fundamentally from the objective of parameter estimation (Mullainathan and Spiess, 2017).

Some of the time-varying variables do improve performance. Output and communication variables also seem to generally help in Indonesia. However, it is worth noting that even for these time-varying measures, it could still be the case that the majority of their variation is cross-sectional. Indeed, the within over between variance for many of the variables in these predictor groups is below 1, indicating that much of the variation lies in the cross-section rather than within-location, over-time.[23] Electoral data generally appears to improve performance in both countries, though the effect varies for Colombia. Note, moreover, that there is not a clear increase in the marginal contribution of these predictors over time. The first out-of-sample year is seldom the worst in terms of the marginal contribution of these time-varying predictors. Likewise, the final year, which uses the largest training set, is seldom the best. Overall, we conclude that time-invariant predictors contribute more consistently to our model's performance in predicting conflict over time.

[FIGURE 1 HERE]

5.2.3 Our Models Predict Violence across Locations

So far, several pieces of evidence point to the difficulty of predicting the specific timing of violence: the relatively poor performance of time-varying predictors; the interchangeability of histories of violence and time-invariant predictors; and our finding that histories of violence predict spikes in violence roughly as well as levels of violence.

In Appendix A.3, we further show that our models perform especially poorly when predicting *within-location*, over-time variation in violence. In this exercise, we attempt to predict deviations in the number of violent incidents in each location from its historical mean. Performance is very poor. These results further underscore the difficulty of predicting within-unit changes in violence, given the available data.

[23] In Appendix A.7, we consider alternative groupings of time-varying predictors that shed light on this distinction between the spatial and temporal variation in time-varying predictors.

Next, we take the opposite approach, forecasting conflict exclusively across locations. To do this, we randomly split subdistricts in Indonesia and municipalities in Colombia into two, equal-sized groups. We pool observations over time, and train our algorithms using all location–years of data in one group of locations, generating predictions for a second group of location–years. Table 4 reports these results. Strikingly, Column 1 shows that overall performance when forecasting across locations is similar to performance when predicting ahead in time.

[TABLE 4 HERE]

When forecasting across locations, some differences emerge in the performance of individual predictor groups. Figure 2 examines which groups of predictors (along with population) best predict out-of-sample violence across locations. Time-invariant predictors remain important. However, time-varying predictors including weather, natural disasters and commodity prices no longer reduce predictive power, and in some cases, improve it substantially.

One notable difference between the across-location and year-ahead predictions is that the training set uses all years of available data in the across-location approach. The algorithms therefore observe the entire relevant distribution of weather, disasters and commodity price fluctuations over the duration of the period. These variables may behave very differently year to year. When we predict violence one year ahead, if the training period includes such shocks while the testing period does not, the lack of common support across these periods may inhibit the predictive power of these variables. Thus, the short time series of the training and testing samples, and the difficulty of generating off-support predictions, may explain why time-varying covariates like weather shocks perform worse in our predictions over time.

[FIGURE 2 HERE]

6 Discussion

Using an unusually long and wide array of annual data, we show that local violence in Indonesia and Colombia appears to be predictable with relatively high levels of accuracy. But that predictability is largely a function of time-invariant, location-specific risk. This is important in and of itself, since hot spots for violence may pose an especially severe risk of further escalation. Machine learning approaches can help to identify these hot spots that would have remained more obscure with simpler forecasting methods. However, the residual variation—year-to-year changes in violence—remains difficult to forecast.

There are several possible explanations for this latter result. For one, it is possible that the time-varying dimensions of violence are simply idiosyncratic and therefore hard to predict. In many cases, conflict is not only inefficient but is an out-of-equilibrium behavior (Fearon, 1995). These deviations from normal, peaceful social competition could be inherently difficult to forecast. Violence may also be hard to predict because it responds endogenously to the strategic calculations of armed actors. For instance, we may observe peace in a particular region precisely because government security forces crudely predicted a high conflict risk there, and allocated resources accordingly. Likewise, a terrorist may decide to attack an area because that is where the attack was least expected.

A variety of measurement problems may also limit model performance. Human-coded violence data, or data from news reports, is state-of-the-art in that it is often the best or only source of information available, but it is nonetheless prone to misclassification or errors of omission. The timing of violence could be a function of factors that are inherently hard-to-observe and measure, such as social grievances or the deterioration of communal trust. Both issues will bedevil prediction exercises of all kinds.

We might also lack a sufficiently long time series to be able to capture time-varying conflict risk, though several signs suggest otherwise. The limited predictive power of shocks in our over-time predictions may reflect a lack of common support in the training and testing samples. If so, then performance could improve with more years of data. Yet, our results hold

even when the training sample is at its longest, and as the training sample gets longer one might be more concerned about the possibility of structural breaks in the violence-generating process. If models need more than a quarter century of rich conflict and risk factor data to yield better predictions, then the practical prospects for early warning are surely limited.

High-frequency data on local conditions and leading indicators of violence are other potential avenues for improvement. For now, few developing country contexts offer such data. But possibilities in the near future include data from social media, mobile phone meta-data, real-time incident data, and media monitoring. We view these as promising avenues for future research seeking to forecast where violence changes over time.

References

Acemoglu, Daron, Carlos Garcia-Jimeno, and James A. Robinson, "State capacity and economic development: A network approach," *American Economic Review*, 2015.

_ , **James A. Robinson, and Rafael J. Santos**, "The Monopoly of Violence: Evidence from Colombia," *Journal of the European Economic Association*, 2013, *11* (S1), 5–44. https://economics.mit.edu/files/10402.

_ , **Leopoldo Fergusson, James A. Robinson, Dario Romero, and Juan F. Vargas**, "The Perils of High-Powered Incentives: Evidence from Colombia's False Positives," Working Paper 22617, National Bureau of Economic Research September 2016. http://dx.doi.org/10.3386/w22617.

Alesina, Alberto and Ekaterina Zhuravskaya, "Segregation and the Quality of Government in a Cross Section of Countries," *American Economic Review*, 2011, *101* (5), 1872–1911.

Angrist, Joshua D. and Adriana D. Kugler, "Rural Windfall or a New Resource Curse? Coca, Income, and Civil Conflict in Colombia," *The Review of Economics and Statistics*, May 2008, *90* (2), 191–215. https://ideas.repec.org/a/tpr/restat/v90y2008i2p191-215.html.

Barron, Patrick, Kai Kaiser, and Menno Pradhan, "Understanding variations in local conflict: Evidence and implications from Indonesia," *World Development*, 2009, *37* (3), 698–713. https://doi.org/10.1016/j.worlddev.2008.08.007.

_ , **Sana Jaffrey, and Ashutosh Varshney**, "How Large Conflicts Subside: Evidence From Indonesia," *Indonesia Social Development Paper*, 2014, (18). https://asiafoundation.org/resources/pdfs/HowLargeConflictsSubside.pdf.

_ , _ , **and** _ , "When Large Conflicts Subside: The Ebbs and Flows of Violence in Post-Suharto Indonesia," *Journal of East Asian Studies*, 2016, *16* (2), 191–217. https://doi.

org/10.1017/jea.2016.6.

Bazzi, Samuel and Christopher Blattman, "Economic shocks and conflict: Evidence from commodity prices," *American Economic Journal: Macroeconomics*, 2014, *6* (4), 1–38. dx.doi.org/10.1257/mac.6.4.1.

_ **and Matthew Gudgeon**, "The Political Boundaries of Ethnic Divisions," *American Economic Journal: Applied Economics*, forthcoming.

_ **, Robert Blair, Christopher Blattman, Oeindrila Dube, Matthew Gudgeon, and Richard M. Peck**, "The Promise and Pitfalls of Conflict Prediction: Evidence from Colombia and Indonesia," Technical Report, National Bureau of Economic Research 2019.

Beck, Nathaniel, Gary King, and Langche Zeng, "Improving Quantitative Studies of International Conflict: A Conjecture," *The American Political Science Review*, March 2000, *94* (1), 21–35. http://dx.doi.org/10.2307/2586378.

Beger, Andreas, Cassy L. Dorff, and Michael D. Ward, "Irregular leadership changes in 2014: Forecasts using ensemble, split-population duration models," *International Journal of Forecasting*, January 2016, *32* (1), 98–111. dx.doi.org/10.1016/j.ijforecast.2015.01.009.

Berger, Daniel, Shankar Kalyanaraman, and Sera Linardi, "Violence and Cell Phone Communication: Behavior and Prediction in Cote D'Ivoire," *Unpublished Manuscript*, 2014. https://ssrn.com/abstract=2526336.

Berman, Nicolas and Mathieu Couttenier, "External shocks, internal shots: the geography of civil conflicts," *Review of Economics and Statistics*, 2013, (0). https://doi.org/10.1162/REST_a_00521.

_ **,** _ **, Dominic Rohner, and Mathias Thoenig**, "This Mine is Mine! How Minerals Fuel Conflicts in Africa," *American Economic Review*, forthcoming. dx.doi.org/10.1257/aer.20150774.

Blair, Robert A. and Nicholas Sambanis, "Forecasting Civil Wars: Theory and Struc-

ture in an Age of "Big Data" and Machine Learning," *Journal of Conflict Resolution*, 2020.

_ , **Christopher Blattman, and Alexandra Hartman**, "Predicting local violence: Evidence from a panel survey in Liberia," *Journal of Peace Research*, March 2017, *54* (2), 298–312. http://dx.doi.org/10.1177/0022343316684009.

Blattman, Christopher and Edward Miguel, "Civil war," *Journal of Economic Literature*, 2010, pp. 3–57. dx.doi.org/10.1257/jel.48.1.3.

Brandt, Patrick T., John R. Freeman, and Philip A. Schrodt, "Real Time, Time Series Forecasting of Inter- and Intra-State Political Conflict," *Conflict Management and Peace Science*, 2011, *28* (1), 41–64. https://doi.org/10.1177/0738894210388125.

Burke, Marshall, Solomon Hsiang, and Edward Miguel, "Climate and Conflict," *Annual Review Economics*, 2015, *7*, 577–617. https://doi.org/10.1146/annurev-economics-080614-115430.

Cederman, Lars-Erik and Nils B. Weidmann, "Predicting armed conflict: Time to adjust our expectations?," *Science*, 2017, *355* (6324), 474–476. http://dx.doi.org/10.1126/science.aal4483.

Celiku, Bledi and Aart Kraay, "Predicting conflict," *World Bank Policy Research Working Paper 8075*, 2017. https://openknowledge.worldbank.org/handle/10986/26847.

Chacon, Mario, "In the Line of Fire: Political Violence and Decentralization in Colombia," *Working Paper*, 2014. https://dx.doi.org/10.2139/ssrn.23866677.

Colaresi, Michael, Håvard Hegre, and Jonas Nordkvelle, "Early ViEWS: A prototype disaggregated, open-source Violence Early-Warning System," *Presented to the American Political Science Association annual convention, Philadelphia*, 2016. http://www.pcr.uu.se/digitalAssets/653/c_653796-l_1-k_earlyviewsapsa2016.pdf.

Dube, Oeindrila and Juan F. Vargas, "Commodity price shocks and civil conflict: Evidence from Colombia," *The Review of Economic Studies*, 2013, *80* (4), 1384–1421.

https://doi.org/10.1093/restud/rdt009.

_ **and Suresh Naidu**, "Bases, Bullets, and Ballots: The Effect of US Military Aid on Political Conflict in Colombia," *The Journal of Politics*, 2015, *77* (1), 249–267. https://www.journals.uchicago.edu/doi/10.1086/679021.

Fearon, James D., "Rationalist explanations for war," *International organization*, 1995, *49* (3), 379–414. https://www.jstor.org/stable/2706903.

Fischer, Gunther, Freddy Nachtergaele, Sylvia Prieler, Harrij van Velthuizen, Luc Verelst, and David Wiberg, "Global Agro-ecological Zones Assessment for Agriculture (GAEZ 2008)," Technical Report, GAEZ 2008. http://www.fao.org/nr/gaez/publications/en/.

Freund, Yoav and Robert E. Schapire, "A Short Introduction to Boosting," in "In Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence" Morgan Kaufmann 1999, pp. 1401–1406. https://dl.acm.org/citation.cfm?id=1624417.

Gartzke, Erik, "War is in the Error Term," *International Organization*, 1999, *53* (3), 567–587. https://doi.org/10.1162/002081899550995.

Gleditsch, Kristian Skrede and Michael D. Ward, "Forecasting is difficult, especially about the future: Using contentious issues to forecast interstate disputes," *Journal of Peace Research*, 2013, *50* (1), 17–31. https://doi.org/10.1177/0022343312449033.

Goldstone, Jack A., Robert H. Bates, David L. Epstein, Ted Robert Gurr, Michael B. Lustik, Monty G Marshall, Jay Ulfelder, and Mark Woodward, "A Global Model for Forecasting Political Instability," *American Journal of Political Science*, January 2010, *54* (1), 190–208. https://doi.org/10.1111/j.1540-5907.2009.00426.x.

Guha, Rishab. and Serena. Ng, "A Machine Learning Analysis of Seasonal and Cyclical Sales in Weekly Scanner Data," Technical Report, National Bureau of Economic Research 2019.

Gurr, Ted Robert and Mark Lichbach, "Forecasting Internal Conflict," *Comparative*

Political Studies, 1986, *19* (1), 3–38. https://doi.org/10.1177/0010414086019001001.

Harff, Barbara, "No Lessons Learned from the Holocaust? Assessing Risks of Genocide and Political Mass Murder Since 1955," *American Political Science Review*, 2003, *97* (1), 57–73. https://www.jstor.org/stable/3118221.

Hastie, Trevor, Robert Tibshirani, and Jerome Friedman, *The Elements of Statistical Learning* Springer Series in Statistics, New York, NY, USA: Springer New York Inc., 2001.

Hegre, Håvard, Joakim Karlsen, Håvard Mokleiv Nygård, Håvard Strand, and Henrik Urdal, "Predicting Armed Conflict, 2010–2050," *International Studies Quarterly*, 2013, *57* (2), 250–270. https://doi.org/10.1111/isqu.12007.

Hegre, Håvard, Halvard Buhaug, Katherine V Calvin, Jonas Nordkvelle, Stephanie T Waldhoff, and Elisabeth Gilmore, "Forecasting civil conflict along the shared socioeconomic pathways," *Environmental Research Letters*, 2016, *11* (5), 054002. dx.doi.org/10.1088/1748-9326/11/5/054002.

Henderson, J. Vernon, Adam Storeygard, and David N. Weil, "Measuring Economic Growth from Outer Space," *American Economic Review*, 2012, *102* (2), 994–1028. dx.doi.org/10.1257/aer.102.2.994.

Historical Memory Group, *"Enough Already!" Colombia: Memories of War and Dignity*, The National Center for Historical Memory, 2013. http://www.centrodememoriahistorica.gov.co/micrositios/informeGeneral/descargas.html.

Hsiang, Solomon M, Marshall Burke, and Edward Miguel, "Quantifying the influence of climate on human conflict," *Science*, 2013, *341* (6151), 1235367.

Indra, Suahasil Nazara, Djoni Hartono, and Sudarno Sumarto, "Roles of Income Polarization, Income Inequality and Ethnic Fractionalization in Social Conflicts: An Empirical Study of Indonesian Provinces, 2002–2012," *Asian Economic Journal*, 2019, *33* (2), 165–190.

Jasny, Barbara R. and Richard Stone, "Prediction and its Limits," *Science*, 2017, *355* (6324), 468–469. dx.doi.org/10.1126/science.355.6324.468.

Martinez-Bravo, Monica, Priya Mukherjee, and Andreas Stegmann, "The non-democratic roots of elite capture: Evidence from Soeharto mayors in Indonesia," *Econometrica*, 2017, *85* (6), 1991–2010.

McGuirk, Eoin and Marshall Burke, "The Economic Origins of Conflict in Africa," Working Paper 23056, National Bureau of Economic Research January 2017.

Miguel, Edward, Shanker Satyanath, and Ernest Sergenti, "Economic shocks and civil conflict: An instrumental variables approach," *Journal of Political Economy*, 2004, *112* (4), 725–753. https://www.journals.uchicago.edu/doi/10.1086/421174.

Montgomery, Jacob, Florian Hollenbach, and Michael Ward, "Improving Predictions using Ensemble Bayesian Model Averaging," *Political Analysis*, 2012, *20*, 271–291. https://doi.org/10.1093/pan/mps002.

Muchlinski, David, David Siroky, Jingrui He, and Matthew Kocher, "Comparing Random Forest with Logistic Regression for Predicting Class-Imbalanced Civil War Onset Data," *Political Analysis*, 2016, *24* (1), 87–103.

Mueller, Hannes and Christopher Rauh, "Reading Between the Lines: Prediction of Political Violence Using Newspaper Text," *American Political Science Review*, 2017, pp. 1–18. dx.doi.org/10.1017/S0003055417000570.

Mullainathan, Sendhil and Jann Spiess, "Machine Learning: An Applied Econometric Approach," *Journal of Economic Perspectives*, May 2017, *31* (2), 87–106. http://dx.doi.org/10.1257/jep.31.2.87.

Perry, Chris, "Machine learning and conflict prediction: a use case," *Stability: International Journal of Security and Development*, 2013, *2* (3), 56. http://doi.org/10.5334/sta.cr.

Pierskalla, Jan H. and Audrey Sacks, "Unpacking the Effect of Decentralized Gov-

ernance on Routine Violence: Lessons from Indonesia," *World Development*, 2017, *90*, 213–228. https://doi.org/10.1016/j.worlddev.2016.09.008.

Restrepo, Jorge, Michael Spagat, and Juan Vargas, "The Dynamics of the Columbian Civil Conflict: A New Dataset," *Homo Oeconomicus*, 2004, *21*, 396–429. https://ssrn.com/abstract=480247.

Sappington, J. Mark, Kathleen M. Longshore, and Daniel B. Thompson, "Quantifying Landscape Ruggedness for Animal Habitat Analysis: A Case Study Using Bighorn Sheep in the Mojave Desert," *The Journal of Wildlife Management*, 2007, *71* (5), 1419–1426. http://dx.doi.org/10.2193/2005-723.

Schrodt, Philip A., "Forecasting Conflict in the Balkans using Hidden Markov Models," in Robert Trappl, ed., *Programming for Peace: Advances in Group Decision and Negotiation, Vol. 2*, Springer, 2006, pp. 161–184. dx.doi.org/10.1214/10-STS330.

Shmueli, Galit, "To explain or to predict?," *Statist. Sci.*, 08 2010, *25* (3), 289–310. dx.doi.org/10.1214/10-STS330.

Srivastava, Nitish, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov, "Dropout: A Simple Way to Prevent Neural Networks from Overfitting," *Journal of Machine Learning Research*, 2014, *15* (56), 1929–1958.

Sundberg, R. and E. Melander, "Introducing the UCDP georeferenced event dataset," *Journal of Peace Research*, 2013, *50* (4), 523–532.

Tadjoeddin, Zulfan, *Explaining collective violence in contemporary Indonesia: from conflict to cooperation*, Springer, 2014.

Tibshirani, Robert, "Regression Shrinkage and Selection Via the Lasso," *Journal of the Royal Statistical Society, Series B*, 1994, *58*, 267–288. https://www.jstor.org/stable/2346178.

Ward, Michael, Nils Metternich, Cassy Dorff, Max Gallop, Florian Hollenbach, Anna Schultz, and Simon Weschle, "Learning from the Past and Stepping into the

Future: Toward a New Generation of Conflict Prediction," *International Studies Review*, 2013, *15* (4), 473–490. https://doi.org/10.1111/misr.12072.

Weidmann, Nils and Michael Ward, "Predicting Conflict in Space and Time," *Journal of Conflict Resolution*, 2010, *54* (6), 883–901. https://doi.org/10.1177/0022002710371669.

Witmer, Frank D.W., Andrew M Linke, John O'Loughlin, Andrew Gettelman, and Arlene Laing, "Subnational violent conflict forecasts for sub-Saharan Africa, 2015–65, using climate-sensitive models," *Journal of Peace Research*, 2017, *54* (2), 175–192. http://dx.doi.org/10.1177/0022343316682064.

Wright, Austin L. and Patrick Signoret, "Climate Shocks, Price Dynamics, and Human Conflict," *Working paper*, 2016. https://www.austinlwright.com/climate-shocks/.

Table 1: Out-of-Sample (One Year Ahead) Performance of Prediction Models, Area Under the Curve (AUC)

	LASSO (1)	Random Forest (2)	Adaptive Boosting (3)	Neural Network (4)	EBMA (5)
	Indonesia (social conflict, 2008-2014)				
Any violent event	0.819	0.818	0.823	0.792	0.823
$\geq$ 5 violent events	0.940	0.935	0.942	0.910	0.941
$\geq$ 1 s.d. increase in events	0.866	0.817	0.852	0.825	0.860
	Colombia (attacks and clashes, 1998-2005)				
Any violent event	0.845	0.847	0.849	0.825	0.850
$\geq$ 5 violent events	0.914	0.911	0.910	0.886	0.915
$\geq$ 1 s.d. increase in events	0.802	0.787	0.796	0.741	0.801

Notes: Each model is trained on all available data preceding the out-of-sample prediction year. Training data starts with 1991 data used to predict 1992 violence in Colombia and 2003 data for 2004 violence in Indonesia. Out-of-sample prediction begins in 1998 in Colombia and 2008 in Indonesia. The AUC is the area under the ROC curve, a measure of the trade-off between the true positive rate and false positive rate at different thresholds. We report average performance over the out-of-sample years.

Table 2: Out-of-Sample Performance Versus Benchmarks

	Baseline EBMA (1)	OLS (2)	Lagged Predictand (3)	Location FE Only (4)	Department/ District FE Only (5)
	Indonesia (social conflict, 2008-2014)				
Any violent event	0.823	0.790	0.687	0.774	0.752
$\geq$ 5 violent events	0.941	0.911	0.808	0.880	0.871
$\geq$ 1 s.d. increase in events	0.860	0.759	0.527	0.694	0.776
	Colombia (attacks and clashes, 1998-2005)				
Any violent event	0.850	0.821	0.743	0.828	0.721
$\geq$ 5 violent events	0.915	0.858	0.748	0.849	0.742
$\geq$ 1 s.d. increase in events	0.801	0.744	0.521	0.683	0.712

Notes: Each model is trained on all data available preceding the out-of-sample prediction year. Training data starts with 1991 data used to predict 1992 violence in Colombia and 2003 data for 2004 violence in Indonesia. Out-of-sample prediction begins in 1998 in Colombia and 2008 in Indonesia. The AUC is the area under the ROC curve, a measure of the trade-off between the true positive rate and false positive rate as we vary the discrimination threshold. We report average performance over the out-of-sample years above. Lagged predictand includes only a single lag of the violence indicator of interest. Fixed effects models are estimated by OLS.

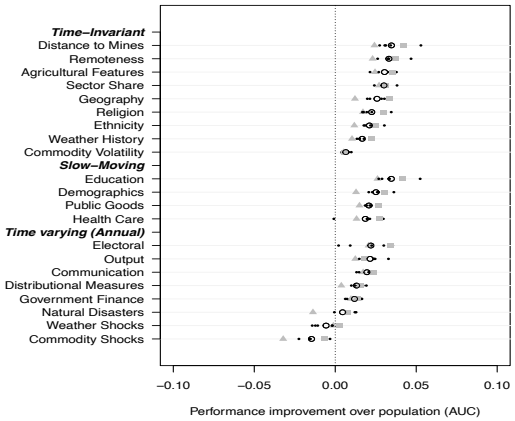
Table 3: Out-of-Sample (One Year Ahead) Performance of the Ensemble (EBMA) Method, Varying Predictor Sets

	Full Predictors (1)	All Past Violence Measures (2)	All Past Violence & Population (3)	Full Excl. Past Violence (4)	Time-Invariant Predictors (5)	Time-Varying Predictors (6)
	Indonesia (social conflict, 2008-2014)					
Any violent event	0.823	0.805	0.815	0.810	0.817	0.789
$\geq$ 5 violent events	0.941	0.939	0.942	0.922	0.931	0.902
$\geq$ 1 s.d. increase in events	0.860	0.845	0.856	0.847	0.856	0.815
	Colombia (attacks and clashes, 1998-2005)					
Any violent event	0.850	0.812	0.838	0.828	0.832	0.763
$\geq$ 5 violent events	0.915	0.905	0.912	0.878	0.880	0.808
$\geq$ 1 s.d. increase in events	0.801	0.765	0.788	0.780	0.781	0.748

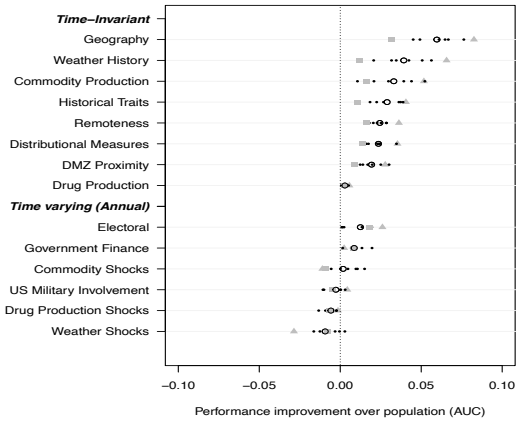
Notes: Each model is trained on all available data preceding the out-of-sample prediction year. Training data starts with 1991 data used to predict 1992 violence in Colombia and 2003 data for 2004 violence in Indonesia. Out-of-sample prediction begins in 1998 in Colombia and 2008 in Indonesia. The AUC is the area under the ROC curve, a measure of the trade-off between the true positive rate and false positive rate at different thresholds. We report average performance over the out-of-sample years above. Past violence measures include breakdowns of events by actors and outcomes such as deaths and damages. Population includes population growth rates and density.

Figure 1: AUC Improvements from Individual Predictor Groups

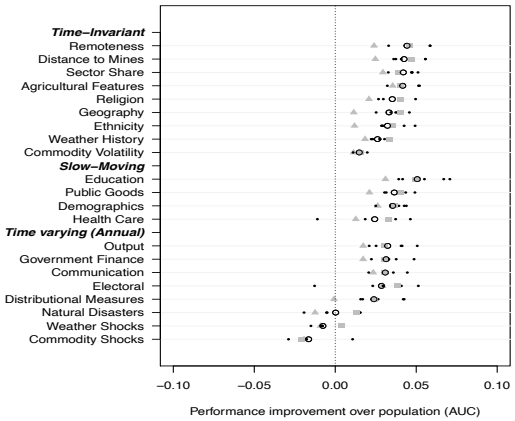
(a) Any violent event (Indonesia)



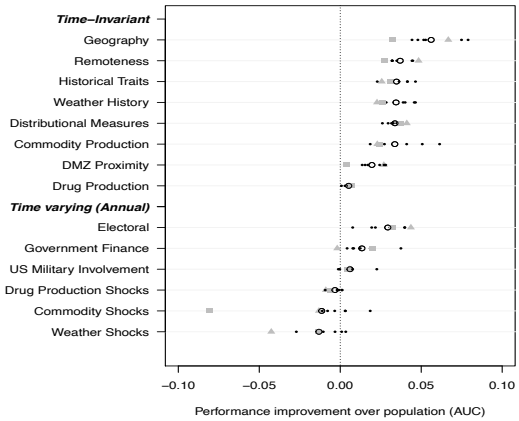
(b) Any violent event (Colombia)



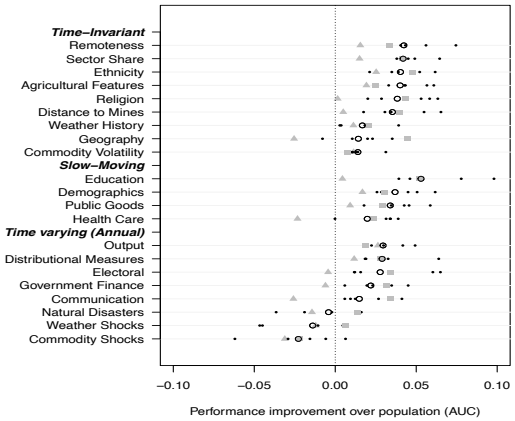
(c) ≥ 5 violent events (Indonesia)



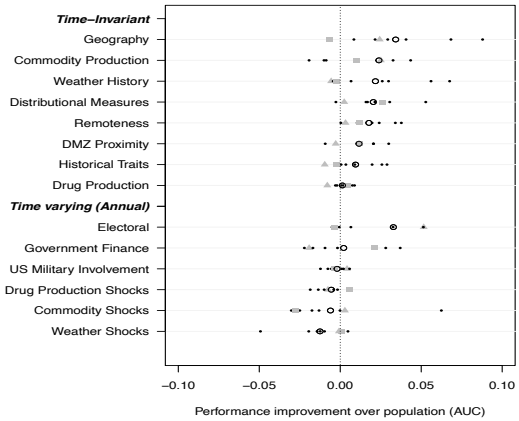
(d) ≥ 5 violent events (Colombia)



(e) ≥ 1 S.D. increase in violent events (Indonesia)



(f) ≥ 1 S.D. increase in violent events (Colombia)



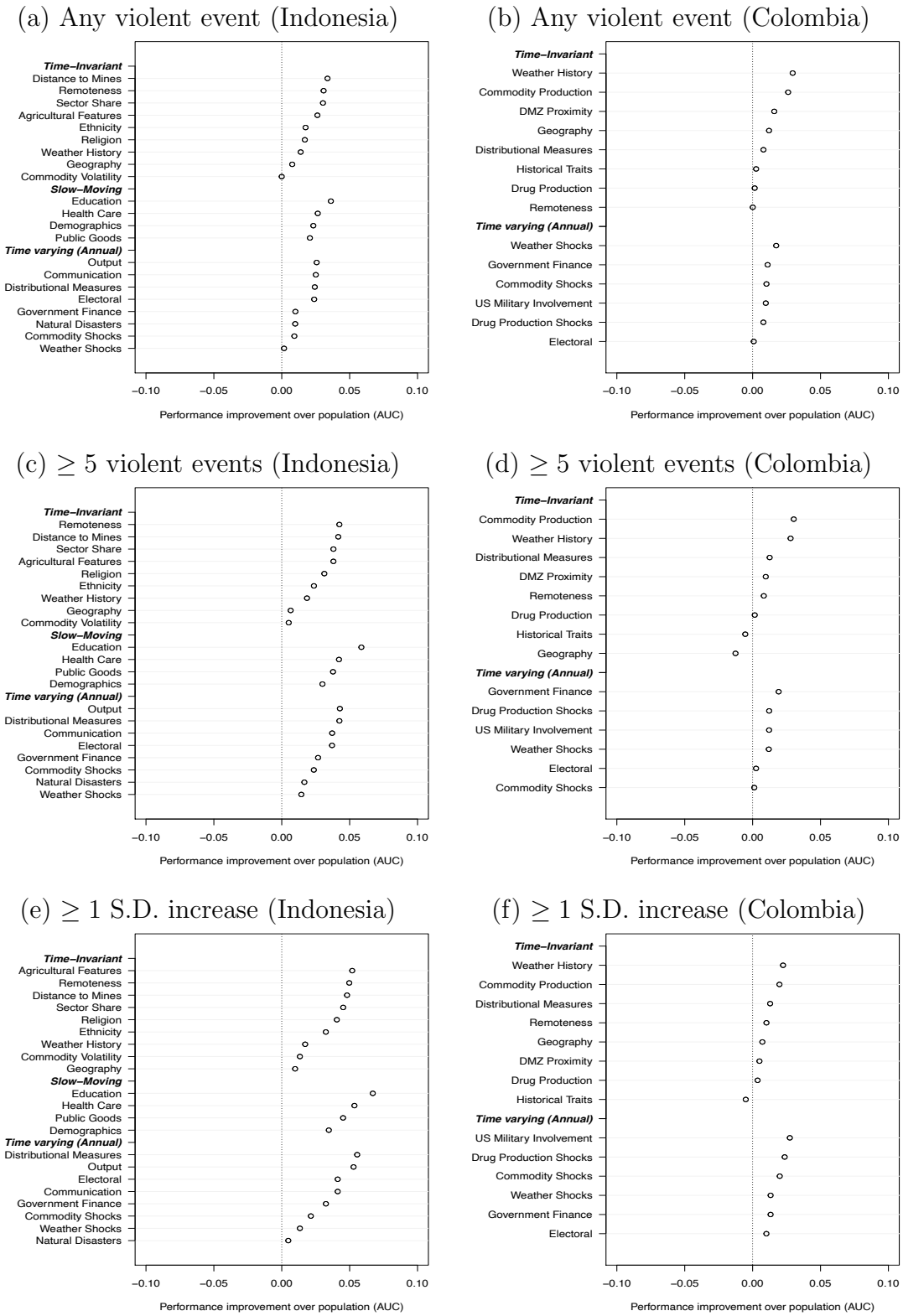
Notes: Dots represent performance in individual years. A triangle (square) denotes the first (last) year of the sample. The hollow circle is average performance across the years. Appendix C.3 lists the variables in each predictor group.

Table 4: Predicting Across Locations

	Full Predictors (1)	All Past Violence Measures (2)	All Past Violence & Population (3)	Full Excl. Past Violence (4)
		Indonesia		
Any violent event	0.827	0.807	0.817	0.810
$\geq$ 5 violent events	0.941	0.935	0.938	0.915
$\geq$ 1 s.d. increase in events	0.864	0.844	0.850	0.855
		Colombia		
Any violent event	0.840	0.795	0.823	0.798
$\geq$ 5 violent events	0.931	0.919	0.926	0.862
$\geq$ 1 s.d. increase in events	0.834	0.786	0.812	0.796

Notes: AUCs for a random test set of locations over time. Algorithms are trained using data from training locations over the entire time span of the datasets. Training data starts with 1991 data used to predict 1992 violence in Colombia and 2003 data for 2004 violence in Indonesia. The AUC is the area under the ROC curve, a measure of the trade-off between the true positive rate and false positive rate as we vary the discrimination threshold.

Figure 2: AUC Improvements from Individual Predictor Groups, Cross-Sectional Prediction



Notes: Performance (AUC) in the single test sample is reported.

Appendix for Online Publication

A Additional Results **1**

A.1 Other Measures of Model Performance . . . 1

A.2 Predicting Event Counts . . . 6

A.3 Predicting Within-Location Risk . . . 9

A.4 Returns to More Detailed Violence Measurement . . . 11

A.5 Predicting New Violence . . . 13

A.6 Alternative Definitions of Escalation . . . 18

A.7 Alternative Predictor Groupings . . . 21

A.8 Alternative Colombia Conflict Data . . . 24

A.8.1 Comparison of U. Rosario and CERAC Datasets . . . 24

A.8.2 Prediction Performance in U. Rosario Data . . . 27

B Methodological Details **30**

B.1 Hyper-parameter Choices . . . 30

B.2 Alternative Cross-Validation . . . 32

B.3 Alternative Neural Networks . . . 35

C Data Appendix **37**

C.1 Indonesia . . . 37

C.1.1 Administrative Levels . . . 37

C.1.2 Violence . . . 37

C.1.3 Comparison to Other Violence Data . . . 40

C.1.4 Covariates . . . 41

C.2 Colombia . . . 49

C.2.1 Administrative Levels . . . 49

C.2.2 Violence 49
C.2.3 Comparison to Other Violence Data 50
C.2.4 Covariates 51
C.3 Predictor Groupings 56
C.3.1 Indonesia 56
C.3.2 Colombia 71
C.4 Correlating Commodity Price and Weather Shocks to Conflict 80

A Additional Results

A.1 Other Measures of Model Performance

In the paper, we measure performance with the AUC. However, practitioners may be interested in other performance statistics. Here we consider choosing various discrimination thresholds, each of which essentially picks a point on the ROC curve. In Table A.1, we report a number of alternative performance statistics. Each of our three main predictands are reported in separate panels (a), (b), and (c). Note that we do not change the performance metric used to choose hyper-parameters. We merely report alternative performance metrics for the same models that are developed in the main results to maximize AUC.

Maximal Accuracy. In the top portion of each panel, we report performance when we set a discrimination threshold to maximize accuracy. We choose this discrimination threshold in our cross-validation routine. We report accuracy (the proportion of all cases correctly predicted), sensitivity (the proportion of incidents correctly predicted), and specificity (the proportion of non-incidents correctly predicted).[1]

Maximal Sensitivity. In the next section of each panel, we choose a different discrimination threshold to maximize sensitivity. Of course, if we predicted violence everywhere, we would achieve a sensitivity of 1, but that would not be a useful prediction. Instead, we choose to maximize sensitivity subject to the constraint that we keep accuracy above 0.5 in the cross-validation process. We observe that, if needed, these models could identify all of the places that experience violence, but this coverage comes at a high cost in terms of accuracy and specificity (false positives).[2]

[1] Observe that as the outcome becomes rarer, accuracy generally increases while sensitivity plummets. This is mechanical: when an outcome is rare, it is easy to predict that it never happens (an uninformative prediction) and achieve high accuracy.

[2] A practitioner might have preferences between these two extremes and consequently choose a point on the ROC that differs from these two. This is precisely why we use the AUC as our benchmark.

Mean Squared Error. We report the mean squared error (MSE). Generally, the MSE is closely correlated with the AUC, but the correspondence is not exact.[3] Using the MSE does not meaningfully change the results of this paper.

Area Under the Precision-Recall Curve. Finally, we report the area under the Precision-Recall Curve or the PR-AUC. This metric measures the trade-off between precision (share of true positives in the sample of positive predictions) and recall (the true positive rate). In contexts such as ours in which data is imbalanced and violence is relatively rare, high PR-AUC's can be difficult to achieve. If there are many non-events, then misclassifying a small share as positive can have large, deleterious effects on precision. Indeed, we see for the rarest events (the > 1 standard deviation spikes) that the PR-AUC is quite low. However, we also see that holding the predictand fixed, comparisons of model performance based on the area under the Precision-Recall curve are largely consistent with the comparisons based on the area under the ROC curve.

[3]This is because the AUC penalizes errors in relative rankings of different locations, while the MSE penalizes errors of prediction according to the magnitude of the difference between actual outcomes and predicted probabilities. While these two types of errors are related, they are distinct. However, the differences often tend to be small. Assessing model performance according to MSE only changes comparisons for borderline cases, which we do not put much emphasis on.

Table A.1: Out-of-Sample Performance of Prediction Models

	Indonesia (social conflict)					**Colombia** (attacks and clashes)				
	LASSO (1)	Random Forest (2)	Adaptive Boosting (3)	Neural Network (4)	EBMA (5)	LASSO (6)	Random Forest (7)	Adaptive Boosting (8)	Neural Network (9)	EBMA (10)
	(a) Any violent event									
EBMA Weight	0.251	0.249	0.253	0.246		0.252	0.248	0.253	0.247	
	Threshold maximizes accuracy									
Accuracy	0.733	0.734	0.742	0.698	0.737	0.770	0.779	0.780	0.748	0.775
Sensitivity	0.745	0.720	0.719	0.686	0.739	0.553	0.629	0.612	0.543	0.582
Specificity	0.719	0.750	0.770	0.714	0.735	0.907	0.873	0.885	0.867	0.894
	Threshold maximizes sensitivity, while accuracy above 50%									
Accuracy	0.546	0.544	0.547	0.542	0.548	0.576	0.529	0.532	0.544	0.545
Sensitivity	0.999	1.000	0.999	1.000	1.000	0.955	0.977	0.977	0.941	0.971
Specificity	0.009	0.003	0.010	0.000	0.013	0.341	0.253	0.256	0.297	0.280
MSE (Brier Score)	0.174	0.174	0.172	0.193	0.173	0.157	0.154	0.151	0.175	0.153
P-R AUC	0.856	0.856	0.860	0.826	0.860	0.795	0.796	0.796	0.770	0.800
AUC	0.819	0.818	0.823	0.792	0.823	0.845	0.847	0.849	0.825	0.850
Dep. Var. Mean	0.542					0.383				
	(b) $\geq$ 5 violent events									
EBMA Weight	0.250	0.249	0.252	0.250		0.250	0.249	0.251	0.250	
	Threshold maximizes accuracy									
Accuracy	0.919	0.929	0.930	0.908	0.928	0.920	0.923	0.922	0.915	0.924
Sensitivity	0.535	0.593	0.601	0.565	0.607	0.342	0.374	0.399	0.315	0.387
Specificity	0.974	0.978	0.978	0.957	0.975	0.984	0.984	0.981	0.982	0.984
	Threshold maximizes sensitivity, while accuracy above 50%									
Accuracy	0.544	0.321	0.465	0.470	0.395	0.521	0.509	0.507	0.284	0.492
Sensitivity	0.959	0.993	0.991	0.969	0.994	0.970	0.982	0.970	0.989	0.982
Specificity	0.483	0.223	0.388	0.396	0.307	0.470	0.457	0.455	0.205	0.437
MSE (Brier Score)	0.062	0.055	0.054	0.073	0.055	0.059	0.058	0.058	0.066	0.057
P-R AUC	0.765	0.773	0.784	0.664	0.782	0.625	0.612	0.604	0.543	0.621
AUC	0.940	0.935	0.942	0.910	0.941	0.914	0.911	0.910	0.886	0.915
Dep. Var. Mean	0.127					0.099				
	(c) $\geq$ 1 s.d. increase in events									
EBMA Weight	0.250	0.250	0.250	0.250		0.250	0.250	0.250	0.250	
	Threshold maximizes accuracy									
Accuracy	0.964	0.964	0.964	0.964	0.964	0.947	0.948	0.947	0.948	0.948
Sensitivity	0.008	0.004	0.018	0.000	0.016	0.012	0.005	0.014	0.000	0.011
Specificity	0.999	0.999	0.999	1.000	0.999	0.998	0.999	0.998	1.000	0.999
	Threshold maximizes sensitivity, while accuracy above 50%									
Accuracy	0.451	0.196	0.404	0.441	0.276	0.536	0.412	0.498	0.295	0.461
Sensitivity	0.956	0.977	0.961	0.902	0.987	0.895	0.943	0.905	0.914	0.929
Specificity	0.432	0.167	0.384	0.423	0.250	0.515	0.383	0.475	0.259	0.434
MSE (Brier Score)	0.031	0.034	0.032	0.037	0.031	0.047	0.047	0.047	0.050	0.046
P-R AUC	0.214	0.170	0.207	0.170	0.217	0.198	0.166	0.181	0.146	0.200
AUC	0.866	0.817	0.852	0.825	0.860	0.802	0.787	0.796	0.741	0.801
Dep. Var. Mean	0.035					0.052				

Notes: Each model is trained on all data available preceding the out-of-sample prediction year. Accuracy is the proportion of subdistricts correctly predicted. Sensitivity is the proportion of subdistricts that actually experience violence for which we predicted violence. Specificity is the proportion of subdistricts that do not actually experience violence, where we accurately predict non-violence. The AUC is the area under the ROC curve, a measure of the trade-off between the true positive rate and false positive rate as we vary the discrimination threshold.

We can also train the models to maximize different performance metrics. Table A.2 reports performance for a set of models where hyper-parameters are chosen to minimize mean squared error rather than to maximize AUC. We see that performance is very similar, as these two accuracy measures are closely related.

Table A.2: Out-of-Sample Performance of Prediction Models

	Indonesia (social conflict)					**Colombia** (attacks and clashes)				
	LASSO (1)	Random Forest (2)	Adaptive Boosting (3)	Neural Network (4)	EBMA (5)	LASSO (6)	Random Forest (7)	Adaptive Boosting (8)	Neural Network (9)	EBMA (10)
	(a) Any violent event									
EBMA Weight	0.251	0.249	0.253	0.246		0.252	0.248	0.252	0.247	
	Threshold maximizes accuracy									
Accuracy	0.732	0.734	0.742	0.698	0.738	0.769	0.779	0.779	0.748	0.776
Sensitivity	0.741	0.720	0.716	0.686	0.735	0.554	0.629	0.611	0.543	0.587
Specificity	0.722	0.750	0.773	0.714	0.743	0.904	0.873	0.884	0.867	0.892
	Threshold maximizes sensitivity, while accuracy above 50%									
Accuracy	0.546	0.544	0.547	0.542	0.548	0.577	0.529	0.534	0.544	0.545
Sensitivity	0.999	1.000	0.999	1.000	0.999	0.954	0.977	0.977	0.941	0.972
Specificity	0.010	0.003	0.011	0.000	0.013	0.344	0.253	0.258	0.297	0.281
MSE (Brier Score)	0.174	0.174	0.172	0.193	0.173	0.158	0.154	0.151	0.175	0.153
P-R AUC	0.856	0.856	0.859	0.826	0.860	0.795	0.796	0.796	0.770	0.800
AUC	0.819	0.818	0.823	0.792	0.823	0.845	0.847	0.849	0.825	0.850
Dep. Var. Mean	0.542					0.383				
	(b) $\geq$ 5 violent events									
EBMA Weight	0.250	0.249	0.251	0.250		0.251	0.250	0.250	0.249	
	Threshold maximizes accuracy									
Accuracy	0.920	0.929	0.929	0.909	0.929	0.920	0.923	0.922	0.919	0.923
Sensitivity	0.524	0.593	0.598	0.526	0.597	0.352	0.374	0.375	0.312	0.371
Specificity	0.978	0.978	0.978	0.964	0.977	0.983	0.984	0.983	0.987	0.984
	Threshold maximizes sensitivity, while accuracy above 50%									
Accuracy	0.554	0.321	0.460	0.400	0.387	0.526	0.509	0.529	0.170	0.462
Sensitivity	0.950	0.993	0.990	0.987	0.995	0.970	0.982	0.973	0.991	0.984
Specificity	0.495	0.223	0.383	0.313	0.297	0.476	0.457	0.480	0.078	0.403
MSE (Brier Score)	0.062	0.055	0.054	0.069	0.055	0.059	0.058	0.058	0.063	0.057
P-R AUC	0.768	0.773	0.785	0.677	0.785	0.625	0.612	0.607	0.572	0.625
AUC	0.940	0.935	0.942	0.912	0.942	0.914	0.911	0.911	0.892	0.917
Dep. Var. Mean	0.127					0.099				
	(c) $\geq$ 1 s.d. increase in events									
EBMA Weight	0.250	0.250	0.250	0.250		0.250	0.250	0.250	0.250	
	Threshold maximizes accuracy									
Accuracy	0.964	0.964	0.964	0.964	0.964	0.948	0.948	0.948	0.948	0.949
Sensitivity	0.012	0.004	0.014	0.000	0.010	0.008	0.005	0.019	0.000	0.011
Specificity	0.999	0.999	0.999	1.000	0.999	0.999	0.999	0.998	1.000	1.000
	Threshold maximizes sensitivity, while accuracy above 50%									
Accuracy	0.456	0.196	0.544	0.340	0.276	0.524	0.412	0.515	0.267	0.451
Sensitivity	0.961	0.977	0.915	0.889	0.978	0.899	0.943	0.894	0.944	0.940
Specificity	0.438	0.167	0.530	0.319	0.250	0.502	0.383	0.494	0.228	0.423
MSE (Brier Score)	0.031	0.034	0.031	0.035	0.031	0.046	0.047	0.047	0.050	0.046
P-R AUC	0.215	0.170	0.206	0.134	0.205	0.205	0.166	0.171	0.148	0.192
AUC	0.862	0.817	0.849	0.780	0.847	0.802	0.787	0.789	0.743	0.798
Dep. Var. Mean	0.035					0.052				

Notes: Each model is trained on all data available preceding the out-of-sample prediction year. Hyper-parameters are chosen to minimize MSE. Accuracy is the proportion of subdistricts correctly predicted. Sensitivity is the proportion of subdistricts that actually experience violence for which we predicted violence. Specificity is the proportion of subdistricts that do not actually experience violence, where we accurately predict non-violence. The AUC is the area under the ROC curve, a measure of the trade-off between the true positive rate and false positive rate as we vary the discrimination threshold.

A.2 Predicting Event Counts

We chose to predict indicators instead of counts in our main analyses for several reasons: they are easy to interpret, performance statistics are also easy to interpret, and they are the most common type of outcome predicted by the conflict literature to date. In this section, however, we examine the predictability of event counts.[4]

Table A.3 reports performance of the ensemble for a number of predictor sets similar to those that we consider in the paper. We report two statistics—the mean squared error and a deviance-based R^2. The deviance-based R^2 is similar to a typical R^2 but is adapted to count data.[5]

Column 1 reports overall performance of the full model. Our predictions explain a decent share of the variation in counts in both countries. Column 2 shows that, as in our prediction of indicators, violence histories perform just as well as the full set of predictors. Column 3 adds population measures to column 2 and shows a small increase in performance.

Column 4 considers all of the predictors that do not directly measure violence. These predictors

---

[4]Predicting event counts requires some changes to the algorithms. In the cases of Lasso and Gradient Boosted Machines, we change the loss function to a Poisson loss function to accommodate the count data. In theory, we could do the same for Random Forests and Neural Networks, but it is technically more challenging since the existing R packages do not accommodate count data. Instead, we use a Gaussian loss function. This presents a problem for the Neural Network algorithm, as its output is linear and therefore cannot take values below zero. While this occurs for some location-years, it is not that common, and we deal with this by left censoring the predictions at zero. For Random Forests, this is not an issue because the algorithm can only make predictions in the convex hull of the predictand values in the training set. We aggregate these predictions using an ensemble model, in this case with weights based on a Poisson distribution for the model likelihood.

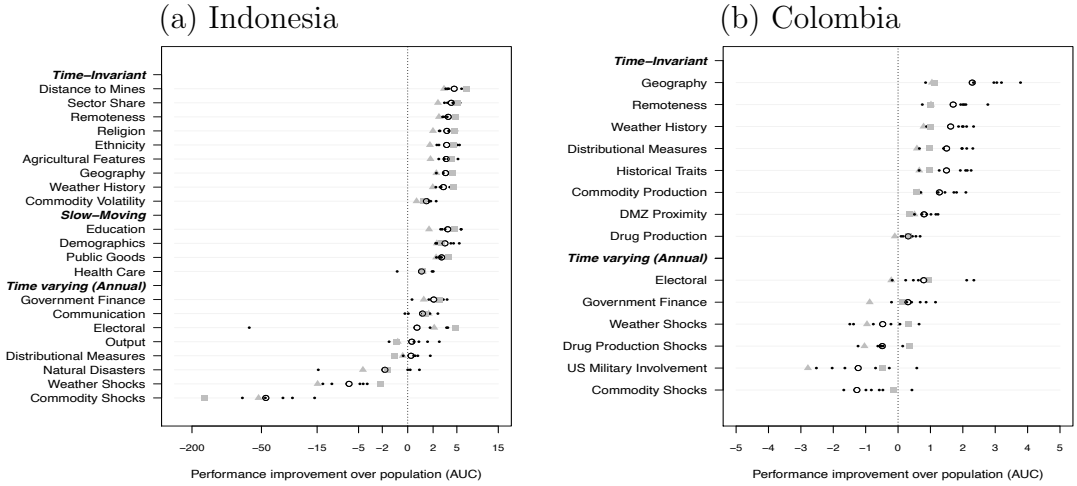
[5]While the typical R^2 measures the share of variance that for which the fitted values account, R^2_{dev} measures the share of deviance that the fitted values explain. Specifically, this measure is $R^2_{dev} = 1 - \frac{\text{Residual Deviance}}{\text{Null Deviance}}$.

perform significantly worse than the full model. This drop in performance is worse than when we are predicting indicators. Columns 5 and 6 further clarify the reason for this. Of the non-violence predictors, the time-varying ones perform more poorly—so much so that their inclusion actually reduces performance relative to time-invariant predictors alone.

This result is in line with what we see in the paper. In the case of predicting indicators, we saw that some of the time-varying predictors actually decreased performance. In this case we see much larger decreases, but that is because the outcome is unbounded and so is the size of possible errors. Therefore, when extrapolating the effect of commodity price shocks, for example, imprecise estimation can lead to large differences between actual and predicted violence in a few cases. And these large errors play an outsized role in contributing to our performance statistics.

Finally, figure A.1 reports a breakdown of specific predictor groups and shows similar patterns to those we see for the prediction of indicators. The deleterious effect of the commodity price movements is particularly severe, where out of sample predictions yield large errors.

Figure A.1: MSE Improvements, Predicting Counts



Notes: Performance in individual years appear as small dots. The first (last) year of the sample is colored gray in order to show the change in performance over time, or lack thereof. The large hollow circle is the average of performance across the years. Note that results for Indonesia are displayed on a symmetric log scale for readability.

Table A.3: Out-of-sample (one year ahead) performance of the ensemble (EBMA) method, varying predictor sets

	Predicting the count of violent events					
	Full Predictors (1)	All Past Violence Measures (2)	All Past Violence & Population (3)	Full Excl. Past Violence (4)	Time-Invariant Predictors (5)	Time-Varying Predictors (6)
	(a) Indonesia					
MSE	11.009	6.958	6.612	21.615	8.219	27.693
R^2_{dev}	0.469	0.666	0.674	0.231	0.659	-0.011
	(b) Colombia					
MSE	7.258	6.356	6.372	8.065	7.437	10.153
R^2_{dev}	0.457	0.519	0.538	0.396	0.484	0.253

Notes: Each model is trained on all data available preceding the out-of-sample prediction year. Training data starts with 1991 data used to predict 1992 violence in Colombia and 2003 data for 2004 violence in Indonesia. Out-of-sample prediction begins in 1998 in Colombia and 2008 in Indonesia. We report average performance over the out-of-sample years above.

A.3 Predicting Within-Location Risk

We construct a new outcome measure that isolates within-location, over-time variation in violence. Specifically, we measure the deviation of the number of incidents in period t from the average number of incidents per year from the start of the panel to period $t-1$. By taking deviations from the historical mean, we remove the average difference in violence across locations, leaving purely within-location variation. Given the continuous outcome, we evaluate performance using the out-of-sample mean squared error (MSE) instead of the AUC.

The results, reported in Table A.4, suggest that our models struggle to predict within-location variation in violence in both Indonesia and Colombia.[6] The first row reports the variance of the dependent variable in each context. In both Colombia and Indonesia, the MSE is only slightly lower than the variance of the dependent variable in the test set, indicating we are able to predict very little of this within-location variation. In both cases, the out-of-sample R^2 close to zero.

[6]These differences arise from the difficulty of predicting within-unit deviations and not changes in the performance metric or the shift to counts. Appendix A.1 shows that our baseline performance is similar using the MSE instead of the AUC, while Appendix A.2 shows that our benchmark model is able to predict a large share of the variation in incident counts.

Table A.4: Predicting Demeaned Number of Violent Events

	Indonesia (1)	Colombia (2)
Var(Dependent Variable)	8.682	7.541
EBMA mean square error (MSE)	8.513	7.428
R^2	0.020	0.015

Notes: Each model is trained on all available data preceding the out-of-sample prediction year. Training data starts with 1991 data used to predict 1992 violence in Colombia and 2003 data for 2004 violence in Indonesia. Out-of-sample prediction begins in 1998 in Colombia and 2008 in Indonesia. We report average mean squared error over the out-of-sample years above.

A.4 Returns to More Detailed Violence Measurement

Our benchmarking exercise found that a model using detailed violence histories performs almost as well as one that includes additional social, economic, and political covariates. This leads us to ask what is the payoff to using richer and more detailed violence data. More granular, accurate, or disaggregated data could improve predictions. For example, a history of small-scale ethnic cleansing could conceivably presage larger scale violence, whereas other kinds of inter-group hostilities might not. If a violence measures conflates these two kinds of violence then its predictive performance will falter. Yet, collecting and coding richer data is costly for policymakers and researchers. Hence, it is useful to explore the returns in terms of predictive performance.

With Indonesia, which has fairly granular violence data, we can conduct this 'experiment' by using more versus less granular violence data, and observe changes in performance. We report results in Table A.5. Column 1 reports the full model with all predictors for comparison. Column 2 reports performance for the lagged dependent variables alone. Column 3 reports performance using measurements of prior aggregate conflict from SNPK, including total number of incidents, total deaths, total injuries, and total property damage. Column 4 further breaks down these incident measures by violence category (e.g., identity violence, resource conflict). And finally, Column 5 includes the lagged total number of killings and indicators for mass unrest reported in *Podes* and the Disaster Information Management System (DIMS).

The AUCs increase with each successive column. Perhaps unsurprisingly, the largest increase comes from the move from columns 2 to 3. The AUCs increase more when disaggregating the subdistrict-level violence categories than when adding additional measures of the same broad episodes of violence as reported in *Podes*. This highlights the potential predictive value of having detailed information on the nature of prior conflict in terms of the key outcomes of contestation.

Table A.5: Out-of-Sample (One Year Ahead) Performance of the Ensemble (EBMA) Method, Varying Data Granularity

	Full Predictors (1)	Only Lagged Indicator (2)	Add Intensity (3)	Disaggregate Violence (4)	All Lagged Violence Data (5)
	Indonesia, 2008-2014				
Any violent event	0.823	0.687	0.785	0.804	0.805
$\geq$ 5 violent events	0.941	0.808	0.934	0.937	0.939
$\geq$ 1 s.d. increase in events	0.860	0.527	0.833	0.846	0.845
Predictor set:					
Lag of dependent variable	Yes	Yes	Yes	Yes	Yes
Total incidents, deaths, injuries	Yes		Yes	Yes	Yes
Disaggregate by violence type	Yes			Yes	Yes
Total village-level killings reported	Yes				Yes
Podes and DIMS	Yes				Yes
Economic and social characteristics	Yes				

Notes: Each model is trained on all data available preceding the out-of-sample prediction year. Training data starts in 2002 in Indonesia. Out-of-sample prediction begins in 2008 in Indonesia. The AUC is the area under the ROC curve, a measure of the trade-off between the true positive rate and false positive rate as we vary the discrimination threshold. We report average performance over the out-of-sample years above.

A.5 Predicting New Violence

In this section, we restrict our attention to locations that are not experiencing violence in the current year, and assess the models' ability to predict violence in the following year. These seem like the cases of violence that would be most important for policymakers to predict. Since our data are events, rather than extended episodes of violence, there is not a natural definition of "onset". We therefore assess predictive performance for previously peaceful locations in two ways.

First, we use our main model forecasts and evaluate prediction performance in the set of location-years that are not currently experiencing conflict. Table A.6 reports performance in these locations without violence. Performance declines significantly, but predictability remains high and correspondingly useful from a policy perspective. Figure A.2 reproduces Figure 1 and shows similar qualitative results. Note that the variability of performance increases across years, since these events are rarer in locations that did not previously experience conflict.

Second, we train new models using only location–years with no current violence. We then evaluate the models using location–years without violence, i.e., the same inclusion criteria for the training sample and the same testing sample as in the first exercise. In Table A.7, we see that, overall, performance declines relative to the prior exercise, which suggests that the models developed with this limited training sample perform worse than our main models. This highlights a bias variance tradeoff: the model that learns from instances with and without violence has access to more training data and consequently lower variance. Figure A.3 again shows similar qualitative results with respect to predictor contributions.

Together, these results point to the difficulty of predicting outbreaks of violence in previously peaceful locations. This is consistent with our findings from other exercises showing the difficulty of forecasting the timing of new conflict events within location.

Table A.6: Performance of Main Prediction Models in Locations without Current Violence, AUC

	LASSO (1)	Random Forest (2)	Adaptive Boosting (3)	Neural Network (4)	EBMA (5)
	Indonesia (social conflict, 2008-2014)				
Any violent event	0.716	0.715	0.718	0.684	0.721
$\geq$ 5 violent events	0.845	0.816	0.816	0.800	0.842
$\geq$ 1 s.d. increase in events	0.809	0.750	0.789	0.772	0.823
	Colombia (attacks and clashes, 1998-2005)				
Any violent event	0.741	0.744	0.752	0.707	0.753
$\geq$ 5 violent events	0.795	0.782	0.770	0.713	0.788
$\geq$ 1 s.d. increase in events	0.791	0.778	0.791	0.703	0.803

Notes: Each model is trained on all available data preceding the out-of-sample prediction year. Training data starts with 1991 data used to predict 1992 violence in Colombia and 2003 data for 2004 violence in Indonesia. Out-of-sample prediction begins in 1998 in Colombia and 2008 in Indonesia. Performance is calculated only in location-years that have not experienced violence in the past year. The AUC is the area under the ROC curve, a measure of the trade-off between the true positive rate and false positive rate at different thresholds. We report average performance over the out-of-sample years.

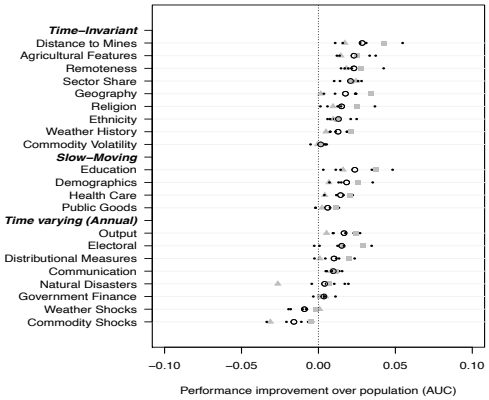
Table A.7: Performance of Prediction Models Trained on Sample without Current Violence, AUC

	LASSO (1)	Random Forest (2)	Adaptive Boosting (3)	Neural Network (4)	EBMA (5)
	Indonesia (social conflict, 2008-2014)				
Any violent event	0.711	0.714	0.712	0.680	0.719
$\geq$ 5 violent events	0.500	0.753	0.775	0.778	0.789
$\geq$ 1 s.d. increase in events	0.500	0.753	0.775	0.778	0.789
	Colombia (attacks and clashes, 1998-2005)				
Any violent event	0.743	0.750	0.751	0.708	0.754
$\geq$ 5 violent events	0.714	0.734	0.796	0.691	0.789
$\geq$ 1 s.d. increase in events	0.783	0.734	0.785	0.728	0.786

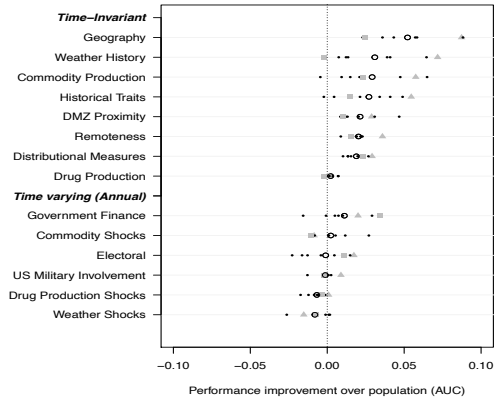
Notes: Each model is trained on data preceding the out-of-sample prediction year. Training data starts with 1991 data used to predict 1992 violence in Colombia and 2003 data for 2004 violence in Indonesia. Data are restricted to location-years that are not currently experiencing violence. Out-of-sample prediction begins in 1998 in Colombia and 2008 in Indonesia. The AUC is the area under the ROC curve, a measure of the trade-off between the true positive rate and false positive rate at different thresholds. We report average performance over the out-of-sample years.

Figure A.2: AUC Improvements from Individual Predictor Groups, New Conflicts

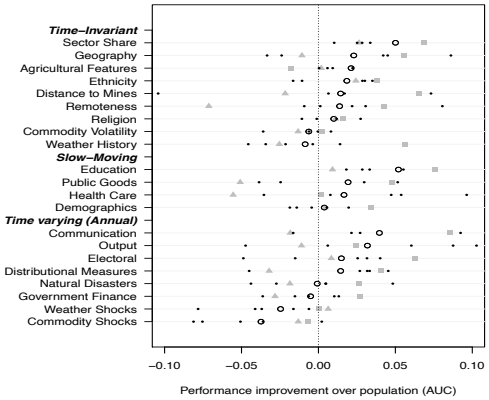
(a) Any violent event (Indonesia)



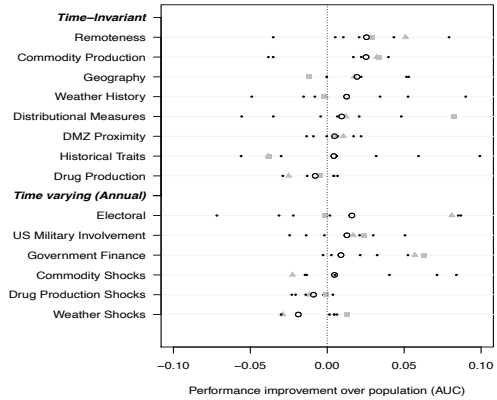
(b) Any violent event (Colombia)



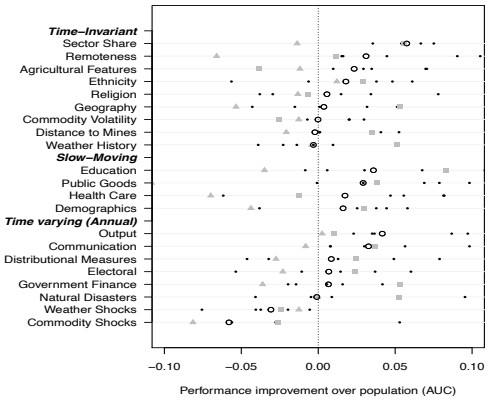
(c) ≥ 5 violent events (Indonesia)



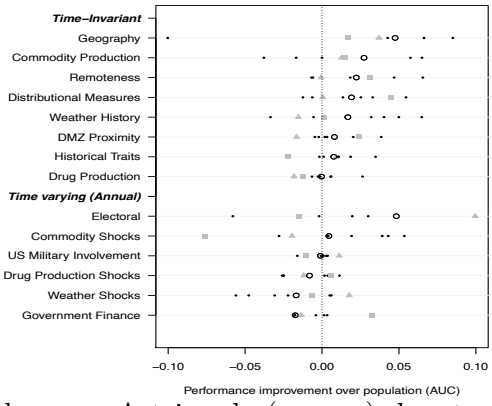
(d) ≥ 5 violent events (Colombia)



(e) ≥ 1 S.D. increase in violent events (Indonesia)



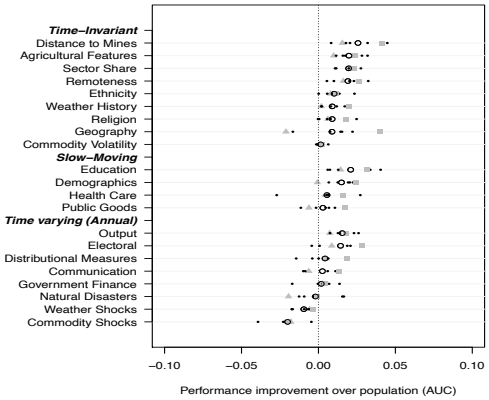
(f) ≥ 1 S.D. increase in violent events (Colombia)



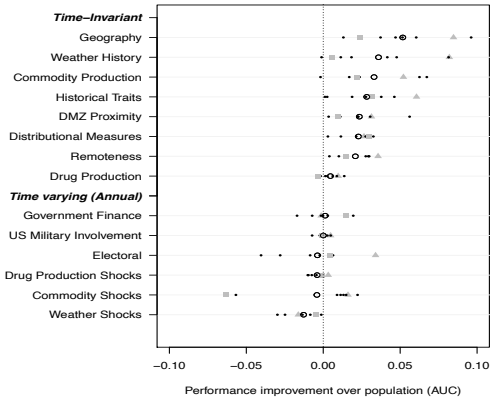
Notes: Dots represent performance in individual years. A triangle (square) denotes the first (last) year of the sample. The hollow circle is average performance across the years. Appendix C.3 lists the variables in each predictor group.

Figure A.3: AUC Improvements from Individual Predictor Groups, Trained with New Conflicts

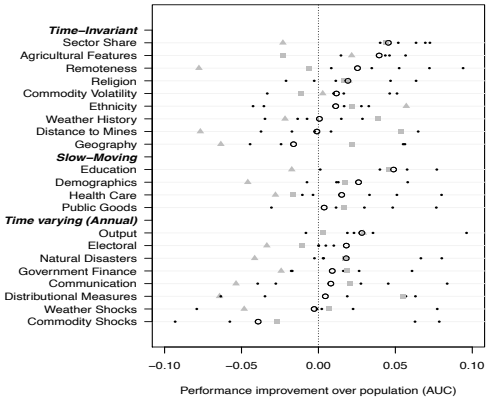
(a) Any violent event (Indonesia)



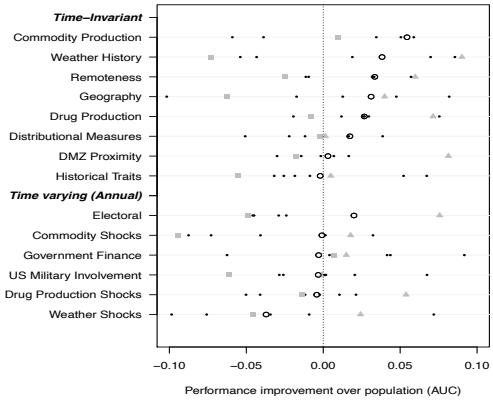
(b) Any violent event (Colombia)



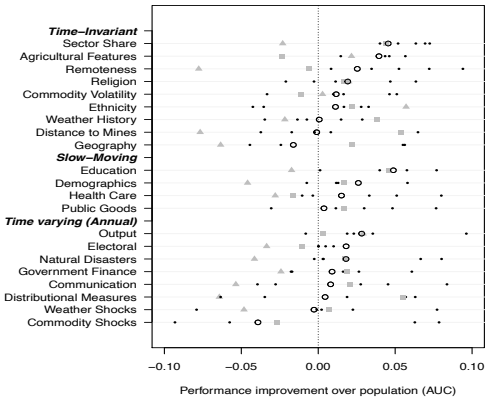
(c) ≥ 5 violent events (Indonesia)



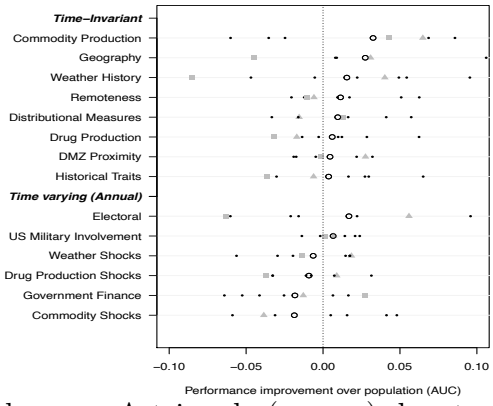
(d) ≥ 5 violent events (Colombia)



(e) ≥ 1 S.D. increase in violent events (Indonesia)



(f) ≥ 1 S.D. increase in violent events (Colombia)



Notes: Dots represent performance in individual years. A triangle (square) denotes the first (last) year of the sample. The hollow circle is average performance across the years. Appendix C.3 lists the variables in each predictor group.

A.6 Alternative Definitions of Escalation

Our decision to predict 1 standard deviation (s.d.) increases in the number of violent events is somewhat arbitrary. Here we consider alternative definitions of our escalation variable. First, we consider a 0.5 s.d. increase in the following year. Then, we consider a 0.5 s.d. and a 1 s.d. increase in the following *two* years. The longer horizon is meant to take into account that some run-ups in violence can occur over longer horizons. We stop at a two-year horizon, because as we measure conflict further into the future, we lose observations for both training and testing, since (i) observations require more data to calculate the given outcome (i.e., we cannot use the last year of data alone); (ii) for a true out-of-sample test, we must skip years that overlap between the training and testing sample (i.e., we cannot use the first year of test data); and (iii) outcomes become increasingly correlated, e.g., the two-year increase in violence from 2008 to 2010 is correlated with the two-year increase from 2009 to 2011.

Table A.8 reports the performance of models predicting these alternative escalation outcomes. Note that that each of these events is a subset of our baseline, so they are more likely to occur. That is, if a location experiences a greater 1 s.d. increase in violence over one year, that location also experiences a greater than 0.5 standard deviation over the next two years. In Indonesia, the rate of 0.5 s.d. escalations over one and two years is is 0.04 and 0.08 respectively, and the rate of 1 s.d. increases over two years is 0.08. In Colombia, the rate of 0.5 s.d. escalations over one and two years is 0.12 and 0.19 respectively, and the rate of 1 s.d. increases over two years is 0.12. Recall that the rates of our main escalation indicators (a one standard deviation increase over one year) are 0.03 in Indonesia and 0.04 in Colombia.

We see that we are able to predict smaller escalations and longer term escalations with a reasonable degree of accuracy. Performance is somewhat better when we move to a two-year horizon for an increase of 1 standard deviation, possibly due to an increase in the frequency of the outcome. Smaller one-year escalations also appear to be harder to predict over one and two year horizons. This may be because less extreme spikes in violence happen for more idiosyncratic reasons, and the outcome measure is noisier. Figure A.4 reports a version of Figure 1 where we see that the added value of the predictor groups is relatively unchanged when we predict different measures of escalation.

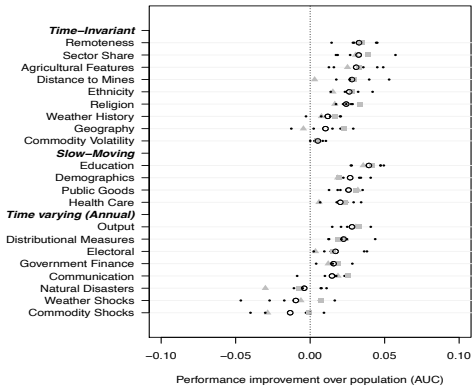
Table A.8: Out-of-Sample Performance of Prediction Models for Alternative Measures of Escalation, AUC

	LASSO (1)	Random Forest (2)	Adaptive Boosting (3)	Neural Network (4)	EBMA (5)
			Indonesia		
$\geq$ 1 s.d. increase in events	0.866	0.817	0.852	0.825	0.860
$\geq$ 0.5 s.d. increase in events	0.786	0.768	0.785	0.718	0.789
$\geq$ 1 s.d. increase in events over 2 years	0.854	0.840	0.852	0.821	0.862
$\geq$ 0.5 s.d. increase in events over 2 years	0.784	0.763	0.778	0.712	0.785
			Colombia		
$\geq$ 1 s.d. increase in events	0.802	0.787	0.796	0.741	0.801
$\geq$ 0.5 s.d. increase in events	0.761	0.750	0.760	0.708	0.762
$\geq$ 1 s.d. increase in events over 2 years	0.800	0.785	0.800	0.769	0.813
$\geq$ 0.5 s.d. increase in events over 2 years	0.743	0.739	0.758	0.693	0.759

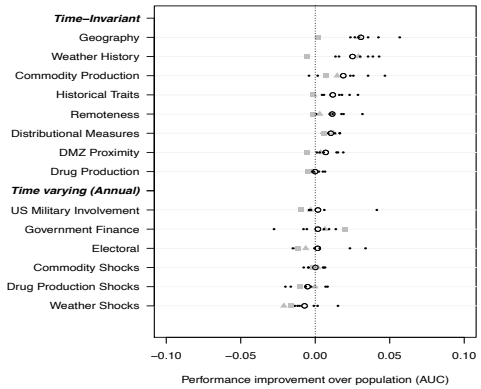
Notes: Prediction models are the same as in Table 1. The sample for the prediction evaluation is restricted to locations without any conflict in the previous year. Each model is trained on all available data preceding the out-of-sample prediction year. Training data starts in 1992 in Colombia and 2004 in Indonesia. Out-of-sample prediction begins in 1998 in Colombia and 2008 in Indonesia. The AUC is the area under the ROC curve, a measure of the trade-off between the true positive rate and false positive rate at different thresholds. We report average performance over the out-of-sample years.

Figure A.4: AUC Improvements from Individual Predictor Groups, Alternative Escalations

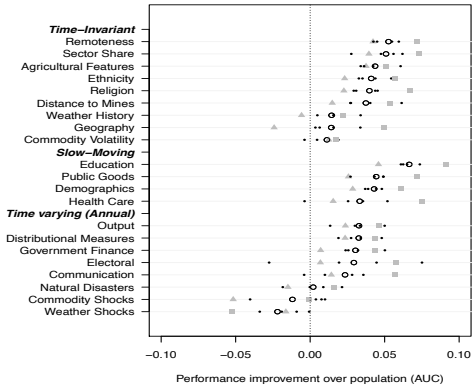
(a) $\geq$ .5 S.D. increase in violent events (Indonesia)



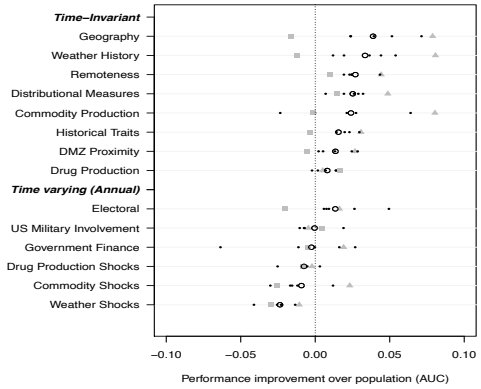
(b) $\geq$ .5 S.D. increase in violent events (Colombia)



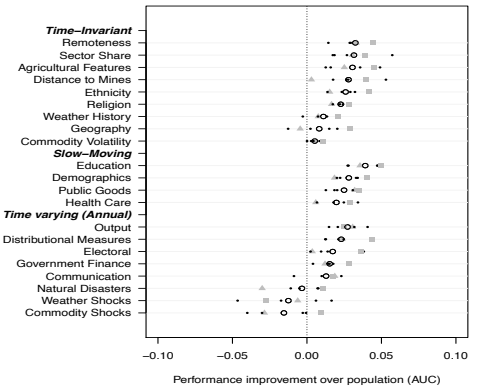
(c) $\geq$ 1 S.D. increase in next 2 years (Indonesia)



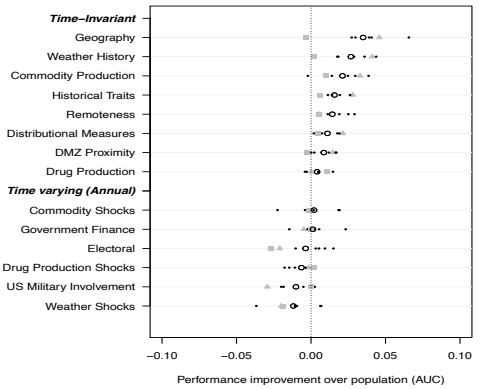
(d) $\geq$ 1 S.D. increase in next 2 years (Colombia)



(e) $\geq$ .5 S.D. increase in next 2 years (Indonesia)



(f) $\geq$ .5 S.D. increase in next 2 years (Colombia)



Notes: Dots represent performance in individual years. A triangle (square) denotes the first (last) year of the sample. The hollow circle is average performance across the years. Appendix C.3 lists the variables in each predictor group.

A.7 Alternative Predictor Groupings

In this section, we consider alternative groupings of our shock predictors. In our main analysis, we find that annual, time-varying predictors generally do not improve our ability to forecast year-ahead violence. Here we consider whether alternative groupings of these shocks affects this conclusion.

Taking as an example a coffee shock, causal studies of conflict have typically focused on the interaction between coffee cultivation (the weight) in some pre-determined time period and time-varying coffee prices (the price), while accounting for the direct effects of both prices and weight. In our main analysis, our 'shock' predictor groups include the interaction term and the time-varying price, while the time-invariant production measure or weight instead enters as part of a separate predictor group (see Appendix C.3).

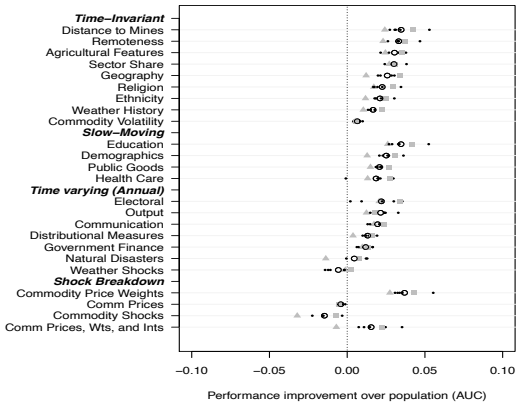
Since our machine learning methods can leverage various interactions between variables, it is possible that reorganizing our shock groupings could change our conclusions about their predictability. To explore this, in Figure A.5, we consider an alternative grouping of the variables. Component variables in "shocks" include a time-invariant weight that varies across locations, a time-varying price that varies over time (either the international price of a commodity or U.S. military spending in the case of the military aid shock in Colombia), and their interaction. For each such "shock" we separately look at the contributions of predictor groups consisting of various components. First, we construct groups consisting only of time invariant weights or only of the annual, time-varying prices. Then, we look at the time-varying prices along with the interaction term (without incorporating the weight component). Finally, we look at the contribution of all three components – the weight, international prices and their interaction.

In Colombia and Indonesia, the commodity variables tell a similar story. On their own, the time-invariant, commodity production measures (the weights) are often predictive of violence and offer substantial improvement over the baseline, population-based model (as we saw in Figure 1 with commodity production predictors in Columbia and distance to mines and agricultural predictors in Indonesia). Unsurprisingly, annual international prices alone do not help to predict one-year-ahead violence across various locations. Adding in the interaction term along with prices (without also including the production weight) also does not improve performance. In contrast, adding in the full group of predictors—weights, prices, and interactions—does have a positive effect on performance.

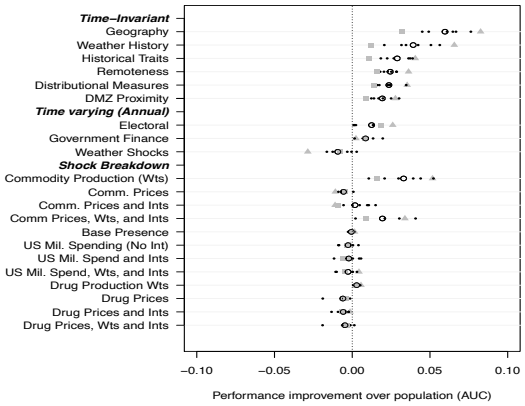
However, the results above suggest that much of the predictive power from incorporating all three components stems from the time-invariant production weights. We conclude that the takeaways from our baseline approach in Figure 1 remain largely insensitive to alternative permutations of predictor groups. For the year-ahead predictions, the time-invariant component of these shocks contribute more to improvements in predictive performance.

Figure A.5: AUC Improvements from Individual Predictor Groups

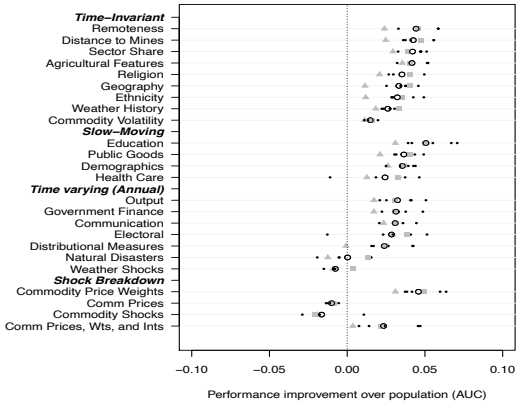
(a) Any violent event (Indonesia)



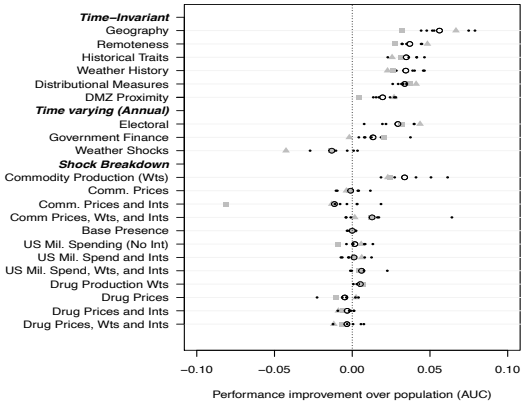
(b) Any violent event (Colombia)



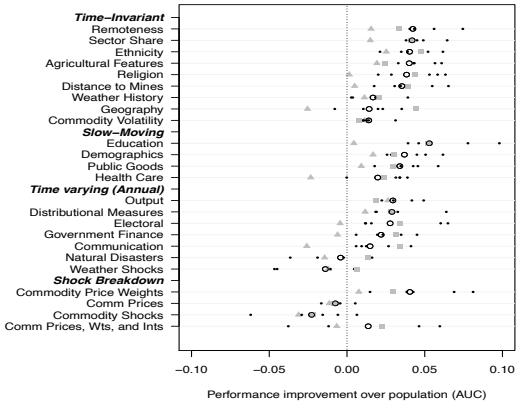
(c) ≥ 5 violent events (Indonesia)



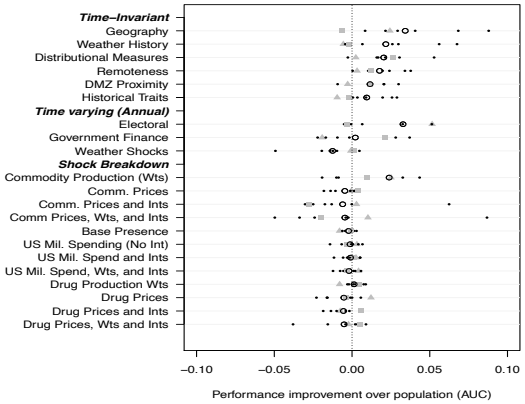
(d) ≥ 5 violent events (Colombia)



(e) ≥ 1 S.D. increase in violent events (Indonesia)



(f) ≥ 1 S.D. increase in violent events (Colombia)



Notes: Dots represent performance in individual years. A triangle (square) denotes the first (last) year of the sample. The hollow circle is average performance across the years. Appendix C.3 lists the variables in each predictor group.

A.8 Alternative Colombia Conflict Data

An alternative dataset on violence in Colombia from U. Rosario has become available in recent years, covering violent events from 1995 to 2014. In theory, the effort is meant to measure the same types of conflicts as the CERAC dataset. However, due to some classification differences, and the inclusion of some additional categories of violence detailed in the Appendix A.8.1, we do not view the data as a direct extension but rather as an alternative dataset. Nevertheless, as a robustness check, we have extended our predictor dataset and our violence dataset, using U.Rosario from 2006 to 2014.

A.8.1 Comparison of U. Rosario and CERAC Datasets

In general, the U. Rosario dataset follows similar protocols for including violent events as the CERAC dataset. However, the two datasets are not perfectly aligned due to at least two coding differences. The U. Rosario dataset appears to include an additional set of political events, which are defined as events targeted toward political actors. Thus, it includes a set of events as attacks that encompass executions, disappearances, personal attacks and kidnapping of political individuals. The CERAC dataset, in contrast, typically did not include events like kidnappings, unless it coincided with other war-related actions that led to the classification of an event as an attack. In addition, the two datasets treat the coding of clashes and attacks differently in the case of complex events that involve a sequence of events involving both attack and clash elements. In particular, when the government military is involved, if there is an event in which the government takes a unilateral action, and this is followed by an exchange of fire with the armed groups, this would be coded as a government attack followed by a clash under CERAC. However, the U. Rosario dataset would instead categorize the event as a clash.

To examine and account for possible discrepancies on the inclusion of events, we construct two series for the number of incidents in the U. Rosario dataset: one that accounts for political events (extension 1) and another that excludes them (extension 2). Figure A.6 shows that both U. Rosario series follow similar trends as CERAC, rising and peaking at 2002, and falling thereafter. When including political events (extension 1), the number of U. Rosario events exceeds the number of CERAC events, consistent with the more liberal inclusion criteria of events targeting political

actors. When political events are removed (extension 2), the CERAC data slightly exceeds the U. Rosario events, consistent with the more liberal coding of complex events under CERAC. Figure A.7 demonstrates similar patterns for our three municipality-level outcome measures: any violent events, ≥ 5 violent events, ≥ 1 s.d. increase in violent events. Overall, the trend similarities in the two data sources suggest that the two data sources provide a similar account of conflict dynamics in Colombia over this period.

Figure A.6: Comparing CERAC and U. Rosario Datasets

Number of Violent Events by Year

(a) including political violence (ext. 1) (b) excluding political violence (ext. 2)

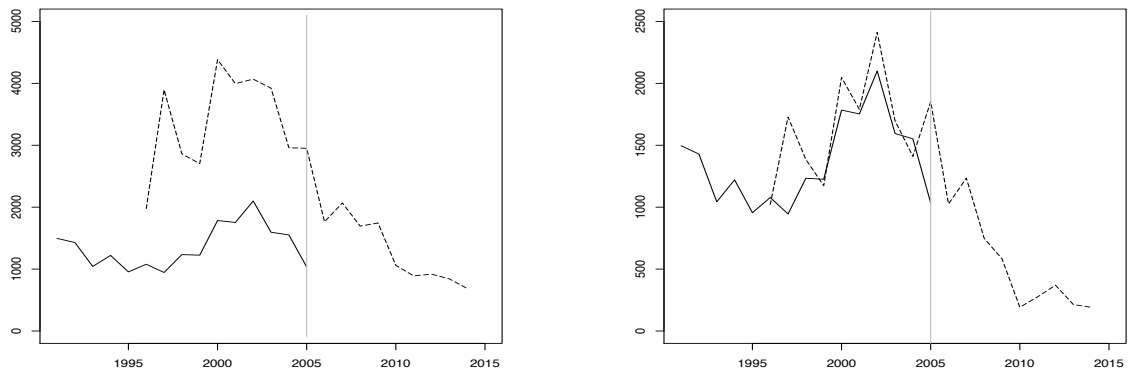
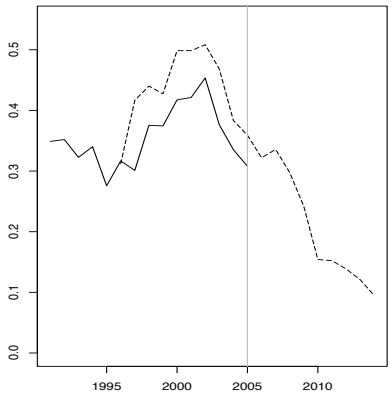


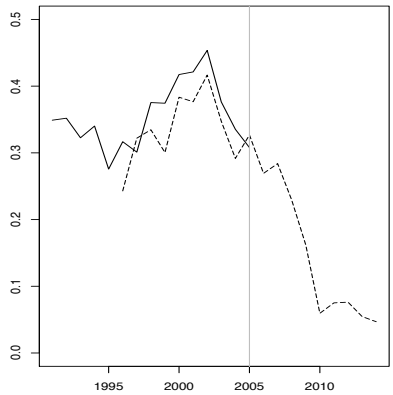
Figure A.7: Comparing CERAC and U. Rosario Datasets

Share of municipalities reporting at least one violent event

(a) including political violence (ext. 1)



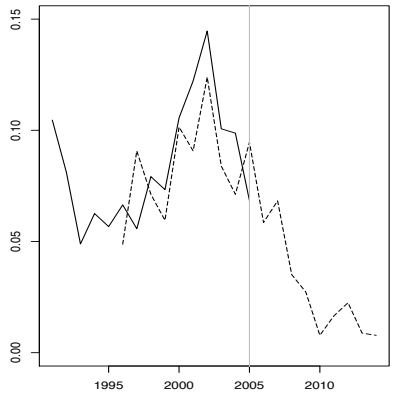
(b) excluding political violence (ext. 2)



Share of municipalities reporting five or more violent events

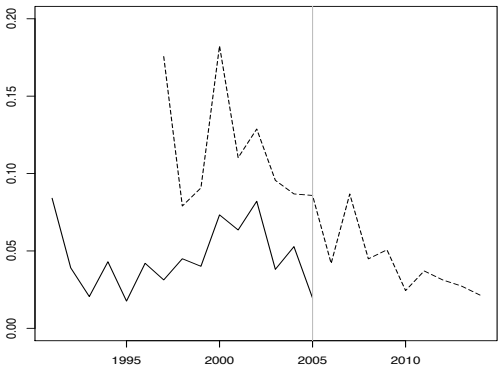
(c) including political violence (ext. 1)

(d) excluding political violence (ext. 2)

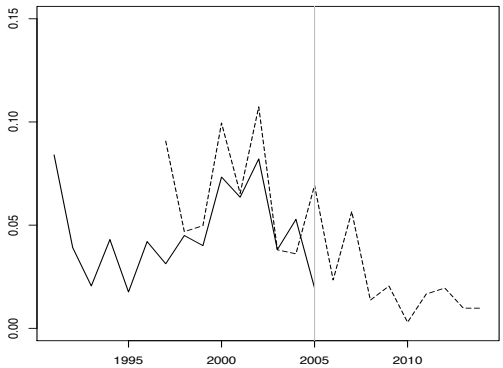


Share of municipalities with ≥ 1 s.d. increase in violent events

(e) including political violence (ext. 1)



(f) excluding political events (ext. 2)



A.8.2 Prediction Performance in U. Rosario Data

Using the now extended data, we construct the same indicators that we use in our main analysis—indicators of any event, more than 5 events, and a 1 standard deviation increase in violent events. We extend our main dataset by using the U. Rosario measures from 2006 onward. We do this for two different sets of events—event counts including political violence and event counts excluding political violence—and generate predictions all the way out to 2014. As in the original analysis with CERAC, we begin training in 1992 and testing in 1998. However, because we are interested in the potential for a longer panel to generate better models, we use all available data for training the model, regardless of the break in the violence series. For example, for our 2008 prediction, the models are trained using all observation-years from 1992 to 2007, including, e.g., municipality–years from 1997 in which violence is measured by CERAC and municipality-years from 2006 in which violence is measured by U. Rosario.

Table A.9 reports average performance statistics. We see that patterns are similar to those for our main violence prediction measure and sample as reported in Table 1. It appears that performance does improve when we extend the series, in particular for the series that excludes political events. The longer data series may decrease model variance, generating better out-of-sample predictions as time goes on. It is also possible that as violence became rarer, it became more concentrated in the locations where we would predict the most risk. Indeed, many of the important predictors for the shorter CERAC series remain important with the extended series combining CERAC and U. Rosario (see Figure A.8). Overall, prediction with this extended series delivers broadly similar findings as our main predictions in the paper with a single, shorter data series.

Table A.9: Extending the Colombia Analysis to 2014 by Combining U. Rosario and CERAC Data

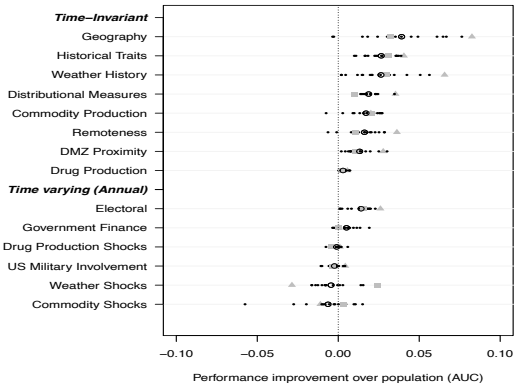
	LASSO (1)	Random Forest (2)	Adaptive Boosting (3)	Neural Network (4)	EBMA (5)
	U. Rosario Incl. Political Violence				
Any violent event	0.863	0.862	0.867	0.847	0.867
$\geq$ 5 violent events	0.925	0.918	0.922	0.905	0.924
$\geq$ 1 s.d. increase in events	0.827	0.831	0.839	0.772	0.839
	U. Rosario Excl. Political Violence				
Any violent event	0.872	0.871	0.875	0.857	0.877
$\geq$ 5 violent events	0.932	0.924	0.929	0.909	0.932
$\geq$ 1 s.d. increase in events	0.828	0.825	0.842	0.793	0.840

Notes: Each model is trained on all available data preceding the out-of-sample prediction year. Training data starts in 1992 in Colombia. Out-of-sample prediction begins in 1998. The AUC is the area under the ROC curve, a measure of the trade-off between the true positive rate and false positive rate at different thresholds. We report average performance over the out-of-sample years.

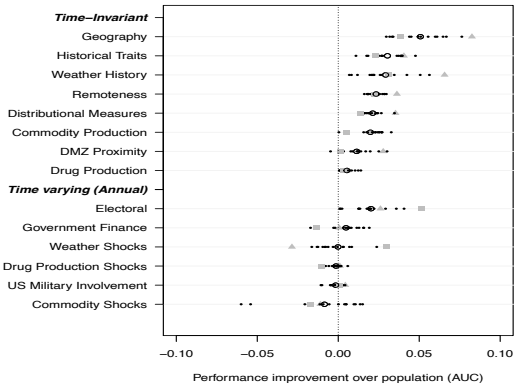
Figure A.8: AUC Improvements from Individual Predictor Groups, Colombia Extensions

Any Violent Event

(a) including political violence

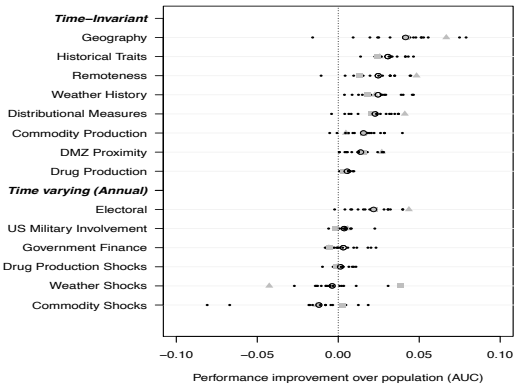


(b) excluding political violence

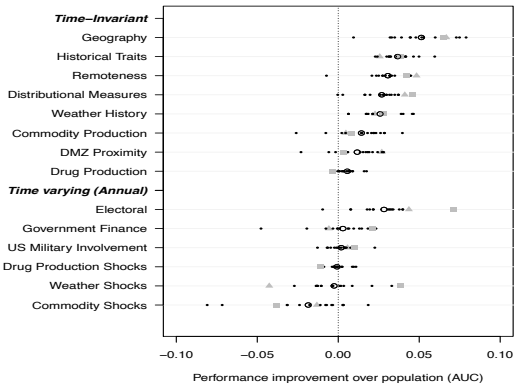


$\geq$ 5 Violent Events

(c) including political violence

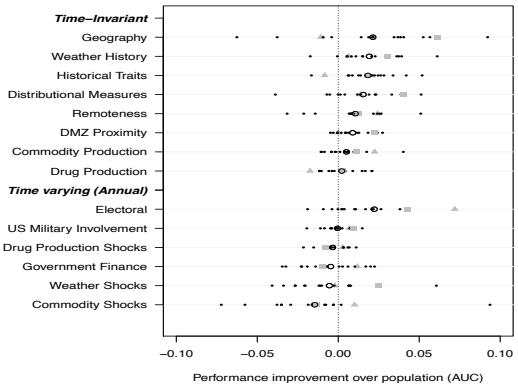


(d) excluding political violence

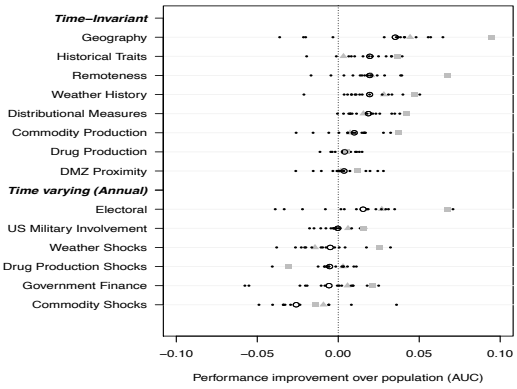


$\geq$ 1 s.d. Increase in Violent Events

(e) including political violence



(f) excluding political violence



Notes: Colombia Extension 1 includes political events in the URosario dataset. Extension 2 ignores these political events. Dots represent performance in individual years. A triangle (square) denotes the first (last) year of the sample. The hollow circle is average performance across the years. Appendix C.3 lists the variables in each predictor group.

B Methodological Details

B.1 Hyper-parameter Choices

Each of the machine learning methods that we use involves a number of hyper-parameter choices. In general, the hyper-parameters govern the degree of flexibility afforded to each algorithm. The optimal choice of these parameters balances the models' ability to uncover complex relationships against the risk that the algorithm over-fits to noise in the data. We choose these parameters using a mixture of rules-of-thumb and 5-fold cross validation. In this section, we detail these choices.

LASSO. We implement a logistic LASSO where predictors are standardized to have a mean of zero and unit variance before they are fed into the fitting algorithm. LASSO and ridge regression are two closely-related penalized regression techniques. While LASSO penalizes the sum of the absolute value of the regression coefficients, ridge regression penalizes the sum of squares. We follow best practices as in Blair et al. (2017) and use a weighted average of the two penalties, where the weight for the LASSO penalty is $\alpha = 0.95$ and the weight on the ridge penalty is $1 - \alpha$.

For each country in each year of estimation, we find the optimal penalty parameter λ by searching over a grid of candidate values and testing the penalties using 5-fold cross validation. The grid includes 140 values from 1 down to 1.8×10^{-8}. The grid was chosen to make sure that optimally chosen values would be interior. We repeat this process 10 times, and take the average optimal value.

Random Forests. These are collections of many trees which are fit to random subsets of the data and then are averaged together. The underlying logic is that the individual trees may be overfit to their respective dataset, but since each tree is fit using a different set of predictors, the overfitting averages out over the entire forest. We choose mean-squared error as the loss function for the random forest. Beyond the choice of the loss function, there are three important hyper-parameters to choose.

First, we must specify a rule governing how large trees can be. We could specify the minimum number of observations per terminal node or the maximum number of terminal nodes. In general, larger trees allow for more complexity but also increase the risk of overfitting (though this risk is

mitigated by averaging over the entire forest). We limit our trees to 60 terminal nodes, as it is within the range suggested by Hastie et al. (2001), and we find little improvement in performance past this number.

Second, we choose the number of trees in the forest. Since each tree is independent, additional trees simply reduce variance of estimates and do not add to bias. As a result, more trees are always better. At the same time, performance gains from additional trees generally diminish quickly, while the computing time costs of fitting individual trees does not. In experimentation using cross-validation, we observed little increase in performance as the number of trees exceeds 100. Therefore, we choose to fit 100 trees in each forest.

Finally, we must choose the number of covariates to be considered at each branch in the trees. We follow the rule-of-thumb of using one-third of the potential covariates at each split (Hastie et al., 2001).

Gradient Boosted Machines. These are comprised of decision trees that are fit sequentially to the residual variation in the predictand that was not predicted by previous trees. Unlike random forests which leverage many overfit trees, gradient boosted machines are meant to learn slowly, with each tree explaining a small amount of additional variation. Therefore, the key parameter is the shrinkage parameter which limits the extent to which each tree can contribute to the machine's overall prediction. Best practice sets these shrinkage parameters as low as possible. We choose a shrinkage parameter $\lambda = 0.1$.

As for random forests, GBM requires the implementer to specify the complexity of trees and the number of trees in the ensemble. We specify the number of terminal nodes as 8 in each tree, a standard parameter value for these models (Hastie et al., 2001). The number of trees to include in the model is the key hyper-parameter that drives the overfitting versus complexity trade-off. If there are too few trees, the predictions will be a very simple function of the predictors, whereas if there are too many trees, the later trees will be fit to noise generated by idiosyncrasies of the training sample. To manage this tradeoff, we choose the number of trees by 5-fold cross validation over a grid of candidate sizes. We choose candidate values between 5 and 150. Though trial and error, we found that the optimal values are within this range. We average the results of 10 such trials to get an optimal number of trees in each year.

Neural Networks. These are built from weighted combinations of features, and an activation function through which these combinations are passed. We use a single hidden layer neural network, and the network is trained via back-propagation. Our neural networks use a sigmoid activation function.

The major parameter governing complexity is the number of nodes to allow in this single layer. We choose the number of nodes by searching over a grid of values and employing 5-fold cross validation to test each candidate number of nodes. As with the other algorithms, we repeat this process 10 times, and choose the average parameter.

Since each predictor has a weight for each node in the hidden layer, training of the neural network can involve the computation of thousands of parameters. The computation costs are magnified during the grid search process. To alleviate this pressure, we preprocess the data by standardizing the predictors and calculating principal components of the predictor set. We use these principal components as predictors. We use 30 principal components in Indonesia and 20 in Colombia. This rotation of the predictor space dramatically increases speed without throwing away much important variation. We also test different neural net architectures that do not require this pre-processing of the data, described in Appendix B.3 below.

Our **Ensemble Bayesian Model Average** is computed by generating a 5-fold cross-validation set of probabilistic predictions for each algorithm using the parameters chosen above. We take these predictions and calculate posterior likelihoods that each model is correct given their predictions and the observed levels of violence. These likelihoods, when normalized to sum to 1 give us weights for our model average. We repeat this process 10 times to get ten sets of weights and average them to aggregate our predictions.

B.2 Alternative Cross-Validation

We choose our hyper-parameters by 5-fold cross-validation, which introduces some randomness into our analysis. With a small sample, we might worry that idiosyncratic splits of data might generate extreme values of hyper-parameters that do not produce strong performance out of sample. For this reason, we repeat the cross-validation procedure 10 times and take the average optimal hyper-

parameters, rounding to the nearest integer where hyper-parameters need to be discrete.

We also compare our baseline models' performance to models trained using hyper-parameters from a single cross-validation run in Table B.1. We see that there are not appreciable returns to our approach, as models using hyper-parameters from a single CV run do just about as well, on average. At the same time, the computational costs of our approach are minimal, so we maintain our conservative strategy of averaging over 10 runs throughout the paper.

Table B.1: Performance with Hyper-Parameters chosen from 1 CV run, AUC

	LASSO (1)	Random Forest (2)	Adaptive Boosting (3)	Neural Network (4)	EBMA (5)
	(a) Any violent event				
	Indonesia				
10 CV Runs (Baseline)	0.819	0.818	0.823	0.792	0.823
1 CV Run	0.818	0.818	0.823	0.786	0.822
	Colombia				
10 CV Runs (Baseline)	0.845	0.847	0.849	0.825	0.850
1 CV Run	0.845	0.847	0.849	0.823	0.850
	(b) $\geq$ 5 violent events				
	Indonesia				
10 CV Runs (Baseline)	0.940	0.935	0.942	0.910	0.941
1 CV Run	0.940	0.935	0.942	0.910	0.940
	Colombia				
10 CV Runs (Baseline)	0.914	0.911	0.910	0.886	0.915
1 CV Run	0.914	0.911	0.911	0.880	0.916
	(c) $\geq$ 1 s.d. increase in events				
	Indonesia				
10 CV Runs (Baseline)	0.866	0.817	0.852	0.825	0.860
1 CV Run	0.865	0.817	0.853	0.807	0.862
	Colombia				
10 CV Runs (Baseline)	0.802	0.787	0.796	0.741	0.801
1 CV Run	0.801	0.787	0.796	0.732	0.800

Notes: Each model is trained on all available data preceding the out-of-sample prediction year. Training data starts in 1992 in Colombia and 2004 in Indonesia. Out-of-sample prediction begins in 1998 in Colombia and 2008 in Indonesia. The AUC is the area under the ROC curve, a measure of the trade-off between the true positive rate and false positive rate at different thresholds. We report average performance over the out-of-sample years.

B.3 Alternative Neural Networks

Neural networks are notoriously difficult to train as there are many choices about network structure that the researcher must make—including number of nodes and number of layers—as well as choices about regularization parameters to prevent over-fitting of weights. These choices are generally made using a combination of rules of thumb and cross-validation. In this section, we explore alternative neural network constructions and compare their performance to our baseline. Our goal is not to conduct an exhaustive search across all potential neural net architectures in order to identify one that strictly dominates the others. Rather, we aim to test a more limited number of equally plausible alternatives to explore whether other approaches are likely to generate better performance than our baseline.

With data sets such as ours, which are relatively small for the machine learning literature, overfitting is a chief concern. As a result, we opt to reduce the dimension of our predictor space using principal components—20 in Colombia and 30 in Indonesia—so that the neural networks will have fewer weights to optimize. In column 2 of B.2 we eliminate the dimension reduction by principal components. We see a drop in performance across the board. Column 3 adds further flexibility by eliminating regularization for the weight-fitting objective function. In this case, performance drops to about 0.5. These overly flexible models perform no better than would be expected from random chance. In column 4, we allow for more complex interactions by adding an additional hidden layer. This added flexibility still underperforms our benchmark model.

Finally, in column 5, we fit a model using the Keras package in R. Keras is a wrapper for the TensorFlow platform, developed by Google AI, which has become widely used in recent years. With this package, we fit a network that consists of two fully connected layers and two dropout layers. The dropout layers randomly set a subset of hidden nodes to zero during the training stage. This procedure guards against overfitting, since the set of active nodes changes stochastically as fitting occurs. This is a substitute for a direct regularization term in the objective function (Srivastava et al., 2014). We see that performance is markedly worse than for our main neural network specification. This table is by no means exhaustive, but it does offer some justification for the conservative approach we take in our baseline model. In general, our environment is not ideal for the deployment of neural networks whose strength comes from their ability to identify complex

patterns in very large datasets.

Table B.2: Alternative Neural Networks, AUC

	Baseline PC Step 1 layer nodes by CV 0 Regularization (1)	No PC Step 1 layer 7 nodes 0.15 Regularization (2)	No PC Step 1 layer 7 nodes 0 Regularization (3)	No PC Step 2 layers 10,5 nodes 0 Regularization (4)	No PC Step 2 layers 256, 128 nodes .5, .3 dropout (5)
			Indonesia		
Any violent event	0.792	0.658	0.636	0.615	0.516
$\geq$ 5 violent events	0.910	0.720	0.500	0.713	0.655
$\geq$ 1 s.d. increase in events	0.825	0.697	0.500	0.656	0.682
			Colombia		
Any violent event	0.825	0.646	0.620	0.662	0.502
$\geq$ 5 violent events	0.886	0.660	0.500	0.662	0.524
$\geq$ 1 s.d. increase in events	0.741	0.596	0.500	0.598	0.519

Notes: Each model is trained on all available data preceding the out-of-sample prediction year. Training data starts in 1992 in Colombia and 2004 in Indonesia. Out-of-sample prediction begins in 1998 in Colombia and 2008 in Indonesia. The AUC is the area under the ROC curve, a measure of the trade-off between the true positive rate and false positive rate at different thresholds. We report average performance over the out-of-sample years.

C Data Appendix

C.1 Indonesia

C.1.1 Administrative Levels

Indonesia is divided into four tiers of government. In 2014, 34 provinces were divided into 514 districts composed of 7,094 subdistricts with more than 80,000 villages. The number of districts and the number of subdistricts have ballooned over time. Concurrent with the wave of decentralization, the Indonesia government created many new districts through a process of redistricting known colloquially as *pemekaran* or blossoming. After remaining steady from 1980 to 1998, the number of new districts ballooned from 302 in 1999 to 514 in 2014. Subdistricts have also split (and a small few amalgamated) over time. Despite the increase in the number of districts, the number of subdistricts per district has increased. The number of villages has remained relatively more steady and thus the number of villages per-district and per-subdistrict decline over time.

The unit of analysis in this paper is the 2000 subdistrict. Since subdistricts have increased over time, we map them back to the larger 2000 borders. The SNPK and most covariate data (unless otherwise specified) is recorded at units of analysis smaller than the 2000 subdistrict level, enabling us to map these data onto the 2000 boundaries.

C.1.2 Violence

The conflict data comes from the Indonesian National Violence Monitoring System (known by its Indonesian acronym SNPK). The data are reported at the 2011 subdistrict level and include incident dates. Subdistrict codes are non-missing in 84% of cases. We aggregate incidents to the 2000 subdistrict borders in each year. Our main conflict measures are binary indicators for any conflict in a given subdistrict–year. Table C.1 presents the violence definitions in the SNPK.

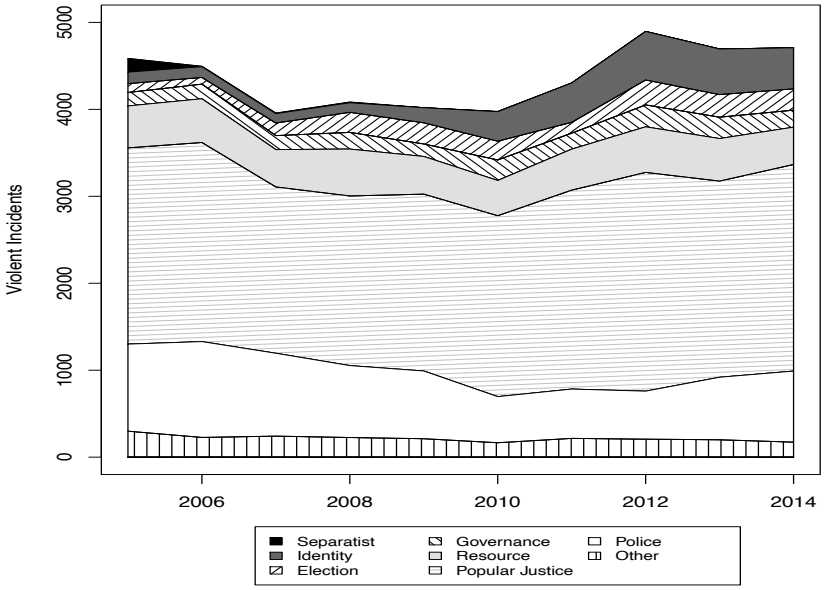
Table C.2 reports the rates at which each of these indicators occur. While around half of subdistricts experience some conflict in a given year, only around 12% experience more than five incidents, and even fewer experience a large increase in the number of events relative to the prior year. Finally, Figure C.2 reports the distribution of conflict counts in subdistrict-years with any

violence. We see tremendous right skew, motivating our approach of predicting subdistricts with five or more incidents of violence.

Table C.1: Violence Categories in the SNPK

Resource Conflict	Violence triggered by resource disputes (land, mining, access to employment, salary, pollution, etc.).
Governance Conflict	Violence is triggered by government policies or programs (public services, corruption, subsidy, region splitting, etc).
Popular Justice Conflict	Violence perpetrated to respond to/punish actual or perceived wrong (group violence only).
Elections and Appointment Conflict	Conflict Violence triggered by electoral competition or bureaucratic appointments.
Separatist Conflict	Violence triggered by efforts to secede from the Unitary State of the Republic of Indonesia (NKRI).
Identity-Based Conflict	Violence triggered by group identity (religion, ethnicity, tribe, etc).
Other Conflict	Violence triggered by other issue.
Violence During Law Enforcement	Violent action taken by members of formal security forces to perform law-enforcement functions (includes use of violence mandated by law as well as violence that exceeds mandate for example torture or extrajudicial-shooting).
Violent Crime	Criminal violence not triggered by prior dispute or directed towards specific targets.
Domestic Violence	Physical violence perpetrated by family member(s) against other family member(s) living under one roof/same house including against domestic workers and violence between cohabiting couples.

Figure C.1: Violence by Category, Indonesia



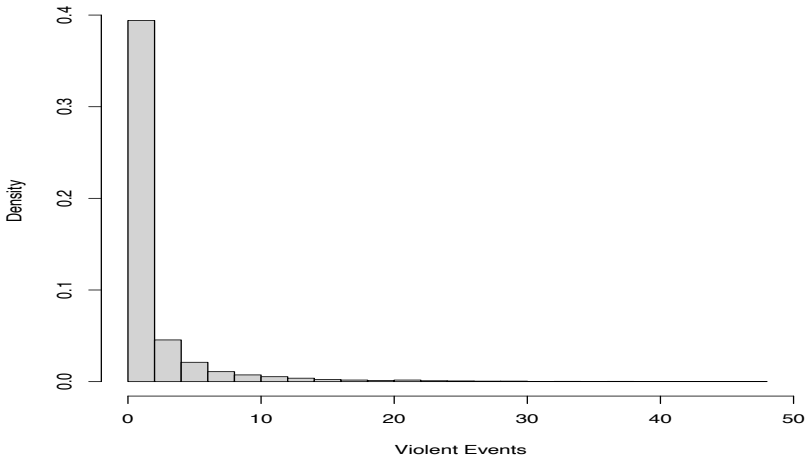
Notes: The graph above plots violent incident counts according to the SNPK. Crimes and domestic violence are not included. The panel is balanced from 2005 onward.

Table C.2: Annual Rates of Conflict in Indonesian Subdistricts

	Any Conflict (1)	≥ 5 Conflict Incidents (2)	≥ 1 std. dev. increase in Incidents (3)
2005	0.540	0.130	
2006	0.537	0.122	0.029
2007	0.509	0.114	0.028
2008	0.550	0.125	0.030
2009	0.527	0.120	0.031
2010	0.536	0.119	0.034
2011	0.531	0.127	0.040
2012	0.573	0.144	0.045
2013	0.541	0.127	0.034
2014	0.539	0.129	0.034

Notes: Conflict incidents above exclude crime and domestic violence.

Figure C.2: Distribution of Violent Events, Indonesia



Notes: This histogram depicts the distribution of violent event counts in subdistrict-years with at least one incident. 19 subdistict-years with over 50 events are excluded from graph for scaling.

C.1.3 Comparison to Other Violence Data

The SNPK offers significantly more comprehensive coverage than a widely used, cross-country, subnational data source. The Uppsala Conflict Data Program (UCDP) Georeferenced Event Data (GED) (Sundberg and Melander, 2013) has been fruitfully deployed in a range of subnational conflict studies and with particular success in sub-Saharan Africa alongside the widely used Armed Conflict Location & Event Data Project (ACLED) data. Both data are available for Indonesia, but ACLED coverage only begins in 2015.

Mapping the UCDP-GED events to our (unbalanced) subdistrict-year panel for 2000–2014, we find very limited coverage of social conflict events in Indonesia. UCDP-GED covers only 328 subdistrict-year incidents between 2000-2014 within the geographical boundaries covered by SNPK. Of these, 316 subdistrict-years also have an incident in SNPK. Meanwhile, 24,784 subdistrict-years have incidents that are recorded in SNPK but not in UCDP-GED. Together, these 24,784 subdistrict-year incidents involve over 24,331 deaths. The more limited coverage by UCDP-GED is explained by both its more narrow focus on large-scale conflict and by its reliance on international news sources and/or English-based ones in Jakarta. The SNPK offers much deeper coverage

precisely because it digitized millions of old newspapers from outlying regions of the country that allowed for coverage of violence that may have otherwise missed the attention of international reporters. Barron et al. (2016) offer a systematic comparison of SNPK to alternative Indonesian sources, consistently finding greater coverage by the SNPK. They attribute the differences to the use of 123 provincial and sub-provincial newspapers, far more than prior efforts.

C.1.4 Covariates

2000 Population Census

We use the following predictors constructed at the 2000 subdistrict level from Indonesia's 2000 universal Population Census.

- **Employment Shares:** Fraction of persons in agriculture, forestry and fishing, industry, services, trade, and transportation. We compute the fraction of persons in self-employment and the fraction that are employers.
- **Demographic Variables:** Share of people in each of the following religions: Muslim, Catholic, Protestant, Hindu, Buddhist, Confucian, and other. Shares of the population that are ethnic Chinese and ethnic Arab. Share of men and share married. Finally, share of persons under 10 and from 10 to 30 years old.
- **Education:** Share of individuals who completed no school and the average years of schooling.
- **Rural Population Share:** Share of villages within the subdistrict classified as rural.
- **Ethnic Fractionalization:** Ethnic fractionalization in district d is given by $F = \sum_{g=1}^{M_e} \pi_g(1-\pi_g)$, where M_e is the number of ethnic groups in the district, and π_g is the population share of group g as reported in the 2000 Census. We observe over 1,000 ethnicities and sub-ethnicities speaking over 400 languages.
- **Ethnic Polarization:** Same as ethnic fractionalization but own group share is emphasized: $P = \sum_{g=1}^{M_e} \pi_g^2(1-\pi_g)$.
- **Religious Fractionalization:** Religious polarization, $RF = \sum_{g=1}^{M_r} \sum_{h=1}^{M_r} \pi_g \pi_h$, where M_r is the number of religious groups, and π_g (π_h) is the population share of group g (h). There

are seven religions recorded in the Census, but in most districts, there is a single cleavage between a Muslim and a non-Muslim group.

- **Religious Polarization:** Same as ethnic polarization but own group share is emphasized: $RP = \sum_{g=1}^{M_e} \pi_g^2(1 - \pi_g)$.
- **Ethnic Residential Segregation:** Following Alesina and Zhuravskaya (2011), we use the 2000 Census to compute Ethnic segregation by comparing ethnic fractionalization at the village level to that of the subdistrict level. Specifically we compute:

$$S = \frac{1}{M-1} \sum_{m=1}^{M} \sum_{s=1}^{S} \frac{t_s}{T} \frac{(\pi_{sm} - \pi_m)^2}{\pi_m}$$

  where M is the number of ethnic groups, T is the total population of the subdistrict, t_s is the population in village s, π_m is the fraction of group m in the district, and π_{sm} is the fraction of group m in subdistrict s. We drop the smallest 1% of ethnic groups so that M remains reasonable (< 25).
- **Interethnic Marriage:** We also construct the share of interethnic marriages in each subdistrict.

Potensi Desa (*Podes*)

We construct the following variables using the quasi-triennial administrative village census *Potensi Desa*, abbreviated *Podes*. We use the 1999, 2002, 2005, 2008 and 2011 rounds.

- **Police and Security:** We compute the number of security and police posts as well as the number of security officers.
- **Population:** Subdistrict population in each round.
- **Health Care:** We construct the number of health care facilities per thousand people, which we classify into hospitals (hospital, maternity hospitals, and polyclinics) and middle care (*puskesmas* and supporting *puskesmas*). We also construct the numbers of doctors, midwives, and traditional birth attendants per thousand.
- **Schooling:** We compute the number of kindergarten, primary, junior high schools, high schools, and universities per thousand. We also calculate the share of villages with any

Islamic boarding schools.

- **Crime and Conflict:** In each wave, we record the share of villages that experienced any crime or conflict. We do the same for each type of crime recorded (theft, robbery, looting, thuggery, arson, rape, murder) and type of conflict (between citizens, security forces, students, tribes). We also include the total number of deaths and injuries resulting from conflict
- **Communication:** In each wave, we record the number of households with a telephone, and the percent of villages with access to public television, a phone stall, and postal access. For the 2005, 2008, and 2011 waves, we also have indicators for whether a strong or weak cell phone signal is present in the village, and in 2011 an indicator for a cell tower in the village. In the 1999 wave, we also have the percent of households with a radio, telephone, television, and satellite dish.
- **Transportation:** We include the percent of villages with a bridge, bus terminal, seaport, road lighting, and a passable road in the 1999 wave. We have the road lighting and passable road indicators in all waves. We also construct the village-population-weighted distance to the district capital in each wave. It can vary due to the creation of new district capital as a result of district proliferation. We also include the land area in 1999.
- **Public Goods and Infrastructure:** We include the percent of villages with access to safe water, and garbage facilities, toilets, gas/electric cooking facilities, and a permanent market. We also construct the number of households with electricity. In 2002, we also compute the population-weighted mean access to irrigation, an indicator for rice being the primary commodity, and the number of small and large rice mills per village.
- **Natural Disasters:** Each wave records the share of villages that experienced a mudslide, flood, fire, earthquake, and other disaster in the past three years.
- **Subdistrict Revenues:** From the 2002 wave, we construct total revenue and expenditures at the subdistrict level. We include the breakdown of revenue into its various sources: the different levels of government, taxes, social organizations, ROSCAs, and other villages.

Whenever available, we use measures from the previous two *Podes* rounds as predictors.

National Socioeconomic Survey (*Susenas*)

The National Socioeconomic Survey (*Susenas*) is an annual survey of about 200,000 nationally representative households. The survey measures household income and expenditures. It is not formally representative at the subdistrict level, so we construct measures at the larger district level.

- **Income and Expenditures:** We use the expenditure data to construct mean and median per-capita total, food, and non-food expenditure. We also calculate unemployment rates and working rates.
- **Income Inequality:** We use the expenditure data to construct an expenditure Gini and the ratio of the 80th percentile to the 20th percentile.
- **Poverty:** We construct the share below the poverty line (P_0) as well as higher order (P_1) and (P_2) poverty measures. We also construct indicators for household participation in the nationally subsidized rice for the poor program (*Raskin*).
- **Health:** We construct the fraction of children and adults with self-reported health problems.
- **Education:** We construct the fraction of 5–21 year olds in school, as well as the percent of adults (15+) who are literate.
- **Communication:** We construct indicators for whether anyone in the household has a telephone, cell phone, computer, or has accessed the internet in the past 3 months (2006–2013).
- **Age:** We also use this data to measure the annual average age.

We use three lags of each of these measures.

Agricultural Census

- **Food Price Shocks:** Using the 2002 Agriculture Census module of *Podes*, we calculate the quantity of each crop produced within the 2000 subdistrict borders. We use the UN Food and Agriculture price series for each crops to construct log changes in crop prices and weight these changes by the 2002 production share. We group into cash crops and food crops. We include three lags of each of these price shocks.

- **Agricultural GDP:** Using the 2002 Agriculture Census module of *Podes*, we calculate the quantity of each crop produced within the 2000 subdistrict borders and then use the UN Food and Agriculture price series to construct total agricultural value in each year.
- **Other:** We also include output shares of important cash and food crops and the share of households with any agricultural land and large agricultural land.

Night Lights

- **Light Intensity:** Annual night light data to proxy for GDP (Henderson et al. (2012)). We use mean stable light intensity at the village level, which ranges from 0 to 63. This attempts to filter out background noise and unstable sources of light. We compute the (population weighted) average light intensity across villages at the 2000 subdistrict boundary level. We include two lags of this measure.

Global Precipitation Climatology Project (GPCP)

The Global Precipitation Climatology Project (GPCP) provides annual rainfall at the district level. We calculate historical averages and annual deviations from the historical average.

University of Delaware Global Climate Resource Database (UDel)

The University of Delaware Global Climate Resource Database provides monthly rainfall and temperature data at the subdistrict level from 1900 onward. We calculate historical averages and annual deviations from the average. We include 3 lags of each of these deviation measures.

Mineral Shocks (SNL/GEM)

Geocoded mine data from the SNL Mine Production Data. For each mineral, we calculate the distance from the center of the subdistrict to the nearest active mine producing that mineral. We multiply log mineral price changes from the World Bank's GEM Commodity Price Database by the inverse of the distance to the nearest mine. We include three lags of these mineral price shocks.

Disaster Information Management System (DIMS)

The Disaster Information Management System lists the major disasters such as floods or volcano eruptions at the district level. We also use this data to construct event counts and deaths from terrorism and unrest. We include two lags of each of these disaster measures.

Database for Policy and Economic Research (DAPOER)

We use the World Bank's Indonesia Database for Policy and Economic Research (DAPOER), which in turn obtains data from the Indonesia Ministry of Finance data, to keep track of total **district** revenue in each year.

- **Total District Revenue Per Capita:** District revenue figures come from the World Bank's Indonesia Database for Policy and Economic Research (DAPOER), which in turn obtains data from the Indonesia ministry of finance data. They are given for each district-year at the time of existence. We aggregate up to the 2000 district boundary. Population data is taken from the same dataset. All figures are inflation adjusted using 2010 as the base year. We also construct district own-revenue, as opposed to revenue from shared rents, grants, and taxes.
- **GDP:** DAPOER also includes annual District level GDP and Agricultural GDP measures. We include three lags of each of these measures.

Political Data (GEC)

Our political data is gathered from documents published by the General Election Commission (GEC), used in Martinez-Bravo et al. (2017) and also provided to us by Audrey Sacks who collected this information from the GEC.

- **Vote Share Fractionalization/Polarization within 2000 subdistrict:** Data on vote share by party and subdistrict in the 1999 district parliamentary (DPRD-II) elections were used—the first of the post-Suharto era—to construct a measure of vote share polarization at the subdistrict level. Forty-eight parties competed in these elections.

- **Vote Shares:** We also use the 1999 parliamentary (DPRD-II) data to construct vote shares for major parties and Islamist parties. We also construct turnout as total votes over population.
- **Time-varying vote Share Fractionalization/Polarization at District Level:** To construct time-varying vote share fractionalization and polarization measures we use national parliamentary (DPR) vote share data in 1999, 2004, 2009 recorded at the district level.
- **Time-varying Party Shares at District Level:** Using this same national parliamentary vote data (DPR), we also retain the votes for certain parties. We consider Golkar and PDIP shares, as well as vote shares for Islamist parties. We also track the number of parties.
- **Direct Election Data:** Direct local elections for district mayors were phased in beginning 2005 across districts and occur every 5 years. We record the date of each of these elections.

We include measures from the latest election as our predictors.

Topographical Variables

- **Slope and Elevation Data**: Topographical variables were created using raster data from the *Harmonized World Soil Database* (HWSD), Version 2.0 (Fischer et al., 2008). The raster files are compiled from high-resolution source data and aggregated to 30 arc-second grids. The terrain, slope, and aspect database provided by HWSD researchers was compiled from a high-resolution digital elevation map constructed by the Shuttle Radar Topography Mission (SRTM).

  Elevation data were computed for each village as the average elevation over the entire village polygon, using raster data from HWSD. Slope and aspect data were also recorded for each village and calculated similarly. Variables equal to the average share of each village corresponding to each slope class (0-2 percent, 2-4 percent, etc.) were constructed using ArcView.
- **Ruggedness**: A 30 arc-second ruggedness raster was computed for Indonesia according to the methodology described by Sappington et al. (2007), and village-level ruggedness was recorded as the average raster value. The authors propose a Vector Ruggedness Measure

(VRM), which captures the distance or dispersion between a vector orthogonal to a topographical plane and the orthogonal vectors in a neighborhood of surrounding elevation planes.

To calculate the measure, one first calculates the x, y, and z coordinates of vectors that are orthogonal to each 30-arc second grid of the Earth's surface. These coordinates are computed using a digital elevation model and standard trigonometric techniques. Given this, a resultant vector is computed by adding a given cell's vector to each of the vectors in the surrounding cells; the neighborhood or window is supplied by the researcher. Finally, the magnitude of this resultant vector is divided by the size of the cell window and subtracted from 1. This results in a dimensionless number that ranges from 0 (least rugged) to 1 (most rugged).8.

- **Soil Quality**: We also make use of the HWSD data for soil quality measures. HWSD provides detailed information on different soil types across the world. The HWSD data for Indonesia is taken from information printed in the FAO-UNESCO Soil Map of the World (FAO 1971-1981), a map printed at a 1:5,000,000 scale. For each subdistrict, we we use the following measures of soil types: percentage of land covered by coarse, medium, and fine soils, percentage of land covered by soils with poor or excessive drainage, average organic carbon percentage, average soil salinity, average soil sodicity, and average topsoil pH.

While each of the above datasets covers the entire country, there are inevitably minor missing data issues as we combine so many sources. Rather than exclude entire predictors or observations, we impute these missing predictors via regression on contemporaneous predictors for which data is available. Violence measurements are not used for the imputation of other variables, nor are any violence measures imputed. The sample is restricted to observations where we have full violence data.

C.2 Colombia

C.2.1 Administrative Levels

The municipality is the second level of administrative authority in Colombia (the first is the Department) and is the fundamental territorial entity in the political-administrative division of the State. It has political, fiscal and administrative autonomy within the framework of the Colombian law.

As of 2015, Colombia has 1,101 municipalities in 32 departments. The departments are composed by municipalities and are also a territorial entity with administrative autonomy. They must perform administrative and coordination functions complementing the municipal action and should serve as intermediaries between the Nation and the municipalities.

C.2.2 Violence

Our main conflict data in Colombia comes from the Conflict Analysis Resource Center (CERAC) which contains data on military confrontation from 1988 to 2005. The data are reported at the event level and episodes are characterized either as bilateral clashes between sides or unilateral attacks from one side against another. We aggregate incidents by municipality-year in our main specification including events involving all three conflict actors: the guerrillas, the paramilitaries, or the government. However, we also consider a specification where the aggregation excludes government attacks or clashes (results available on request). Our main dependent variables are binary indicators for any event in a given municipality-year; more of than five events and more than 1 standard deviation increase in violence.

Table C.3 presents a descriptive analysis of the dependent variable (total onsets, any incident, more than five incidents or 'high', and greater than a 1 s.d. increase in violent events or 'spike') for Colombia. We cover the period between 1992 to 2005 at the municipality level and see that violence rates are rising over the period and then subsiding. Figure C.3 depicts the distribution of violent event counts in municipality-years that experience violence. We observe a significant right tail with a handful of municipalities experiencing dozens of events each year.

Appendix A.8 described an additional data source on violence in Colombia running through

2014.

C.2.3 Comparison to Other Violence Data

Similar to our coverage analysis of the SNPK data, the CERAC dataset offers a more comprehensive coverage of Colombian armed conflict event compared to other cross-country, subnational data sources. The UCDP-GED is available for Colombia since 1998 whereas the ACLED is only available beginning in 2019. After geolocating violent events in UCDP-GED to match them to Colombian municipalities, we find limited coverage of conflict events. In the period of 1998-2005, UCDP-GED covers 1,431 municipality-year incidents. 1,199 of these municipality-years also have an incident in CERAC. CERAC, on the other hand, has 1,933 municipality-years that have incidents but are not in UCDP-GED. These is equivalent to 7,247 casualities not recorded in UCDP-GED but present in CERAC.

Table C.3: Annual Rates of Conflict in Colombian Municipalities

	Any Conflict (1)	≥ 5 Conflict Incidents (2)	≥ 1 std. dev. increase in Incidents (3)
1992	0.352	0.081	0.039
1993	0.323	0.049	0.021
1994	0.340	0.063	0.043
1995	0.276	0.057	0.018
1996	0.317	0.066	0.042
1997	0.301	0.056	0.031
1998	0.375	0.079	0.045
1999	0.374	0.073	0.040
2000	0.417	0.106	0.073
2001	0.421	0.122	0.064
2002	0.454	0.145	0.082
2003	0.376	0.101	0.038
2004	0.335	0.099	0.053
2005	0.308	0.068	0.020

Notes: Conflict incidents above include paramilitary attacks, guerrilla attacks, government attacks, and bilateral clashes between these groups.

C.2.4 Covariates

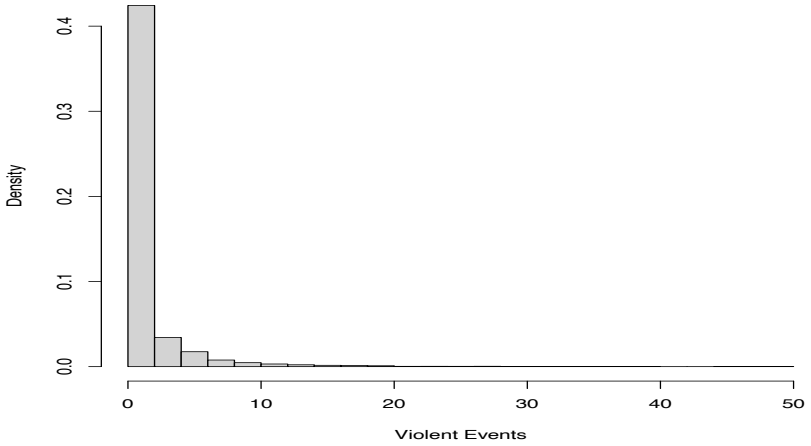
Centro de Estudios Sobre Desarrollo Económico (CEDE)

The Centro de Estudios Sobre Desarrollo Económico (CEDE) at the University of Los Andes maintains a warehouse of data that we use as a main source of municipality-level annual covariates. Many of the datasets below are found via CEDE.

National Administrative Department of Statistics (DANE)

Reported by CEDE, the National Administrative Department of Statistics (DANE) constructs projections of population in each municipality each year using Census data. We use this log population,

Figure C.3: Distribution of Violent Events, Colombia



Notes: This histogram depicts the distribution of violent event counts in subdistrict-years with at least one incident.

and we create population density measures as well as indicators that a given municipality has over 250,000 inhabitants and that the municipality is part of a larger metro area.

DANE also reports measures of deprivation based on the 1993 Census. These include the percent of households with unmet basic needs such as education and housing, a land ownership inequality index, and a life quality survey-based index.

CEDE also provides measures of paved roads, unpaved roads, and rivers in 1995, as well as measures slope, elevation and ruggedness.

National Planning Department (NPD)

Again, via CEDE, fiscal revenue and spending data from the National Planning Department (NPD) are collected. Specifically, information on municipal income, spending, deficit and transfers. We include three lags of these measures.

Minisry of Agriculture

The Ministry of Agriculture supplies data on the suitability of areas for various crops. We use data on the suitability of lands for sugar cane and palm oil production.

Commodities and International Prices (Dube and Vargas, 2013)

We use commodity data compiled by (Dube and Vargas, 2013). We construct shocks as the interaction between the log commodity price change and measures of local commodity production at a fixed point in time. We include three lags of each of these commodity price shocks.

- Internal **coffee** prices are tabulated by the Colombian Coffee Growers Federation (NCFG), and hectares of coffee cultivation is measured in the 1997 coffee growers' survey.
- Data on 1988 **oil** production comes from the Ministerio de Minas y Energia (MME). The international price of oil comes from the IMF's International Financial Statistics.
- **Coal** reserve presence is reported by a 1978 U.S. Geological Survey. The international price of coal comes from Global Financial Data.
- **Precious metal** presence is measured by 1978 mining applications to the MME. An hectares of minable land are interacted with **gold**, **silver**, and **platinum** prices from Global Financial Data.

We consider both the local production and international price of coffee, oil, coal, silver, platinum, and precious metals.

Dube and Vargas (2013) obtain data on coca cultivation for Colombia from two sources: Direccion Nacional de Estupefacientes (DNE) and from the United Nations Office of Drug Control (UNODC). DNE reports illicit drug cultivation in 1994. The UNODC uses satalite imagery to estimate annual coca production in Colombia. UNODC also reports U.S. cocaine prices. We interact log changes in U.S. cocaine prices with coca production in 1994. We include three lags of each of these time-varying drug production variables.

Historical Variables Acemoglu et al. (2015)

From Acemoglu et al. (2015), we include measures of colonial institutions and infrastructure, such as number of crown employees, presence of colonial cities and royal roads, population in 1843, slave share of the population in 1843, number of Indians in 1560, number of *encomiendas* in 1560, colonial gold mines, foundation dates, and population in 1843.

University of Delaware Global Climate Resource Database (UDel)

The University of Delaware Global Climate Resource Database provides monthly rainfall and temperature data at the subdistrict level from 1900 onward. We calculate historical averages and annual deviations from the average. We include three lags of each of these weather deviations.

Registraduria Nacional del Estado Civil ("National Registry")

Information on mayoral and congressional elections (lower and upper house) were used from the entire period of analysis. Universidad de los Andes compiled a database of electoral results since 1958 and has been updating it until 2014. The original data comes from the Registraduria Nacional del Estado Civil ("National Registry")[1]. In particular, we generate parties' vote shares, turnout, and time dummies for electoral periods.

Each of this elections represent different levels of political power in Colombia, at the local, regional and national level. We consider measures of concentration, polarization and fractionalization index for competitiveness of the elections, margin between winner and runner-up, party's vote share and political leaning. We include measures from the latest elections as predictors.

Distance to DMZ (Delimitarized Zone)

The Caguan DMZ was a delimitarized zone of 42,000 square kilometers in southern Colombia authorized by the government to negotiate a peace process with the FARC-EP. The region was made up by the municipalities of Vista Hermosa, La Macarena, La Uribe, Mesetas, and San Vicente del Caguan. The DMZ started in January of 1999 and ended in February 2002. Its existence coincides with the escalation of the conflict in Colombia. Therefore, we calculate the distance of each municipality in Colombia to the DMZ as a covariate.

US Military Aid

We use the dataset created by Dube and Naidu (2015) of US military aid and Colombian military bases. One feature of US military aid is that it is disbursed to particular Colombian military

[1]Universidad de los Andes CEDE makes the database publicly available through its' database website (https://datoscede.uniandes.edu.co).

brigades, each of which is attached to and operates out of a particular government military base. We consider the natural log of US military and antinarcotics aid to Colombia interacted with a dummy that defines if a particular municipality has a military base. In total, we covered 34 municipalities with military bases, of which 32 appear in the sample for which the conflict data is available. We include three lags of the military aid variable.

As in Indonesia, we restrict our sample to observations with full violence data. There are inevitably minor missing data issues as we merge other covariates. Rather than exclude entire predictors or observations, we impute these missing predictors via regression on contemporaneous predictors for which data is available. Violence measurements are not used for the imputation of other variables, nor are any violence measures imputed.

C.3 Predictor Groupings

C.3.1 Indonesia

Table C.4: Violence Predictors (Indonesia): SNPK

	(1) Mean	(2) Median	(3) 25th Pct.	(4) 75th Pct.	(5) Corr. w/ Any	(6) Within/ Between Var.	(7) % Miss	(8) Source
Incidents per 10,000, social conflict	2.61	0.14	0.00	0.50	0.05	2.31	0.0	SNPK
Any social conflict	0.53	1.00	0.00	1.00	0.38	1.17	0.0	SNPK
More than 5 social conflict incidents	0.13	0.00	0.00	0.00	0.32	0.85	0.0	SNPK
More than 1 s.d. increase in social conflict incidents	0.03	0.00	0.00	0.00	0.15	1.88	0.0	SNPK
Number of social conflicts	2.18	1.00	0.00	2.00	0.31	0.66	0.0	SNPK
Deaths per 10,000, social conflict	2.15	0.00	0.00	0.00	0.03	2.88	0.0	SNPK
Injuries per 10,000, social conflict	1.45	0.00	0.00	0.45	0.05	2.07	0.0	SNPK
Destroyed buildings per 10,000, social conflict	0.19	0.00	0.00	0.00	0.01	3.13	0.0	SNPK
Incidents per 10,000, ALL	4.50	0.73	0.21	1.68	0.06	1.72	0.0	SNPK
Incidents per 10,000, Resource	0.08	0.00	0.00	0.00	0.02	2.32	0.0	SNPK
Incidents per 10,000, Governance	0.05	0.00	0.00	0.00	0.04	1.91	0.0	SNPK
Incidents per 10,000, Election	0.09	0.00	0.00	0.00	0.03	2.13	0.0	SNPK
Incidents per 10,000, Identity	0.06	0.00	0.00	0.00	0.03	1.28	0.0	SNPK
Incidents per 10,000, Popular Justice	0.34	0.00	0.00	0.19	0.05	1.53	0.0	SNPK
Incidents per 10,000, Law Enforcement	0.15	0.00	0.00	0.00	0.03	1.96	0.0	SNPK
Incidents per 10,000, Crime	1.69	0.42	0.00	1.03	0.07	1.04	0.0	SNPK
Incidents per 10,000, Domestic	0.20	0.00	0.00	0.12	0.05	1.61	0.0	SNPK
Incidents per 10,000, Separatist	1.78	0.00	0.00	0.00	0.04	3.04	0.0	SNPK
Incidents per 10,000, Other	0.06	0.00	0.00	0.00	0.02	2.59	0.0	SNPK
Deaths per 10,000, ALL	2.63	0.00	0.00	0.24	0.03	2.70	0.0	SNPK
Deaths per 10,000, Resource	0.01	0.00	0.00	0.00	0.00	3.26	0.0	SNPK
Deaths per 10,000, Governance	0.00	0.00	0.00	0.00	0.01	3.50	0.0	SNPK
Deaths per 10,000, Election	0.01	0.00	0.00	0.00	0.01	3.52	0.0	SNPK
Deaths per 10,000, Identity	0.03	0.00	0.00	0.00	0.02	1.57	0.0	SNPK
Deaths per 10,000, Popular Justice	0.03	0.00	0.00	0.00	0.01	3.47	0.0	SNPK
Deaths per 10,000, Law Enforcement	0.02	0.00	0.00	0.00	-0.00	3.49	0.0	SNPK
Deaths per 10,000, Crime	0.44	0.00	0.00	0.12	0.03	2.26	0.0	SNPK
Deaths per 10,000, Domestic	0.04	0.00	0.00	0.00	0.02	3.14	0.0	SNPK
Deaths per 10,000, Separatist	2.04	0.00	0.00	0.00	0.03	2.92	0.0	SNPK
Deaths per 10,000, Other	0.01	0.00	0.00	0.00	0.01	3.52	0.0	SNPK
Injuries per 10,000, ALL	2.31	0.37	0.00	1.13	0.07	1.56	0.0	SNPK
Injuries per 10,000, Resource	0.06	0.00	0.00	0.00	0.04	2.97	0.0	SNPK
Injuries per 10,000, Governance	0.04	0.00	0.00	0.00	0.01	3.25	0.0	SNPK
Injuries per 10,000, Election	0.07	0.00	0.00	0.00	0.01	2.82	0.0	SNPK
Injuries per 10,000, Identity	0.06	0.00	0.00	0.00	0.04	1.72	0.0	SNPK
Injuries per 10,000, Popular Justice	0.41	0.00	0.00	0.17	0.05	1.69	0.0	SNPK
Injuries per 10,000, Law Enforcement	0.16	0.00	0.00	0.00	0.02	2.61	0.0	SNPK
Injuries per 10,000, Crime	0.75	0.00	0.00	0.52	0.08	1.28	0.0	SNPK
Injuries per 10,000, Domestic	0.11	0.00	0.00	0.00	0.05	2.21	0.0	SNPK
Injuries per 10,000, Separatist	0.58	0.00	0.00	0.00	0.03	3.37	0.0	SNPK
Injuries per 10,000, Other	0.08	0.00	0.00	0.00	0.01	3.51	0.0	SNPK
Damaged buildings per 10,000, ALL	0.62	0.00	0.00	0.00	0.01	3.02	0.0	SNPK
Damaged buildings per 10,000, Resource	0.04	0.00	0.00	0.00	0.00	3.15	0.0	SNPK
Damaged buildings per 10,000, Governance	0.02	0.00	0.00	0.00	0.02	2.88	0.0	SNPK
Damaged buildings per 10,000, Election	0.02	0.00	0.00	0.00	0.02	3.05	0.0	SNPK
Damaged buildings per 10,000, Identity	0.05	0.00	0.00	0.00	0.01	2.73	0.0	SNPK
Damaged buildings per 10,000, Popular Justice	0.03	0.00	0.00	0.00	0.02	2.56	0.0	SNPK
Damaged buildings per 10,000, Law Enforcement	0.00	0.00	0.00	0.00	-0.01	3.51	0.0	SNPK
Damaged buildings per 10,000, Crime	0.11	0.00	0.00	0.00	0.04	2.63	0.0	SNPK
Damaged buildings per 10,000, Domestic	0.00	0.00	0.00	0.00	0.01	3.48	0.0	SNPK
Damaged buildings per 10,000, Separatist	0.33	0.00	0.00	0.00	0.01	3.43	0.0	SNPK
Damaged buildings per 10,000, Other	0.03	0.00	0.00	0.00	0.01	3.45	0.0	SNPK
Destroyed buildings per 10,000, ALL	0.23	0.00	0.00	0.00	0.02	2.99	0.0	SNPK
Destroyed buildings per 10,000, Resource	0.02	0.00	0.00	0.00	-0.00	3.17	0.0	SNPK
Destroyed buildings per 10,000, Governance	0.00	0.00	0.00	0.00	0.01	3.52	0.0	SNPK
Destroyed buildings per 10,000, Election	0.01	0.00	0.00	0.00	0.01	3.37	0.0	SNPK
Destroyed buildings per 10,000, Identity	0.02	0.00	0.00	0.00	-0.00	3.29	0.0	SNPK
Destroyed buildings per 10,000, Popular Justice	0.01	0.00	0.00	0.00	0.01	3.17	0.0	SNPK
Destroyed buildings per 10,000, Law Enforcement	0.00	0.00	0.00	0.00	-0.01	3.52	0.0	SNPK
Destroyed buildings per 10,000, Crime	0.04	0.00	0.00	0.00	0.02	3.07	0.0	SNPK
Destroyed buildings per 10,000, Domestic	0.00	0.00	0.00	0.00	-0.00	3.49	0.0	SNPK
Destroyed buildings per 10,000, Separatist	0.13	0.00	0.00	0.00	0.01	3.19	0.0	SNPK
Destroyed buildings per 10,000, Other	0.00	0.00	0.00	0.00	0.01	3.51	0.0	SNPK
Total active papers, SNPK	51.46	20.79	10.60	33.79	0.00	1.08	0.0	SNPK

Table C.5: Violence Predictors (Indonesia): *Podes*

	(1) Mean	(2) Median	(3) 25th Pct.	(4) 75th Pct.	(5) Corr. w/ Any	(6) Within/ Between Var.	(7) % Miss	(8) Source
Theft, % villages	0.04	0.02	0.01	0.05	0.15	0.53	0.0	Podes
Robbery, % villages	0.00	0.00	0.00	0.00	0.13	0.86	0.0	Podes
Looting incident, % villages	0.00	0.00	0.00	0.00	0.04	0.83	0.0	Podes
Thuggery, % villages	0.01	0.00	0.00	0.01	0.10	0.94	0.0	Podes
Arson, % villages	0.00	0.00	0.00	0.00	0.00	1.54	0.0	Podes
Rape, % villages	0.00	0.00	0.00	0.00	0.08	0.99	0.0	Podes
Murder, % villages	0.00	0.00	0.00	0.00	0.10	1.15	0.0	Podes
Fight among citizens, % villages	0.00	0.00	0.00	0.00	0.06	1.20	0.0	Podes
Fights with security officers, % villages	0.00	0.00	0.00	0.00	0.04	1.20	0.0	Podes
Conflict among students, % villages	0.00	0.00	0.00	0.00	0.06	1.07	0.0	Podes
Conflict between tribes, % villages	0.00	0.00	0.00	0.00	0.02	1.33	0.0	Podes
Indicator for any conflict, % villages	0.00	0.00	0.00	0.00	0.08	1.09	0.0	Podes
Deaths per 10000	0.02	0.00	0.00	0.00	0.01	1.63	0.0	Podes
Injuries per 10000	0.21	0.00	0.00	0.00	0.00	1.39	0.0	Podes
Security Post Present,% villages	0.79	0.95	0.67	1.00	0.14	0.53	0.0	Podes
Police Post Present,% villages	0.15	0.10	0.06	0.17	0.17	0.45	0.0	Podes
Number of security officers	283.42	234.00	117.00	402.00	0.14	0.36	0.0	Podes
Terrorism (Count)	0.00	0.00	0.00	0.00	0.02	2.60	0.0	DIMS
Terrorism (Deaths)	0.00	0.00	0.00	0.00	0.01	2.83	0.0	DIMS

Table C.6: Violence Predictors (Indonesia): DIMS

	(1) Mean	(2) Median	(3) 25th Pct.	(4) 75th Pct.	(5) Corr. w/ Any	(6) Within/ Between Var.	(7) % Miss	(8) Source
Terrorism (Count)	0.00	0.00	0.00	0.00	0.02	2.60	0.0	DIMS
Terrorism (Deaths)	0.00	0.00	0.00	0.00	0.01	2.83	0.0	DIMS
Unrest (Count)	0.01	0.00	0.00	0.00	0.03	2.70	0.0	DIMS
Unrest (Deaths)	0.01	0.00	0.00	0.00	-0.01	3.48	0.0	DIMS

Table C.7: Population Predictors (Indonesia)

	(1) Mean	(2) Median	(3) 25th Pct.	(4) 75th Pct.	(5) Corr. w/ Any	(6) Within/ Between Var.	(7) % Miss	(8) Source
Population (ln)	10.41	10.54	9.93	10.99	0.27	0.28	2.3	Census
Population (ln)	10.63	10.65	10.13	11.10	0.42	0.17	0.0	Podes
Number of villages	18.76	15.00	10.00	22.00	0.07	0.02	0.7	Census
Population growth (ln ch.), change over last 2 Podes	0.05	0.04	0.00	0.10	0.06	1.81	0.0	Podes
Number of villages	19.79	16.00	11.00	23.00	0.06	0.25	0.0	Podes

Table C.8: Religion Predictors (Indonesia)

	(1) Mean	(2) Median	(3) 25th Pct.	(4) 75th Pct.	(5) Corr. w/ Any	(6) Within/ Between Var.	(7) % Miss	(8) Source
Religious Fractionalization	0.15	0.04	0.00	0.27	-0.03	0.08	2.3	Census
Religious Polarization	0.07	0.02	0.00	0.13	-0.02	0.07	2.3	Census
Muslim share	0.75	0.95	0.59	1.00	0.18	0.10	2.3	Census
Catholic share	0.08	0.01	0.00	0.04	-0.14	0.19	2.3	Census
Protestant share	0.14	0.01	0.00	0.13	-0.11	0.10	2.3	Census
Hindu share	0.01	0.00	0.00	0.00	-0.06	0.59	2.3	Census
Buddhist share	0.01	0.00	0.00	0.00	0.11	0.60	2.3	Census
Confucian share	0.01	0.00	0.00	0.00	-0.08	0.54	2.3	Census
Christian share	0.22	0.03	0.00	0.29	-0.17	0.10	2.3	Census
Any Islamic Boarding School, % villages	0.24	0.12	0.00	0.41	0.19	0.02	0.7	Podes

Table C.9: Ethnicity Predictors (Indonesia)

	(1) Mean	(2) Median	(3) 25th Pct.	(4) 75th Pct.	(5) Corr. w/ Any	(6) Within/ Between Var.	(7) % Miss	(8) Source
Ethnic Fractionalization	0.30	0.15	0.02	0.58	0.08	0.15	2.3	Census
Ethnic Polarization	0.09	0.07	0.01	0.16	0.08	0.12	2.3	Census
Ethnic Segregation (village-kecamatan comparison)	0.43	0.53	0.00	0.85	0.08	0.12	2.3	Census
Chinese share	0.01	0.00	0.00	0.00	0.11	0.60	2.3	Census
Arab share	0.00	0.00	0.00	0.00	0.08	0.18	2.3	Census
Interethnic Marriage, share	0.19	0.12	0.03	0.30	0.21	0.14	2.3	Census

Table C.10: Education Predictors (Indonesia)

	(1) Mean	(2) Median	(3) 25th Pct.	(4) 75th Pct.	(5) Corr. w/ Any	(6) Within/ Between Var.	(7) % Miss	(8) Source
Mean years of schooling	7.96	7.76	7.29	8.43	0.35	0.15	2.3	Census
No school share	0.35	0.34	0.27	0.42	-0.26	0.12	2.3	Census
Kindergarten present, per thsd.	0.40	0.36	0.19	0.55	-0.03	0.62	0.0	Podes
Elementary School present, per thsd.	0.97	0.89	0.67	1.19	-0.37	0.28	0.0	Podes
High School present, per thsd.	0.08	0.07	0.04	0.11	0.09	0.61	0.0	Podes
Madrasa present, per thsd.	0.21	0.07	0.00	0.26	0.04	0.63	0.0	Podes
University present, per thsd.	0.01	0.00	0.00	0.01	0.19	0.72	0.0	Podes
In school, % school age	0.70	0.70	0.66	0.74	0.12	0.75	0.0	Susenas
Literate, % population over 15yo	0.90	0.92	0.86	0.96	0.13	0.28	0.0	Susenas

Table C.11: Remoteness Predictors (Indonesia)

	(1) Mean	(2) Median	(3) 25th Pct.	(4) 75th Pct.	(5) Corr. w/ Any	(6) Within/ Between Var.	(7) % Miss	(8) Source
Rural population share	0.78	0.94	0.67	1.00	-0.37	0.12	2.3	Census
Main Road Electrified, % villages	0.61	0.71	0.24	1.00	0.26	0.03	0.7	Census
Road Access Year Round, % villages	1.05	1.00	1.00	1.09	-0.08	0.04	0.7	Census
Distance to province capital	85.39	69.55	36.19	115.97	-0.23	0.00	0.0	Podes
Distance to district capital	47.59	27.63	12.78	55.01	-0.23	0.57	0.0	Podes
Bridge present, % villages	0.54	0.57	0.29	0.82	0.05	0.13	3.9	Podes
Bus terminal present, % villages	0.05	0.00	0.00	0.07	0.17	0.23	3.9	Podes
Seaport present, % villages	0.04	0.00	0.00	0.01	-0.05	0.29	3.9	Podes
Land area	32805.15	11856.30	4722.80	34865.00	-0.09	0.15	3.9	Podes
Light Transport present, % villages	0.46	0.40	0.09	0.83	0.27	0.54	0.0	Podes
Road Pass present, % villages	0.88	1.00	0.86	1.00	0.25	0.40	0.0	Podes

Table C.12: Demographics Predictors (Indonesia)

	(1) Mean	(2) Median	(3) 25th Pct.	(4) 75th Pct.	(5) Corr. w/ Any	(6) Within/ Between Var.	(7) % Miss	(8) Source
Male share	0.50	0.50	0.49	0.51	-0.00	0.21	2.3	Census
Married share	0.44	0.44	0.40	0.48	-0.03	0.20	2.3	Census
Share of population under 10	0.02	0.02	0.02	0.02	-0.23	0.16	2.3	Census
Share of population from 10 to 30	0.04	0.04	0.03	0.04	-0.02	0.29	2.3	Census
Average age	28.66	27.85	26.44	30.67	0.02	0.19	0.0	Susenas

Table C.13: Health Care Predictors (Indonesia)

	(1) Mean	(2) Median	(3) 25th Pct.	(4) 75th Pct.	(5) Corr. w/ Any	(6) Within/ Between Var.	(7) % Miss	(8) Source
Hospital present, per thsd.	0.00	0.00	0.00	0.00	0.15	0.91	0.0	Podes
Maternal Hospital present, per thsd.	0.01	0.00	0.00	0.00	0.14	1.13	0.0	Podes
Medical School Polytechnic present, per thsd.	0.03	0.00	0.00	0.03	0.09	1.27	0.0	Podes
Puskesmas present, per thsd.	0.04	0.03	0.00	0.05	-0.17	1.16	0.0	Podes
Pustu present, per thsd.	0.13	0.06	0.01	0.17	-0.20	1.03	0.0	Podes
Doctor present, per thsd.	0.15	0.08	0.04	0.17	0.19	0.51	0.0	Podes
MidWife present, per thsd.	0.56	0.43	0.26	0.71	-0.05	0.79	0.0	Podes
Traditional Birth present, per thsd.	1.06	0.66	0.25	1.35	-0.25	0.80	0.0	Podes
Health problem, % children	0.41	0.41	0.33	0.48	0.03	0.89	0.0	Susenas
Serious health problem, % children	0.27	0.26	0.21	0.32	-0.02	1.02	0.0	Susenas
Health problem, % >5 year old	0.29	0.28	0.23	0.34	0.01	0.92	0.0	Susenas
Serious health problem, % >5 year old	0.17	0.16	0.12	0.20	-0.06	0.92	0.0	Susenas

Table C.14: Geography Predictors (Indonesia)

	(1) Mean	(2) Median	(3) 25th Pct.	(4) 75th Pct.	(5) Corr. w/ Any	(6) Within/ Between Var.	(7) % Miss	(8) Source
North facing slope, percent	16.96	16.76	8.05	23.34	-0.14	0.03	0.2	HWSD
East facing slope, percent	16.82	16.89	8.28	22.96	-0.14	0.01	0.2	HWSD
South facing slope, percent	15.59	15.17	6.50	22.42	-0.18	0.02	0.2	HWSD
Slope below 2%, percent	33.71	23.10	6.51	58.37	0.22	0.02	0.2	HWSD
West facting slope, percent	16.35	16.46	7.95	23.18	-0.17	0.02	0.2	HWSD
Elevation, meters	234.54	109.94	34.32	308.11	-0.13	0.01	0.2	HWSD
Forest coverage, percent	28.38	26.04	10.20	43.39	-0.19	0.09	19.1	HWSD
Grass and scrub coverage, percent	18.77	19.27	11.29	25.25	-0.24	0.12	19.1	HWSD
Vector ruggedness, 3x3 window	0.23	0.24	0.13	0.32	-0.13	0.03	0.3	HWSD
Vector ruggedness, 5x5 window	0.34	0.36	0.20	0.48	-0.15	0.04	0.8	HWSD
Vector ruggedness, 9x9 window	0.45	0.49	0.29	0.63	-0.18	0.11	4.2	HWSD
Slope between 0-0.5%, percent	5.33	2.25	0.48	8.06	0.17	0.03	0.2	HWSD
Slope between 0.5-2%, percent	28.38	20.79	5.98	50.36	0.22	0.02	0.2	HWSD
Slope between 2-5%, percent	22.83	20.48	10.69	33.64	0.09	0.02	0.2	HWSD
Slope between 5-8%, percent	13.46	11.84	4.15	21.28	-0.13	0.02	0.2	HWSD
Slope between 8-16%, percent	7.43	6.32	0.36	12.80	-0.20	0.01	0.2	HWSD
Slope between 16-30%, percent	12.74	7.10	0.02	24.37	-0.21	0.02	0.2	HWSD
Slope between 30-45%, percent	5.93	1.31	0.00	9.76	-0.16	0.02	0.2	HWSD
Slope above 45%, percent	3.33	0.20	0.00	4.15	-0.10	0.03	0.2	HWSD
Nutrient availability	1.74	1.33	1.00	2.69	-0.03	0.11	19.1	HWSD
Nutrient retention capacity	1.49	1.32	1.00	2.00	-0.05	0.13	19.1	HWSD
Favorable rooting conditions	1.28	1.00	1.00	1.34	-0.10	0.16	19.1	HWSD
Oxygen availability to roots	1.18	1.00	1.00	1.23	-0.00	0.17	19.1	HWSD
Excess salts	0.97	1.00	1.00	1.00	-0.04	0.19	19.1	HWSD
Toxicity	0.97	1.00	1.00	1.00	-0.04	0.19	19.1	HWSD
Workability	1.54	1.04	1.00	2.00	-0.07	0.15	19.1	HWSD
Organic Carbon (%)	2.47	1.67	1.30	2.81	-0.02	0.03	0.4	HWSD
Topsoil Salinity	0.11	0.10	0.10	0.10	-0.07	0.03	0.4	HWSD
Topsoil Sodicity	1.47	1.40	1.25	1.65	0.12	0.03	0.4	HWSD
Topsoil pH	5.88	6.01	5.07	6.63	-0.01	0.03	0.4	HWSD
Topsoil Gypsium	0.00	0.00	0.00	0.00	0.01	0.00	0.4	HWSD
Topsoil Base Saturation	60.98	54.84	43.98	75.47	0.02	0.04	0.4	HWSD
Coarse texture soil, percent	0.07	0.00	0.00	0.04	-0.04	0.02	0.4	HWSD
Medium texture soils, percent	0.59	0.62	0.49	0.75	0.00	0.04	0.4	HWSD
Fine texture soils, percent	0.34	0.30	0.17	0.49	0.03	0.04	0.4	HWSD
Soil w/ very poor drainage, percent	0.09	0.00	0.00	0.17	0.10	0.04	0.4	HWSD
Soil w/ poor drainage, percent	0.19	0.12	0.02	0.35	0.11	0.03	0.4	HWSD
Soil w/ imperfect drainage, percent	0.13	0.01	0.00	0.14	-0.10	0.03	0.4	HWSD
Soil w/ moderately good drainage, percent	0.53	0.50	0.30	0.80	-0.03	0.04	0.4	HWSD
Soil w/good drainage, percent	0.02	0.00	0.00	0.00	0.04	0.02	0.4	HWSD
Soil w/ somewhat excessive drainage, percent	0.04	0.00	0.00	0.00	-0.07	0.02	0.4	HWSD

Table C.15: Public Goods Predictors (Indonesia)

	(1) Mean	(2) Median	(3) 25th Pct.	(4) 75th Pct.	(5) Corr. w/ Any	(6) Within/ Between Var.	(7) % Miss	(8) Source
Permanent Market Present, % villages	0.18	0.13	0.07	0.25	0.14	0.04	0.7	Podes
Safe water present, % villages	0.15	0.00	0.00	0.16	0.23	0.30	0.0	Podes
Garbage disposal present, % villages	0.13	0.00	0.00	0.09	0.31	0.23	0.0	Podes
Toilets present, % villages	0.69	0.82	0.43	1.00	0.22	0.67	0.0	Podes
Cooking facilities present, % villages	0.16	0.00	0.00	0.08	0.17	1.20	0.0	Podes
Has electricity, % households	0.64	0.72	0.43	0.89	0.30	0.44	0.0	Podes
Without electrictiy, % households	0.10	0.03	0.00	0.14	-0.11	1.30	0.0	Podes

Table C.16: Sector Shares Predictors (Indonesia)

	(1) Mean	(2) Median	(3) 25th Pct.	(4) 75th Pct.	(5) Corr. w/ Any	(6) Within/ Between Var.	(7) % Miss	(8) Source
Trading share	0.04	0.03	0.02	0.06	0.31	0.13	2.3	Census
Self-employed share	0.14	0.13	0.09	0.18	-0.07	0.14	2.3	Census
Agriculture output share	0.29	0.31	0.18	0.42	-0.38	0.12	2.3	Census
Services output share	0.07	0.05	0.03	0.09	0.36	0.16	2.3	Census
Employer share	0.09	0.07	0.03	0.14	-0.24	0.21	2.3	Census
Population share in agriculture	0.21	0.20	0.09	0.32	-0.31	0.15	2.3	Census
Population share in forestry	0.08	0.06	0.02	0.11	-0.18	0.22	2.3	Census
Population share in industry	0.02	0.01	0.00	0.03	0.17	0.30	2.3	Census
Population share in trade	0.04	0.03	0.02	0.06	0.31	0.13	2.3	Census
Population share in service	0.06	0.04	0.02	0.08	0.35	0.16	2.3	Census
Population share in transport	0.01	0.01	0.00	0.01	0.32	0.18	2.3	Census

Table C.17: Distance to Mines Predictors (Indonesia)

	(1) Mean	(2) Median	(3) 25th Pct.	(4) 75th Pct.	(5) Corr. w/ Any	(6) Within/ Between Var.	(7) % Miss	(8) Source
Distance to nearest Bauxite mine	632.65	611.33	479.02	802.41	0.06	0.00	0.0	SNL
Distance to nearest Coal mine	152.17	107.38	59.21	163.82	-0.02	0.00	0.0	SNL
Distance to nearest Gold mine	58.98	46.41	26.61	72.96	-0.06	0.00	0.0	SNL
Distance to nearest IronOre mine	125.14	88.66	62.74	146.54	-0.08	0.00	0.0	SNL
Distance to nearest Nickel mine	412.61	320.62	196.41	434.30	0.06	0.00	0.0	SNL
Distance to nearest Tin mine	519.98	433.28	251.94	787.78	-0.04	0.00	0.0	SNL
Distance to nearest Zinc mine	311.11	295.87	122.64	460.07	-0.07	0.00	0.0	SNL
Distance to nearest Silver mine	640.47	545.76	446.26	947.37	0.10	0.00	0.0	SNL

Table C.18: Agricultural Features Predictors (Indonesia)

	(1) Mean	(2) Median	(3) 25th Pct.	(4) 75th Pct.	(5) Corr. w/ Any	(6) Within/ Between Var.	(7) % Miss	(8) Source
Access to Irrigation, % households	0.38	0.28	0.00	0.73	-0.00	0.11	2.8	Podes
Has small rice mill, % villages	0.18	0.06	0.00	0.25	-0.11	0.04	0.7	Podes
Has large rice mill, % villages	0.06	0.00	0.00	0.05	0.03	0.02	0.7	Podes
Major cash crop output share	0.32	0.15	0.01	0.61	-0.16	0.03	0.7	Podes
Major crop output share	0.53	0.57	0.16	0.90	-0.14	0.03	0.7	Podes
Major food crop output share	0.38	0.30	0.03	0.70	0.02	0.03	0.7	Podes
Share of HHs with any ag land	0.64	0.73	0.45	0.89	-0.39	0.19	9.8	Podes
Share of HHs with Ag land >0.1 Ha	0.59	0.65	0.36	0.85	-0.37	0.20	10.6	Podes
Rice is primary commodity, % villages	0.40	0.33	0.00	0.75	-0.08	0.04	0.7	Podes

Table C.19: Communication Predictors (Indonesia)

	(1) Mean	(2) Median	(3) 25th Pct.	(4) 75th Pct.	(5) Corr. w/ Any	(6) Within/ Between Var.	(7) % Miss	(8) Source
Telephone present, % respondents	0.07	0.04	0.02	0.08	0.27	0.46	0.0	Susenas
Cell phone present, % respondents	0.53	0.58	0.27	0.79	0.16	1.61	0.0	Susenas
Computer present, % respondents	0.07	0.05	0.02	0.09	0.25	0.67	0.0	Susenas
Accessed internet in past 3 months, % respondents	0.07	0.05	0.02	0.10	0.26	0.69	0.0	Susenas
Telephone, % households	0.00	0.00	0.00	0.00	0.17	0.50	0.0	Podes
Public TV present, % villages	0.05	0.04	0.00	0.07	0.19	0.45	0.0	Podes
Phone Stall Present, % villages	0.03	0.01	0.00	0.04	0.23	0.46	0.0	Podes
Postal access, % villages	0.02	0.01	0.00	0.02	0.15	0.51	0.0	Podes
Strong cell signal present, % villages	0.05	0.04	0.01	0.07	0.21	0.33	0.0	Podes
Weak cell signal present, % villages	0.02	0.01	0.00	0.02	-0.19	0.85	0.0	Podes
Cell tower present, % villages	0.03	0.02	0.01	0.04	0.21	0.00	0.0	Podes
Radio, % households	0.01	0.00	0.00	0.01	-0.05	0.45	3.9	Podes
Satellite dish, % households	0.04	0.02	0.00	0.05	-0.00	0.20	3.9	Podes
Television, % households	0.27	0.21	0.09	0.40	0.31	0.25	3.9	Podes

Table C.20: Natural Disaster Predictors (Indonesia)

	(1) Mean	(2) Median	(3) 25th Pct.	(4) 75th Pct.	(5) Corr. w/ Any	(6) Within/ Between Var.	(7) % Miss	(8) Source
Mudslide, % villages	0.08	0.00	0.00	0.08	-0.11	0.05	0.7	Podes
Flood, % villages	0.21	0.10	0.00	0.35	-0.02	0.04	0.7	Podes
Earthquake, % villages	0.11	0.00	0.00	0.00	0.03	0.06	0.7	Podes
Fire, % villages	0.12	0.00	0.00	0.14	-0.06	0.03	0.7	Podes
Other disaster, % villages	0.06	0.00	0.00	0.04	-0.08	0.03	0.7	Podes
Land abrasions (Count)	0.03	0.00	0.00	0.00	0.03	2.25	0.0	DIMS
Land abrasions (Deaths)	0.00	0.00	0.00	0.00	-0.03	2.89	0.0	DIMS
Disease outbreaks (Count)	0.01	0.00	0.00	0.00	-0.00	2.76	0.0	DIMS
Disease outbreaks (Deaths)	0.01	0.00	0.00	0.00	-0.02	3.56	0.0	DIMS
Droughts (Count)	0.20	0.00	0.00	0.00	-0.01	2.28	0.0	DIMS
Earthquakes (Count)	0.03	0.00	0.00	0.00	0.01	2.45	0.0	DIMS
Earthquakes (Deaths)	56.43	0.00	0.00	0.00	0.01	3.52	0.0	DIMS
Floods (Count)	0.83	0.00	0.00	1.00	0.06	1.48	0.0	DIMS
Floods (Deaths)	0.22	0.00	0.00	0.00	-0.01	2.64	0.0	DIMS
Forest Fires (Count)	0.01	0.00	0.00	0.00	-0.01	3.32	0.0	DIMS
Industrial Accidents (Count)	0.01	0.00	0.00	0.00	0.02	2.07	0.0	DIMS
Industrial Accidents (Deaths)	0.00	0.00	0.00	0.00	0.02	2.55	0.0	DIMS
Tornadoes (Count)	0.42	0.00	0.00	0.00	0.01	1.87	0.0	DIMS
Tornadoes (Deaths)	0.02	0.00	0.00	0.00	-0.01	3.54	0.0	DIMS
Transport accidents (Count)	0.02	0.00	0.00	0.00	-0.00	2.76	0.0	DIMS
Transport accidents (Deaths)	0.06	0.00	0.00	0.00	-0.03	2.52	0.0	DIMS
Tsunami (Count)	0.04	0.00	0.00	0.00	0.03	2.33	0.0	DIMS
Tsunami (Deaths)	56.39	0.00	0.00	0.00	0.01	3.52	0.0	DIMS
Volcanoes (Count)	0.01	0.00	0.00	0.00	0.00	2.48	0.0	DIMS
Deaths from any disaster	113.27	0.00	0.00	0.00	0.01	3.52	0.0	DIMS
Total number of disasters	1.88	1.00	0.00	3.00	0.03	1.35	0.0	DIMS

Table C.21: Electoral Predictors (Indonesia)

	(1) Mean	(2) Median	(3) 25th Pct.	(4) 75th Pct.	(5) Corr. w/ Any	(6) Within/ Between Var.	(7) % Miss	(8) Source
Mainstream party (Golkar/PDIP) share, 1999 local elec. (DPRD-II)	0.56	0.58	0.45	0.69	-0.10	0.10	4.8	GEC
Islamist Party Share, 1999 local elec. (DPRD-II)	0.14	0.11	0.04	0.21	0.11	0.14	4.8	GEC
Vote share Fractionalization, 1999 local elec. (DPRD-II)	0.69	0.72	0.65	0.78	0.13	0.08	4.8	GEC
Vote share Polarization, 1999 local elec. (DPRD-II)	0.16	0.16	0.15	0.18	-0.05	0.14	4.8	GEC
Number of Votes Over Population, 1999 local elec. (DPRD-II)	0.53	0.52	0.46	0.58	-0.03	0.27	5.8	GEC
District-Head Direct Election in Given Year	0.32	0.00	0.00	1.00	-0.04	0.44	0.0	GEC
Number of Votes Over Population, national elec. 99,04,09 (DPR)	0.47	0.49	0.38	0.55	0.04	0.60	1.4	GEC
Mainstream party (Golkar/PDIP) share, national elec. 99,04,09	0.32	0.30	0.21	0.42	-0.18	0.79	0.0	GEC
Islamist Party Share, national elec. 99,04,09 (DPR)	0.16	0.15	0.09	0.22	0.16	0.49	0.0	GEC
Vote share Fractionalization, national elec. 99,04,09 (DPR)	0.85	0.86	0.82	0.89	0.05	1.19	0.0	GEC
Vote share Polarization, national elec. 99,04,09 (DPR)	0.11	0.11	0.09	0.13	-0.02	1.21	0.0	GEC
Number of Parties, national elec. 99,04,09 (DPR)	32.66	38.00	24.00	38.00	0.02	39.05	0.0	GEC

Table C.22: Government Finance Predictors (Indonesia)

	(1) Mean	(2) Median	(3) 25th Pct.	(4) 75th Pct.	(5) Corr. w/ Any	(6) Within/ Between Var.	(7) % Miss	(8) Source
Total Revenue (000s Rp), (ln)	13.03	13.07	12.31	13.74	0.12	0.28	7.0	Podes
Total revenue within village resources, (ln)	11.94	12.12	10.85	13.16	0.13	0.29	12.9	Podes
Revenue from taxes, (ln)	9.74	9.88	8.79	10.77	0.12	0.43	22.3	Podes
Revenue from social organizations, (ln)	10.79	10.87	9.72	11.89	0.17	0.32	15.9	Podes
Revenue from ROSCAs, (ln)	10.17	10.27	9.14	11.29	0.07	0.38	21.3	Podes
Revenue from other within village, (ln)	9.29	9.29	8.11	10.48	0.11	0.49	30.6	Podes
Revenue from higher government admin, (ln)	12.24	12.18	11.60	12.80	0.05	0.33	7.4	Podes
Revenue from central government, (ln)	11.54	11.62	10.85	12.26	-0.05	0.39	25.8	Podes
Revenue from provincial government, (ln)	9.19	9.21	8.08	10.28	0.10	0.66	51.0	Podes
Revenue from district government, (ln)	10.75	10.87	9.68	11.86	0.09	0.47	19.9	Podes
Revenue from district government, (ln)	11.97	11.97	11.23	12.72	0.13	0.27	8.2	Podes
Routine Expenditures, (ln)	12.31	12.40	11.56	13.16	0.06	0.36	9.4	Podes
Total Revenue District per-cap (ln)	14.12	14.08	13.63	14.59	-0.04	1.34	38.4	DAPOER
Self-Generated Kab. Rev. per-capita (ln)	11.14	11.19	10.61	11.72	0.08	1.73	38.4	DAPOER

Table C.23: Output Predictors (Indonesia)

	(1) Mean	(2) Median	(3) 25th Pct.	(4) 75th Pct.	(5) Corr. w/ Any	(6) Within/ Between Var.	(7) % Miss	(8) Source
GDP from 2002 ag. census and curr. comm. pr., (ln)	22.17	22.58	21.04	23.74	-0.13	0.14	9.3	Podes Ag Module/FAO
Nighttime light intensity	9.73	2.87	0.06	9.94	0.34	0.16	2.3	Henderson et al. (2012)
Unemployment rate	0.04	0.03	0.01	0.05	0.08	0.47	6.4	Susenas
Total GDP District (ln)	2.45	2.44	1.92	2.89	0.14	1.08	36.5	DAPOER
Ag GDP District per-cap (ln)	0.98	1.22	0.67	1.66	-0.14	0.79	36.5	DAPOER
Median PCE, (ln)	12.65	12.66	12.34	12.97	0.23	1.04	0.0	Susenas
Mean PCE, (ln)	12.80	12.80	12.45	13.13	0.25	1.04	0.0	Susenas
Unemployed, % working age	0.05	0.04	0.03	0.06	0.04	1.35	0.0	Susenas

Table C.24: Distributional Measures Predictors (Indonesia)

	(1) Mean	(2) Median	(3) 25th Pct.	(4) 75th Pct.	(5) Corr. w/ Any	(6) Within/ Between Var.	(7) % Miss	(8) Source
Wage Gini	0.25	0.24	0.20	0.29	0.18	1.16	0.0	Susenas
Poverty (P_0)	16.54	15.04	10.31	21.59	-0.11	0.46	0.0	Susenas
Poverty (P_1)	2.88	2.36	1.51	3.78	-0.09	0.68	0.0	Susenas
Poverty (P_2)	0.78	0.57	0.35	1.00	-0.07	0.90	0.0	Susenas
PCE Inequality 90-10, (ln)	1.08	1.06	0.86	1.28	0.18	1.17	0.0	Susenas
PCE Inequality 80-20, (ln)	0.70	0.68	0.54	0.83	0.18	1.19	0.0	Susenas
Worked, % working age	0.64	0.64	0.59	0.69	-0.20	0.54	0.0	Susenas
Mean months bought subsidized rice	0.53	0.55	0.34	0.70	-0.09	0.40	0.0	Susenas

Table C.25: Commodity Shock Predictors (Indonesia)

	(1) Mean	(2) Median	(3) 25th Pct.	(4) 75th Pct.	(5) Corr. w/ Any	(6) Within/ Between Var.	(7) % Miss	(8) Source
Cash crop pr. shock	0.02	0.00	-0.01	0.04	0.00	5.29	0.7	Podes Ag Module/FAO
Maj cash crop pr. shock	0.01	0.00	-0.00	0.02	-0.02	4.99	0.7	Podes Ag Module/FAO
Food pr. shock	0.02	0.00	-0.00	0.02	0.00	4.55	0.7	Podes Ag Module/FAO
Bauxite price shock	-0.00	-0.00	-0.00	0.00	0.01	84.31	0.0	SNL/GEM
Coal price shock	0.00	0.00	-0.00	0.00	-0.01	6.06	0.0	SNL/GEM
Copper price shock	0.00	-0.00	-0.00	0.00	-0.03	3.38	0.0	SNL/GEM
Gold price shock	0.00	0.00	0.00	0.01	0.01	1.17	0.0	SNL/GEM
IronOre price shock	0.00	0.00	0.00	0.00	-0.00	4.26	0.0	SNL/GEM
Nickel price shock	0.00	-0.00	-0.00	0.00	-0.00	9.99	0.0	SNL/GEM
Tin price shock	0.00	0.00	-0.00	0.00	-0.02	2.91	0.0	SNL/GEM
Zinc price shock	0.00	-0.00	-0.00	0.00	-0.01	12.42	0.0	SNL/GEM
Silver price shock	0.00	0.00	0.00	0.00	-0.05	2.50	0.0	SNL/GEM
Avocado Price, (ln ch.)	0.00	0.03	-0.10	0.04	0.01	8.69	0.0	FAO
Live Buffalo Price, (ln ch.)	0.03	-0.01	-0.06	0.07	-0.05	5.16	0.0	FAO
Cabbage Price, (ln ch.)	0.05	-0.00	-0.02	0.10	0.01	15.05	0.0	FAO
Carrot Price, (ln ch.)	0.09	0.10	-0.04	0.20	0.00	9.42	0.0	FAO
Cashew Price, (ln ch.)	0.11	0.08	0.02	0.13	0.05	7.18	0.0	FAO
Cassava Price, (ln ch.)	0.06	0.02	-0.03	0.14	0.01	91.71	0.0	FAO
Live Chicken Price, (ln ch.)	0.04	-0.00	-0.04	0.15	-0.03	16.78	0.0	FAO
Chili Pepper Price, (ln ch.)	0.05	0.03	-0.11	0.22	0.01	271.30	0.0	FAO
Cocoa Price, (ln ch.)	0.02	-0.02	-0.09	0.11	0.03	8.31	0.0	FAO
Coffee Price, (ln ch.)	0.05	0.00	-0.04	0.08	0.02	11.32	0.0	FAO
Cucumber Price, (ln ch.)	0.10	0.05	-0.03	0.11	-0.05	5.49	0.0	FAO
Eggplant Price, (ln ch.)	0.04	0.02	-0.02	0.06	0.01	21.58	0.0	FAO
Green Bean Price, (ln ch.)	0.14	0.02	-0.01	0.15	-0.01	23.91	0.0	FAO
Groundnut Price, (ln ch.)	0.06	0.08	0.00	0.11	0.04	5.30	0.0	FAO
Maize Price, (ln ch.)	0.04	0.02	-0.01	0.04	0.02	7.86	0.0	FAO
Mango Price, (ln ch.)	-0.02	-0.04	-0.13	0.10	0.02	60.98	0.0	FAO
Palm Oil Price, (ln ch.)	0.03	0.02	-0.03	0.08	0.00	8.80	0.0	FAO
Orange Price, (ln ch.)	-0.01	-0.01	-0.02	0.00	0.02	5.31	0.0	FAO
Papaya Price, (ln ch.)	0.00	-0.01	-0.03	0.05	0.01	435.33	0.0	FAO
Pepper Price, (ln ch.)	0.01	0.05	-0.04	0.11	0.02	28.05	0.0	FAO
Potato Price, (ln ch.)	0.03	-0.01	-0.07	0.10	0.03	14.53	0.0	FAO
Rubber Price, (ln ch.)	0.01	0.02	-0.07	0.11	-0.02	22.71	0.0	FAO
Soybean Price, (ln ch.)	0.02	0.00	-0.03	0.03	0.01	47.42	0.0	FAO
Sweet Potato Price, (ln ch.)	0.06	0.02	-0.03	0.10	0.04	18.56	0.0	FAO
Tomato Price, (ln ch.)	0.04	0.02	-0.03	0.14	0.01	19.81	0.0	FAO
Bauxite Price, (ln ch.)	-0.02	-0.03	-0.10	0.12	-0.02	16.43	0.0	GEM
Coal Price, (ln ch.)	0.06	0.01	-0.14	0.28	-0.01	22.58	0.0	GEM
Copper Price, (ln ch.)	0.08	-0.00	-0.09	0.34	-0.03	11.37	0.0	GEM
Gold Price, (ln ch.)	0.12	0.15	0.07	0.17	0.00	18.76	0.0	GEM
Iron Ore Price, (ln ch.)	0.08	0.07	0.04	0.51	-0.01	18.75	0.0	GEM
Nickel Price, (ln ch.)	0.00	-0.04	-0.25	0.36	-0.02	12.35	0.0	GEM
Tin Price, (ln ch.)	0.12	0.15	-0.17	0.37	0.00	52.24	0.0	GEM
Zinc Price, (ln ch.)	0.03	-0.06	-0.07	0.23	-0.02	23.47	0.0	GEM
Silver Price, (ln ch.)	0.12	0.06	0.01	0.28	0.00	96.84	0.0	GEM

Table C.26: Commodity Volatility Predictors (Indonesia)

	(1) Mean	(2) Median	(3) 25th Pct.	(4) 75th Pct.	(5) Corr. w/ Any	(6) Within/ Between Var.	(7) % Miss	(8) Source
Cash crop shock, std. dev.	0.07	0.06	0.01	0.11	-0.11	0.06	0.7	Podes Ag Module/FAO
Food crop shock, std. dev.	0.04	0.03	0.00	0.06	0.02	0.04	0.7	Podes Ag Module/FAO
Major cash crop shock, std. dev.	0.04	0.01	0.00	0.08	-0.16	0.04	0.7	Podes Ag Module/FAO

Table C.27: Weather Shock Predictors (Indonesia)

	(1) Mean	(2) Median	(3) 25th Pct.	(4) 75th Pct.	(5) Corr. w/ Any	(6) Within/ Between Var.	(7) % Miss	(8) Source
Temperature, log deviation from average	0.01	0.01	0.00	0.02	0.01	0.69	0.0	UDel
Rainfall, log deviation from average	-0.10	-0.09	-0.24	0.09	0.04	1.39	0.0	UDel

Table C.28: Weather History Predictors (Indonesia)

	(1) Mean	(2) Median	(3) 25th Pct.	(4) 75th Pct.	(5) Corr. w/ Any	(6) Within/ Between Var.	(7) % Miss	(8) Source
Rainfall, 1900-1990 average	224.22	221.46	184.44	260.17	-0.15	0.00	0.0	UDel
Temperature, 1900-1990 average	25.32	25.93	24.32	26.66	0.06	0.00	0.0	UDel
Rainfall, 1900-1990 standard deviation	0.19	0.18	0.14	0.22	0.03	0.00	0.0	UDel
Temperature, 1900-1990 standard deviation	0.01	0.01	0.01	0.02	0.05	0.00	0.0	UDel

C.3.2 Colombia

Table C.29: Violence Predictors (Colombia)

	(1) Mean	(2) Median	(3) 25th Pct.	(4) 75th Pct.	(5) Corr. w/ Any	(6) Within/ Between Var.	(7) % Miss	(8) Source
Guerilla attacks, number	0.62	0.00	0.00	0.00	0.33	0.94	0.0	CERAC
Paramilitary attacks, number	0.09	0.00	0.00	0.00	0.20	1.67	0.0	CERAC
Clashes, number	0.54	0.00	0.00	0.00	0.35	1.19	0.0	CERAC
Casualties	2.23	0.00	0.00	1.00	0.27	1.47	0.0	CERAC
Government attacks, number	0.10	0.00	0.00	0.00	0.22	1.78	0.0	CERAC
Paramilitary initiated massacres	0.06	0.00	0.00	0.00	0.17	1.89	0.0	CERAC
Guerilla initiated massacres	0.01	0.00	0.00	0.00	0.09	2.29	0.0	CERAC
Paramilitary attacks on infrastructure	0.00	0.00	0.00	0.00	0.05	2.76	0.0	CERAC
Paramilitary attacks, non-infrastructure	0.09	0.00	0.00	0.00	0.19	1.72	0.0	CERAC
All attacks, number	0.81	0.00	0.00	1.00	0.35	0.88	0.0	CERAC
Violent events (attacks and clashes)	1.36	0.00	0.00	1.00	0.38	0.88	0.0	CERAC
Any violent event, indicator	0.36	0.00	0.00	1.00	0.46	1.18	0.0	CERAC
Five or more violent events, indicator	0.09	0.00	0.00	0.00	0.35	1.19	0.0	CERAC
More than 1 s.d. increase in non-crime incidents	0.05	0.00	0.00	0.00	0.23	2.27	0.0	CERAC

Table C.30: Population Predictors (Colombia)

	(1) Mean	(2) Median	(3) 25th Pct.	(4) 75th Pct.	(5) Corr. w/ Any	(6) Within/ Between Var.	(7) % Miss	(8) Source
log population, millions	-4.27	-4.34	-4.97	-3.68	0.30	0.08	0.0	CEDE
Large municipalities (pop. exceeds 250K)	0.02	0.00	0.00	0.00	0.13	0.00	0.0	CEDE
Population density	0.00	0.00	0.00	0.00	0.04	0.09	0.0	CEDE
Population Growth (ln ch.)	0.01	0.01	-0.01	0.02	0.04	0.91	0.0	CEDE

Table C.31: Historical Traits Predictors (Colombia)

	(1) Mean	(2) Median	(3) 25th Pct.	(4) 75th Pct.	(5) Corr. w/ Any	(6) Within/ Between Var.	(7) % Miss	(8) Source
Total Crown Employees	0.23	0.00	0.00	0.00	0.11	0.00	3.6	Acemoglu et al. 2015
Non-military Crown Employees	0.23	0.00	0.00	0.00	0.11	0.00	3.6	Acemoglu et al. 2015
Colonial State Presence Index	0.57	0.00	0.00	1.00	0.06	0.00	3.6	Acemoglu et al. 2015
City Status Dummy	0.04	0.00	0.00	0.00	0.09	0.00	3.6	Acemoglu et al. 2015
Distance to royal roads	29.09	14.39	6.29	32.51	0.11	0.00	3.6	Acemoglu et al. 2015
Share of slaves in population in 1843	0.01	0.00	0.00	0.00	0.12	0.00	3.6	Acemoglu et al. 2015
Slave presence in 1843 dummy	0.42	0.00	0.00	1.00	0.07	0.00	3.6	Acemoglu et al. 2015
Number of Indians in 1560	1.99	0.00	0.00	5.14	-0.13	0.00	3.6	Acemoglu et al. 2015
Number of encomiendas in 1560	0.32	0.00	0.00	0.69	-0.07	0.00	3.6	Acemoglu et al. 2015
Presence of encomiendas in 1560 dummy	0.32	0.00	0.00	1.00	-0.11	0.00	3.6	Acemoglu et al. 2015
Presence of colonial gold mines in 1560 dummy	0.04	0.00	0.00	0.00	0.10	0.00	3.6	Acemoglu et al. 2015
Foundation date	1783.59	1799.00	1701.00	1888.00	0.05	0.00	3.7	Acemoglu et al. 2015
Population in 1843 (mean imputed)	2990.87	2892.24	2087.00	3073.00	0.05	0.00	3.6	Acemoglu et al. 2015
Dummy for missing population in 1843	0.41	0.00	0.00	1.00	0.05	0.00	3.6	Acemoglu et al. 2015

Table C.32: Geography Predictors (Colombia)

	(1) Mean	(2) Median	(3) 25th Pct.	(4) 75th Pct.	(5) Corr. w/ Any	(6) Within/ Between Var.	(7) % Miss	(8) Source
Surface area, sq. km.	1088.29	289.95	133.25	709.77	0.08	0.11	0.0	CEDE
Proportion of Land Uninhabitable	0.16	0.00	0.00	0.30	0.03	0.00	3.2	CEDE
% optimal sutiable for sugar cane	0.36	0.00	0.00	1.00	-0.02	0.00	0.7	Agriculture Ministry
% suboptimal suitable for sugar cane	0.11	0.00	0.00	0.00	0.03	0.00	0.7	Agriculture Ministry
% suitable for sugar cane	0.47	0.25	0.00	1.00	-0.01	0.00	0.7	Agriculture Ministry
% no suitable for sugar cane	0.62	1.00	0.00	1.00	0.10	0.00	0.7	Agriculture Ministry
% optimal sutiable for palm	0.43	0.00	0.00	1.00	-0.04	0.00	0.7	Agriculture Ministry
% suboptimal suitable for palm	0.25	0.00	0.00	0.50	0.01	0.00	0.7	Agriculture Ministry
% suitable for palm	0.68	1.00	0.25	1.00	-0.04	0.00	0.7	Agriculture Ministry
% no suitable for palm	0.41	0.50	0.00	0.75	0.14	0.00	0.7	Agriculture Ministry
Slope Mean	35.62	36.63	21.04	46.94	0.08	0.00	32.3	CEDE
Slope Standard Deviation	19.50	21.06	15.74	23.09	0.10	0.00	32.3	CEDE
Percentage of slightly flat terrain	0.20	0.14	0.01	0.26	0.04	0.00	32.3	CEDE
Percentage of slightly sloped terrain	0.05	0.02	0.00	0.08	0.02	0.00	32.3	CEDE
Percentage of moderately sloped terrain	0.07	0.05	0.01	0.07	-0.07	0.00	32.3	CEDE
Percentage of strongly sloped terrain	0.15	0.12	0.04	0.19	-0.06	0.00	32.3	CEDE
Percentage of slightly steep terrain	0.14	0.14	0.05	0.18	-0.04	0.00	32.3	CEDE
Percentage of moderately steep terrain	0.30	0.33	0.09	0.41	0.07	0.00	32.3	CEDE
Percentage stongly steep terrain	0.11	0.08	0.00	0.15	0.01	0.00	32.3	CEDE
Mean of water availability	2.30	2.27	1.90	2.60	0.15	0.00	32.3	CEDE
Standard Deviation of water availability	0.53	0.51	0.45	0.64	0.02	0.00	32.3	CEDE
Meters of main rivers	11436.58	0.00	0.00	5503.38	0.02	0.00	2.3	CEDE
Meters of secondary rivers	23726.09	381.10	0.00	17630.90	0.09	0.00	2.3	CEDE
Meters of tertiary rivers	12113.59	0.00	0.00	8811.40	0.17	0.00	2.3	CEDE
Density of Primary Rivers, km/km2	33.62	0.00	0.00	12.71	-0.01	0.00	3.2	CEDE
Density of Secondary Rivers, km/km2	28.93	0.00	0.00	40.34	0.04	0.00	3.2	CEDE
Density of Tertiary Rivers, km/km2	22.40	0.00	0.00	24.12	0.02	0.00	3.2	CEDE
Flat land share	0.02	0.00	0.00	0.00	0.01	0.00	1.8	CEDE
Hill share	0.21	0.00	0.00	0.32	0.06	0.00	1.8	CEDE
Mountain share	0.68	0.92	0.25	1.00	-0.01	0.00	1.8	CEDE
Valley share	0.14	0.00	0.00	0.18	0.04	0.00	1.8	CEDE
Water bodies share	0.00	0.00	0.00	0.00	0.00	0.00	1.8	CEDE
Hilly terrain proportion	0.27	0.28	0.02	0.44	0.05	0.00	0.3	CEDE
Mountainous terrain proportion	0.10	0.06	0.00	0.16	0.04	0.00	0.3	CEDE
Rugged terrain proportion	0.01	0.00	0.00	0.01	0.07	0.00	0.3	CEDE
Maximum slope	43.63	50.98	14.86	65.60	0.21	0.00	0.4	CEDE

Table C.33: Remoteness Predictors (Colombia)

	(1) Mean	(2) Median	(3) 25th Pct.	(4) 75th Pct.	(5) Corr. w/ Any	(6) Within/ Between Var.	(7) % Miss	(8) Source
Paved Roads, km	5749.71	0.00	0.00	7352.70	0.15	0.00	1.8	CEDE
Unpaved Roads, km	2493.58	0.00	0.00	0.00	0.11	0.00	1.8	CEDE
Paved Secondary Roads, km	3533.15	0.00	0.00	1807.24	0.10	0.00	1.8	CEDE
Unpaved Secondary Roads, km	14256.91	6858.80	0.00	20421.30	0.15	0.00	1.8	CEDE
Dirt Roads, km	16590.36	4618.80	0.00	18743.80	0.14	0.00	1.8	CEDE
Paved Roads Density, km/km^2	19.47	0.00	0.00	28.18	-0.05	0.00	4.7	CEDE
Unpaved Roads Density, km/km^2	4.64	0.00	0.00	0.00	0.02	0.00	4.7	CEDE
Paved Secondary Roads Density, km/km^2	13.71	0.00	0.00	6.05	-0.09	0.00	4.7	CEDE
Unpaved Secondary Roads Density, km/km^2	46.53	22.51	0.00	72.47	-0.09	0.00	4.7	CEDE
Dirt Roads Density, km/km^2	40.05	16.08	0.00	57.52	-0.05	0.00	4.7	CEDE

Table C.34: Distributional Measures Predictors (Colombia)

	(1) Mean	(2) Median	(3) 25th Pct.	(4) 75th Pct.	(5) Corr. w/ Any	(6) Within/ Between Var.	(7) % Miss	(8) Source
Gini Coefficient for Land Holdings	0.72	0.72	0.64	0.76	0.06	0.00	16.6	DANE
Unmet Basic Needs Index, 1993	59.80	52.41	38.24	68.73	0.12	0.00	0.0	DANE
Life Quality Index	55.80	54.20	46.80	64.30	0.10	0.00	1.9	DANE

Table C.35: DMZ Proximity Predictors (Colombia)

	(1) Mean	(2) Median	(3) 25th Pct.	(4) 75th Pct.	(5) Corr. w/ Any	(6) Within/ Between Var.	(7) % Miss	(8) Source
Municipality is part of Demilitarized Zone	0.00	0.00	0.00	0.00	0.05	0.00	0.0	Authors' Calculations
Distance to DMZ	99.39	91.80	55.42	127.33	0.01	0.00	1.6	Authors' Calculations

Table C.36: Government Finance Predictors (Colombia)

	(1) Mean	(2) Median	(3) 25th Pct.	(4) 75th Pct.	(5) Corr. w/ Any	(6) Within/ Between Var.	(7) % Miss	(8) Source
Total Government Revenue, Mil. Pesos per Cap.	4.91	5.11	4.02	5.72	-0.04	2.05	9.3	CEDE
Total Government Spending, Mil. Pesos per Cap.	4.90	5.11	3.99	5.72	-0.04	2.08	9.0	CEDE
Current Government Revenue, Mil. Pesos per Cap.	4.03	3.97	3.50	4.49	-0.11	1.26	9.3	CEDE
Current Government Spending, Mil. Pesos per Cap.	3.77	3.82	3.18	4.33	-0.06	1.53	9.2	CEDE
Current Government Deficit, Mil. Pesos per Cap.	0.01	0.01	-0.00	0.03	-0.01	3.71	9.0	CEDE
Capital Spending, Mil. Pesos per Cap.	4.43	4.72	3.30	5.43	-0.03	2.35	9.3	CEDE
Tax Revenue, Mil. Pesos per Cap.	2.52	2.43	1.47	3.45	-0.03	1.29	9.5	CEDE
Nontax Revenue, Mil. Pesos per Cap.	1.90	1.80	0.96	2.72	-0.03	1.57	10.0	CEDE
Functional Government Spending, Mil. Pesos per Cap.	3.69	3.73	3.11	4.26	-0.07	1.47	9.4	CEDE
Capital Spending (Remaining Investments), Mil. Pesos per Cap.	3.19	3.85	1.08	4.83	0.01	4.06	12.3	CEDE
Land tax revenue, Mil. Pesos per Cap.	1.89	1.79	1.02	2.66	-0.11	1.14	10.2	CEDE
Industry tax revenue, Mil. Pesos per Cap.	1.28	0.89	0.27	1.91	0.02	1.08	12.3	CEDE
Other tax revenue, Mil. Pesos per Cap.	1.46	1.13	0.29	2.44	0.00	1.65	10.6	CEDE
Personel spending, Mil. Pesos per Cap.	3.03	3.05	2.42	3.59	-0.07	1.43	9.8	CEDE
Capital Revenue, Mil. Pesos per Cap.	3.83	4.64	1.92	5.38	-0.01	3.50	27.4	CEDE
Total Deficit, Mil. Pesos per Cap.	-0.01	0.00	-0.02	0.01	0.00	6.59	9.0	CEDE
Financing Spending, Mil. Pesos per Cap.	0.01	-0.00	-0.01	0.02	-0.00	6.59	9.0	CEDE
Current Transfer Income, Mil. Pesos per Cap.	3.34	3.31	2.89	3.78	-0.17	0.92	11.7	CEDE
Spending on Capital Formation, Mil. Pesos per Cap.	3.90	3.97	3.10	4.67	-0.08	1.64	9.4	CEDE
Internal or External Credit, Mil. Pesos per Cap.	-0.00	-0.00	-0.00	0.01	-0.00	3.53	26.1	CEDE
Balance Sheet Resources, Mil. Pesos per Cap.	0.01	-0.00	-0.02	0.01	-0.00	4.47	9.0	CEDE
General Expenditure, Mil. Pesos per Cap.	2.61	2.58	2.03	3.14	-0.10	1.30	9.4	CEDE
Spending for Transfers, Mil. Pesos per Cap.	1.87	1.86	0.81	2.65	-0.02	2.05	13.1	CEDE
National Expenditure, Mil. Pesos	35266.08	8228.00	4378.00	16382.00	0.07	0.00	0.7	CEDE

Table C.37: Electoral Predictors (Colombia)

	(1) Mean	(2) Median	(3) 25th Pct.	(4) 75th Pct.	(5) Corr. w/ Any	(6) Within/ Between Var.	(7) % Miss	(8) Source
Mayor Senate Election Participation Rate	0.33	0.32	0.25	0.40	-0.16	0.66	1.9	CEDE
Chamber Election Participation Rate	0.23	0.21	0.15	0.29	-0.09	0.76	0.2	CEDE
Senate Election Participation Rate	0.23	0.21	0.14	0.29	-0.10	0.96	0.2	CEDE
National Election, indicator	0.29	0.00	0.00	1.00	-0.01	Inf	0.0	CEDE
Presidential Election, indicator	0.21	0.00	0.00	0.00	-0.01	Inf	0.0	CEDE
Local Election, indicator	0.43	0.00	0.00	1.00	-0.01	Inf	0.0	CEDE
Mayoral Election, indicator	0.36	0.00	0.00	1.00	-0.01	Inf	0.0	CEDE
Herfindahl Index, Chamber	4101.27	3679.04	2770.71	4871.35	0.04	0.74	0.2	CEDE
Factionalization Index, Chamber	0.64	0.65	0.54	0.74	0.09	0.75	0.2	CEDE
Polarization Index, Chamber	0.19	0.18	0.16	0.20	0.08	0.65	0.2	CEDE
Herfindahl Index, Senate	3461.82	3064.05	2358.01	4102.43	0.03	0.88	0.2	CEDE
Factionalization Index, Senate	0.71	0.71	0.61	0.78	0.10	0.81	0.2	CEDE
Polarization Index, Senate	0.18	0.17	0.16	0.19	0.07	0.64	0.2	CEDE
Winner's Margin, Chamber	0.32	0.27	0.12	0.48	-0.00	0.92	0.7	CEDE
Winner's Margin, Senate	0.29	0.26	0.11	0.44	0.01	1.36	0.7	CEDE
Winner's Share, Mayor	0.58	0.54	0.47	0.66	0.01	1.55	2.8	CEDE
Winner's Share, Chamber	0.50	0.48	0.37	0.63	-0.01	1.07	0.7	CEDE
Second Party's Share, Chamber	0.18	0.18	0.11	0.25	-0.02	0.99	0.7	CEDE
Winner's Share, Senate	0.46	0.45	0.34	0.57	-0.01	1.33	0.7	CEDE
Second Party's Share, Senate	0.17	0.17	0.10	0.23	-0.05	1.37	0.7	CEDE
Winning Party Leans Other, Chamber, Indicator	0.03	0.00	0.00	0.00	0.02	1.50	0.3	CEDE
Winning Party Leans Right, Chamber, Indicator	0.29	0.00	0.00	1.00	-0.11	0.70	0.3	CEDE
Winning Party Leans Left, Chamber, Indicator	0.69	1.00	0.00	1.00	0.12	0.73	0.3	CEDE
Winning Party Leans Other, Senate, Indicator	0.05	0.00	0.00	0.00	0.03	1.48	0.3	CEDE
Winning Party Leans Right, Senate, Indicator	0.20	0.00	0.00	0.00	-0.06	1.06	0.3	CEDE
Winning Party Leans Left, Senate, Indicator	0.76	1.00	1.00	1.00	0.06	1.09	0.3	CEDE
Second Party Leans Other, Chamber, Indicator	0.14	0.00	0.00	0.00	0.07	1.06	0.3	CEDE
Second Party Leans Right, Chamber, Indicator	0.51	1.00	0.00	1.00	0.01	1.06	0.3	CEDE
Second Party Leans Left, Chamber, Indicator	0.37	0.00	0.00	1.00	-0.04	1.00	0.3	CEDE
Second Party Leans Other, Senate, Indicator	0.20	0.00	0.00	0.00	0.06	1.32	0.3	CEDE
Second Party Leans Right, Senate, Indicator	0.50	0.00	0.00	1.00	-0.04	1.18	0.3	CEDE
Second Party Leans Left, Senate, Indicator	0.32	0.00	0.00	1.00	0.01	1.14	0.3	CEDE
Winning Party Other, Mayor, Indicator	0.31	0.00	0.00	1.00	0.06	1.47	1.9	CEDE
Winning Party Liberal, Mayor, Indicator	0.43	0.00	0.00	1.00	0.04	0.90	1.9	CEDE
Winning Party Conservative, Mayor, Indicator	0.28	0.00	0.00	1.00	-0.09	0.88	1.9	CEDE
Winning Party Other, Chamber, Indicator	0.18	0.00	0.00	0.00	0.03	1.46	0.2	CEDE
Winning Party Liberal, Chamber, Indicator	0.62	1.00	0.00	1.00	0.09	0.90	0.2	CEDE
Winning Party Conservative, Chamber, Indicator	0.21	0.00	0.00	0.00	-0.11	0.85	0.2	CEDE
Winning Party Other, Senate, Indicator	0.20	0.00	0.00	0.00	0.02	1.59	0.2	CEDE
Winning Party Liberal, Senate, Indicator	0.69	1.00	0.00	1.00	0.04	1.27	0.2	CEDE
Winning Party Conservative, Senate, Indicator	0.13	0.00	0.00	0.00	-0.06	1.41	0.2	CEDE
Second Party Other, Chamber, Indicator	0.43	0.00	0.00	1.00	0.05	1.22	0.2	CEDE
Second Party Liberal, Chamber, Indicator	0.26	0.00	0.00	1.00	-0.06	1.14	0.2	CEDE
Second Party Conservative, Chamber, Indicator	0.33	0.00	0.00	1.00	0.03	1.28	0.2	CEDE
Second Party Other, Senate, Indicator	0.52	1.00	0.00	1.00	0.04	1.67	0.2	CEDE
Second Party Liberal, Senate, Indicator	0.22	0.00	0.00	0.00	-0.01	1.54	0.2	CEDE
Second Party Conservative, Senate, Indicator	0.28	0.00	0.00	1.00	-0.01	1.61	0.2	CEDE

Table C.38: U.S. Military Involvement Predictors (Colombia)

	(1) Mean	(2) Median	(3) 25th Pct.	(4) 75th Pct.	(5) Corr. w/ Any	(6) Within/ Between Var.	(7) % Miss	(8) Source
Log U.S. military aid to Colombia	0.01	0.18	-1.04	0.81	-0.03	Inf	0.0	USAID
Log U.S. narcotics aid to Colombia	0.22	0.11	-0.01	0.74	-0.00	Inf	0.0	USAID
Log U.S. military and narcotics aid to Colombia	0.09	0.11	-0.41	0.71	-0.01	Inf	0.0	USAID
Base Presence	0.03	0.00	0.00	0.00	0.11	0.06	0.0	Dube and Naidu 2015
Base Presence X US mil. and narc. assistance (ln ch.)	0.00	0.00	0.00	0.00	0.01	12.80	0.0	USAID/Dube and Naidu 2015
Base Presence X US military assistance (ln ch.)	-0.00	0.00	0.00	0.00	-0.01	47.80	0.0	USAID/Dube and Naidu 2015
Base Presence X US narcotics assistance (ln ch.)	0.01	0.00	0.00	0.00	0.02	5.20	0.0	USAID/Dube and Naidu 2015

Table C.39: Commodity Production Predictors (Colombia)

	(1) Mean	(2) Median	(3) 25th Pct.	(4) 75th Pct.	(5) Corr. w/ Any	(6) Within/ Between Var.	(7) % Miss	(8) Source
Oil production, 1988, 00,000s bar/day	0.00	0.00	0.00	0.00	0.07	0.00	0.0	MME
Coal Reserves, 1978, indicator	0.32	0.00	0.00	1.00	-0.07	0.00	0.0	USGS
Precious metal mining, 1978, hectares	589.02	0.00	0.00	0.00	0.09	0.00	0.0	Jacome (1978)
Gold production, Department, 1987	2605.27	223.15	0.00	555.26	0.15	0.00	0.3	Igneominas
Coal production, Department, 1990	638.18	0.00	0.00	948.00	-0.16	0.00	1.2	Igneominas
Coffee Cultivation, thsd. hectares, 1997	0.85	0.05	0.00	0.96	0.14	0.00	1.9	NFCG

Table C.40: Commodity Shock Predictors (Colombia)

	(1) Mean	(2) Median	(3) 25th Pct.	(4) 75th Pct.	(5) Corr. w/ Any	(6) Within/ Between Var.	(7) % Miss	(8) Source
Coffee Cultivation (000s hec., 1997) X Coffee Price (log ch.)	-0.04	0.00	-0.04	0.00	-0.04	4.36	1.9	NFCG
Oil Production (00,000s barrels/day, 1988) X Oil Price (log ch.)	0.00	0.00	0.00	0.00	0.00	30.28	0.0	IFS/MME
Mining (000s hectares, 1978) X Gold Price (log ch.)	-14.01	0.00	0.00	0.00	-0.02	5.28	0.0	GFD/Jacome (1978)
Coal Reserves Indicator (1978) X Oil Price (log ch.)	-0.00	0.00	0.00	0.00	-0.02	33.11	0.0	IMF/USGS
Mining (000s hectares, 1978) X Silver Price (log ch.)	0.61	0.00	0.00	0.00	-0.00	37.22	0.0	GFD/Jacome (1978)
Mining (000s hectares, 1978) X Platinum Price (log ch.)	10.66	0.00	0.00	0.00	0.01	10.19	0.0	GFD/Jacome (1978)
log internal coffee price, thsds 2006 pesos/lb	-0.04	-0.07	-0.16	0.06	-0.04	Inf	0.0	NFCG
log int'l oil price, thsds 2006 pesos/barrel	0.01	-0.11	-0.17	0.16	0.03	Inf	0.0	IFS
log int'l gold price, millions 2006 pesos/ounce	-0.02	-0.04	-0.10	0.07	0.04	Inf	0.0	GFD
log int'l coal price, thsds 2006 pesos/ton	-0.01	-0.02	-0.12	0.02	0.00	Inf	0.0	IMF
log int'l silver price, millions 2006 pesos/ounce	0.00	0.03	-0.10	0.08	-0.00	Inf	0.0	GFD
log int'l platinum price, millions 2006 pesos/ounce	0.02	-0.03	-0.08	0.06	0.04	Inf	0.0	GFD

Table C.41: Drug Production Predictors (Colombia)

	(1) Mean	(2) Median	(3) 25th Pct.	(4) 75th Pct.	(5) Corr. w/ Any	(6) Within/ Between Var.	(7) % Miss	(8) Source
Poppy Production, hectares, 1994	19.89	0.00	0.00	0.00	0.07	0.00	0.0	DNE
Poppy Production, indicator, 1994	0.14	0.00	0.00	0.00	0.08	0.00	0.0	DNE
Coca Production, hectares, 1994	70.12	0.00	0.00	0.00	0.08	0.00	0.0	DNE
Coca Production, indicator, 1994	0.05	0.00	0.00	0.00	0.12	0.00	0.0	DNE
Erradicated hectares, 1994	3.78	0.00	0.00	0.00	0.05	0.00	0.0	DNE
Erradication occured, indicator, 1994	0.01	0.00	0.00	0.00	0.07	0.00	0.0	DNE

Table C.42: Drug Shock Predictors (Colombia)

	(1) Mean	(2) Median	(3) 25th Pct.	(4) 75th Pct.	(5) Corr. w/ Any	(6) Within/ Between Var.	(7) % Miss	(8) Source
Coca Production in Colombia, hectares	0.05	0.04	-0.07	0.17	-0.00	Inf	0.0	UNODC
Coke Production in Colombia, tons	-0.07	0.01	-0.29	0.06	-0.04	Inf	0.0	UNODC
Eradication in Colombia, hectares	0.37	0.39	-0.11	0.67	-0.03	Inf	0.0	UNODC
Coca Leaf Production in Colombia, tons	0.09	0.01	-0.06	0.25	-0.02	Inf	0.0	UNODC
Cocaine Production in Colombia, tons	0.14	0.14	-0.04	0.27	-0.04	Inf	0.0	NA
Cocaine Retail Price, Europe, log change	-0.02	-0.01	-0.12	0.05	-0.03	Inf	0.0	UNODC
Cocaine Wholesale Price, Europe, log change	-0.05	-0.05	-0.07	-0.00	-0.04	Inf	0.0	UNODC
Cocaine Wholesale Price, U.S., log change	-0.03	-0.01	-0.08	0.07	-0.03	Inf	0.0	UNODC
Coca Production in Bolivia, hectares	-0.04	0.00	-0.06	0.08	-0.03	Inf	0.0	UNODC
Coca Production in Peru, hectares	-0.06	-0.00	-0.20	0.06	-0.01	Inf	0.0	UNODC
Coca Production in Bolivia, Colombia, and Peru, hectares	-0.02	-0.02	-0.08	0.03	-0.01	Inf	0.0	UNODC
coca leaf production, world total	-0.01	-0.01	-0.08	0.08	-0.03	Inf	0.0	NA
Coke Production in Bolivia, tons	-0.04	0.01	-0.11	0.15	-0.04	Inf	0.0	UNODC
Coke Production in Peru, tons	0.14	0.14	-0.04	0.27	-0.04	Inf	0.0	UNODC
Coke Production in Bolivia, Colombia, and Peru, tons	0.01	0.00	-0.06	0.07	-0.05	Inf	0.0	UNODC
Coke Production in Bolivia and Peru, tons	-0.06	0.03	-0.21	0.06	-0.04	Inf	0.0	UNODC
Eradication in Bolivia, hectares	0.00	-0.07	-0.39	0.28	0.00	Inf	0.0	UNODC
Coca Leaf Production in Bolivia, tons	-0.05	-0.00	-0.12	0.06	-0.03	Inf	0.0	UNODC
Cocaine Production in Bolivia, tons	-0.05	0.01	-0.11	0.15	-0.04	Inf	0.0	UNODC
cocaine exports to switzerland from colombia	0.09	0.05	-0.25	0.47	-0.01	Inf	0.0	NA
Eradication in Peru, hectares	0.02	-0.00	-0.39	0.63	0.02	Inf	0.0	UNODC
Coca Leaf Production in Peru, tons	-0.05	0.03	-0.31	0.12	-0.04	Inf	0.0	UNODC
Cocaine Production in Colombia, tons	-0.04	0.05	-0.29	0.06	-0.04	Inf	0.0	UNODC
cocaine exports to switzerland from peru	-0.07	0.08	-0.25	0.18	-0.02	Inf	0.0	NA
U.S. Wholesale Price (log ch.) X Hect. Coca Prod. 1994	1.85	0.00	0.00	0.00	0.01	3.98	0.0	DNE/UNODC

Table C.43: Weather History Predictors (Colombia)

	(1) Mean	(2) Median	(3) 25th Pct.	(4) 75th Pct.	(5) Corr. w/ Any	(6) Within/ Between Var.	(7) % Miss	(8) Source
Rainfall, 1900-1990 average	168.37	154.98	115.96	201.61	0.16	0.00	0.2	UDel
Rainfall, 1900-1990 S.D.	0.15	0.14	0.12	0.17	-0.13	0.00	0.2	UDel
Temperature, 1900-1990 average	21.07	21.03	17.27	25.37	0.14	0.00	0.2	UDel
Temperature, 1900-1990 S.D.	0.03	0.03	0.02	0.03	-0.07	0.00	0.2	UDel

Table C.44: Weather Shock Predictors (Colombia)

	(1) Mean	(2) Median	(3) 25th Pct.	(4) 75th Pct.	(5) Corr. w/ Any	(6) Within/ Between Var.	(7) % Miss	(8) Source
Rainfall, log deviation from average	-0.09	-0.08	-0.17	-0.00	0.04	3.47	0.2	UDel
Temperature, log deviation from average	0.01	0.01	0.00	0.03	-0.04	2.42	0.2	UDel

C.4 Correlating Commodity Price and Weather Shocks to Conflict

Our finding that time varying shocks add little to our predictive power in forecasting conflict over time is surprising. To ascertain that this does not stem from data quality issues, in this section, we look more closely at two specific types of shocks constructed in our data — weather and commodity price shocks. Specifically, we verify that these shocks correlate with conflict outcomes in ways that are consistent with prior findings in the literature on causes of conflict. To this end, we regress each of our outcomes—both the year-ahead and contemporaneous conflict outcomes—on the price and weather shocks as well as year and location fixed effects. We report the coefficient estimates with standard errors clustered at the sub-district/municipality level.

Table C.45 reports the results of these regressions for Indonesia. Rainfall has a statistically negative association with future conflict escalations but exhibits no other correlations. Temperature shocks have a statistically significant positive association with the incidence of five or more events in the current year and in the following year. This is consistent with Hsiang et al. (2013) who find more extreme weather associated with a higher incidence of conflict across countries and with Wright and Signoret (2016) who find that positive temperature shocks increase the likelihood of (most types of) intergroup conflict within Indonesia.

Food crop prices do not exhibit a consistent relationship with any indicators of conflict. This is somewhat inconsistent with McGuirk and Burke (2017) who find that higher food prices in Sub-Saharan Africa are positively associated with conflict in areas that are net consumers. Of course, our shocks are weighted by production, so we might expect the opposite effect, as commodity prices raise demand for labor. Major cash crops, on the other hand, have a negative association with indicators of five or more events and with future escalations. Since they are not consumed locally, the likely mechanism is an increase in wages and the opportunity cost of conflict. These results, that correlations are statistically significant for some but not all indicators, underscore that the relationship between weather and conflict incidence is complex—with different correlations with the intensive and extensive margin as well as different correlations with levels versus changes.

Table C.46 shows a similar exercise for Colombia. We do not find a statistically significant

relationship between weather and conflict. In examining the correlations with commodity prices, we look at coffee and oil price shocks, as in Dube and Vargas (2013). However, our specifications differ along three dimensions. First, to account for endogeneity, their paper instruments coffee prices with the coffee production levels of other countries — as Colombia was a major producer of coffee in the international market over 1988-2005. Second they examine effects on number of conflict events in each municipality-year, while we examine indicators for the occurrence of any conflict event, more than five events and an escalation. Third, they examine effects on contemporaneous conflict, while we examine effects contemporaneously and in the following year.

Consistent with Dube and Vargas (2013), we find that coffee price shocks are negatively associated with current conflict. However, we observe an insignificant association with future conflict. With oil prices, we observe positive significant positive correlations with indicators of five or more events and escalations of conflict one year ahead.

These results are not meant to comment on the mechanisms that cause conflict in Indonesia nor Colombia. As discussed above, the conflict literature has taken more sophisticated approaches to estimating these relationships that takes into account endogeneity and other modeling concerns. Instead, we emphasize that the commodity price and weather shocks we examine are correlated with conflict in ways that are largely consistent with results on the literature examining the causes of conflict. To the extent that there is variation in the strength of this correlation, this variation may reflect, in part, nuanced underlying relationships between time varying shocks and conflict outcomes.

Table C.45: Indonesia Shocks with Fixed Effects

	Type of Shock							
	Rainfall		Temperature		Food Price		Major Cash Crop Price	
Dep. Var.: conflict in	t	$t+1$	t	$t+1$	t	$t+1$	t	$t+1$
	(a) Any violent event							
shock in t	0.004	-0.002	-0.001	-0.009	-0.003	-0.003	-0.005	0.001
	(0.006)	(0.005)	(0.007)	(0.008)	(0.004)	(0.004)	(0.004)	(0.004)
	(b) $\geq$ 5 violent events							
shock in t	-0.002	-0.004	0.018***	0.006*	-0.003	0.004	-0.003**	-0.003**
	(0.003)	(0.003)	(0.004)	(0.004)	(0.003)	(0.003)	(0.001)	(0.001)
	(c) $\geq$ 1 s.d. increase in events							
shock in t	-0.003	-0.006**	0.003	-0.001	0.001	0.003	-0.002	-0.002*
	(0.003)	(0.002)	(0.003)	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)

Notes: Right hand-side variables are standardized to have mean zero and standard deviation 1. Standard errors clustered at the sub-district level. Sub-district and year fixed effects included but not reported. Price shocks are defined as 1 year log change. Food prices are weighted by each food crops share of agriculture in 2002, as are cash crop prices. Weather shocks are defined as log differences from historical means.

Table C.46: Colombia Shocks with Fixed Effects

	Type of Shock							
	Rainfall		Temperature		Coffee Price		Oil Price	
Dep. Var.: conflict in	t	$t+1$	t	$t+1$	t	$t+1$	t	$t+1$
	(a) Any violent event							
shock in t	0.003	0.000	0.005	0.013*	-0.001	0.003	-0.003	0.000
	(0.005)	(0.005)	(0.007)	(0.007)	(0.004)	(0.004)	(0.002)	(0.001)
	(b) $\geq$ 5 violent events							
shock in t	0.000	-0.004	0.003	0.002	-0.008***	-0.001	0.002	0.008***
	(0.003)	(0.003)	(0.004)	(0.004)	(0.002)	(0.002)	(0.001)	(0.002)
	(c) $\geq$ 1 s.d. increase in events							
shock in t	0.003	-0.002	0.002	0.001	-0.002	-0.004	0.001	0.009***
	(0.002)	(0.002)	(0.003)	(0.003)	(0.002)	(0.002)	(0.001)	(0.001)

Notes: Right hand-side variables are standardized to have mean zero and standard deviation 1. Standard errors clustered at the municipality level. Municipality and year fixed effects included but not reported. Price shocks are defined as 1 year log change. Coffee prices are weighted by coffee cultivation in 1997, oil prices are weighted by 1988 oil production. Weather shocks are defined as log differences from historical means.