Forecasting Civil Wars*

*Acknowledgements: The research described in this project was sponsored by the Political Instability Task Force (PITF). The PITF is funded by the Central Intelligence Agency. The views expressed herein are the Principal Investigators' alone and do not represent the views of the US Government.

Abstract

We use event data to show that a parsimonious model grounded in theoretical intuitions from the literature on conflict escalation can forecast civil war onsets with high levels of accuracy and over much shorter intervals than was previously possible. Our theoretically-motivated model outperforms more atheoretical alternatives, and also outperforms models based on "structural characteristics" (such as regime type) alone. Our results raise the possibility of a more direct connection between theory and forecasting than is typically assumed, and suggest that process can substitute for structure over short forecasting windows. We assess predictive performance retrospectively, but also provide a list of the 30 countries at highest risk of onset in the first half of 2016. This list was submitted to the Evidence in Governance and Politics network as part of a formal pre-analysis plan—the first of its kind in the forecasting literature—and will be updated every six months.

1 Introduction

Outbreaks of civil war are rarely a surprise. They tend to be more like volcano eruptions than earthquakes, with low-level violence beginning long before conflict crosses the threshold usually associated with civil war. Can we model these dynamics of escalation in the months preceding onset? If so, then we could anticipate not only where but also when civil wars are most likely to occur.

We answer this question using event data from the World-Wide Integrated Crisis Early Warning System (ICEWS). Event data is unique in that it allows us to forecast over intervals much shorter than was previously possible, and to capture escalation dynamics that may change on a monthly, weekly or even daily basis. Most previous attempts to forecast civil war have been based on what we might call a country's "fundamentals"—structural characteristics such as per capita GDP, natural resources, regime type, or ethnolinguistic fractionalization.[1] Some of these variables have been shown to be statistically significant correlates of civil war, but most are poor predictors (Ward, Greenhill and Bakke 2010).

This is not surprising. Structural characteristics change slowly, if at all, and so cannot capture the rapidly evolving dynamics of escalation that typically precede the onset of civil war. Event data allows us to model escalation with a degree of precision that is impossible to replicate using structural characteristics alone.[2]

This process of escalation remains poorly understood and severely understudied. In an early review, Brubaker and Laitin (1998, 426) noted the "lack [of] strong evidence showing that higher levels of conflict (measured independently of violence) lead to higher levels of violence. Even where violence is clearly rooted in preexisting conflict, it should not be treated

[1]For a recent exception, see Montgomery, Hollenbach and Ward (2012).

[2]Systematically addressing these dynamics is the main motivation behind the "contentious politics" research program led by McAdam, Tarrow and Tilly (2001). Dynamic perspectives on conflict escalation have been proposed in the social movement literature (Lichbach 1987; Tarrow 1989, 1998), but few of them are quantitative and most are unsuitable for forecasting.

as a natural, self-explanatory outgrowth of such conflict, something that occurs automatically when the conflict reaches a certain intensity, a certain 'temperature.' Violence is not a quantitative degree of conflict but a qualitative form of conflict, with its own dynamics." Almost twenty years later, Brubaker and Laitin's claim still rings true: there is little empirical evidence demonstrating that higher levels of social conflict predict higher levels of violence, or that higher levels of violence predict the onset of civil war. Our goal is to help fill this gap.

We make five contributions. First, we show that ICEWS can be used to predict the onset of civil war remarkably accurately, with an area under the curve (AUC) ranging from 0.83 to 0.85, and over considerably shorter intervals (1-month or 6-month) than was possible in earlier models (e.g. Goldstone et al. 2010). We view these results as especially promising given the rarity of the dependent variable when operationalized at the sub-annual level. Our model relies on random forests, an algorithm that has the dual advantages of being well established in the forecasting and machine learning literatures but also easily comprehensible for non-specialists in both the academic and policy communities. Recent research has also shown that random forests outperforms more familiar logit models when forecasting civil wars using structural characteristics (Muchlinski et al. 2015).

Second, we show that a relatively parsimonious model informed by prominent theories of escalation (especially Lichbach 1987) performs better than more atheoretical alternatives. While this may seem unsurprising, it is not at all obvious that theory is necessary for forecasting given the availability of "big" datasets like ICEWS and the ease of implementing reliable but computationally intensive machine learning techniques for algorithmic (and thus atheoretical) variable selection and parameter estimation. Even models based on event data have tended to rely on coarse aggregations of often wildly disparate events into "Quad" counts (Leetaru and Schrodt 2014) or "Goldstein" scales (Goldstein 1992). Indeed, this approach is now sufficiently common that the publicly-available version of ICEWS contains both Quad counts and Goldstein scales for every event in the dataset.

We show, however, that a model composed of "made-to-order," theoretically-motivated predictors—coded to capture the dynamics of escalation with as much conceptual precision as possible—outperforms models based on Quad counts or Goldstein scales alone. Our model also outperforms an even more atheoretical alternative, in which we attempt to forecast civil wars using *all* events between government, opposition and rebel groups in the ICEWS dataset—over 1,000 predictors in all. To our knowledge, ours is the first attempt to directly compare the performance of theoretical and atheoretical approaches to predicting civil wars (or any other outcome in political science, for that matter).

Forecasting is sometimes criticized as an exercise in data mining with no relevance for social science theory. While our results are by no means conclusive—there are many more approaches to prediction, both theoretical and atheoretical, than the ones we test here—they suggest the possibility of a closer connection between theory and forecasting than is typically assumed, even in an era of big data and virtually limitless computational power.

Moreover, several of the patterns we observe in the relative importance of our theoretically-motivated predictors are consistent with the logic of escalation. For example, we find that low-level violence between government and opposition groups is by far the most reliable predictor of civil war onset, but that its importance relative to the other predictors is much more marked over 1-month rather than 6-month windows. This is consistent with the logic of escalation, as we would expect low-level violence to become an increasingly important predictor of higher-level violence as onset draws near. We also find that events occurring between government and opposition groups are generally more important than those occurring between government and rebels. This, too, is consistent with the logic of escalation: conflict between government and opposition should precede conflict between government and rebels, which should occur primarily during (rather than before) civil war.

Third, we assess the value-added of structural characteristics in a model otherwise composed of ICEWS-based predictors alone. Other studies have found that structural characteristics can be used to identify countries that are "immune" from risk in the context of split-population models (Beger, Dorff and Ward 2016). Our results confirm these findings. We compare five approaches to incorporating structural characteristics and show that a random forests analog to split-population modeling is the best of these alternatives. More surprisingly, however, we find that the inclusion of structural characteristics only marginally improves the performance of models based on ICEWS alone. We conclude that over these shorter forecasting windows, process may substitute for structure, providing all the information needed to forecast accurately.

Fourth, in addition to more conventional performance metrics (like the AUC), we attempt to present our results in ways that we believe should be more useful for practitioners and policymakers, and consider alternate analytical approaches that might be used to help determine whether a change in risk from one period to the next is significant enough to warrant attention and, potentially, intervention.

Finally, we use our model to generate a list of 30 countries at highest risk of civil war in the first half of 2016. "True" prospective forecasting of this sort remains rare in political science, even among forecasters themselves (Montgomery, Hollenbach and Ward 2012), who typically evaluate their models retrospectively, often through cross-validation. This leaves open the possibility of "fishing" for results—ensuring accuracy by over-fitting through multiple unreported rounds of training and testing on the same dataset. Following the lead of experimental social scientists in attempting to safeguard against these risks, we specified our predictions and the parameters of our model in a pre-analysis plan submitted to the Evidence in Governance and Politics (EGAP) network.[3] Every six months we will update our predictions for the next six month window and retrospectively evaluate the accuracy of our predictions from the previous window.

To our knowledge this is the first effort (at least within political science) to pre-register

[3]To preserve anonymity, we have requested that our pre-analysis plan not be listed on EGAP's registry until the peer review process is complete. A current copy of the plan is included in the appendix.

forecasts in a formal, public-available scholarly venue, and in a way that cannot be subsequently revised without documentation and approval by a third party. We hope this will serve as impetus for other forecasters to do the same, thus helping to ensure that the accuracy of our models can be easily and transparently evaluated over time.

2 Theories of civil war onset

2.1 Dynamics of escalation

When do claim-making groups resort to civil war to redress grievances or satisfy political goals? The literature offers a number of intuitions; we leverage three in particular. First, violence is often a response to repression by governments, including non-violent restrictions on group rights. Second, conflict typically simmers at low levels long before it boils over into civil war. And third, governments can interrupt this process of escalation by accommodating group demands, thus (potentially) preventing civil war from occurring. We explore these insights in turn, then use ICEWS to code proxies for the dynamics they describe.

Why might (non-violent) government repression provoke opposition groups to violence? There are multiple mechanisms and no consensus in the literature.[4] Most directly, both violent and non-violent repression may exacerbate grievances, thus intensifying the risk that violence will occur (Eckstein 1965). More indirectly, non-violent repression may radicalize the public, widening the pool of potential recruits willing to use violence against the state (Goldstein 1983; Kitschelt 1985). Alternatively, repression may deter moderates while simultaneously convincing radicals of the futility of non-violence (Gurr 1970; Tarrow 1989). The relationship between rebellion and repression is mutually reinforcing. As Gurr and Moore (1997, 1085) argue, repression and mobilization evolve simultaneously and in response to

[4]For a review of early empirical results in the literature on U.S. protest movements, see Wilson (1976). For a more recent review—part of an excellent study of protest movements in Western Europe—see Della Porta (1995).

one another, and jointly determine subsequent levels of rebellion.

An early attempt to reconcile competing hypotheses is Lichbach's (1987) model of escalation. In this model, groups want to redress grievances, and will choose the strategy that they perceive to be most effective and least costly to that end. Lichbach's key insight is that these groups must minimize fixed costs[5] and variable costs[6] simultaneously. Governments affect the repertoire of violent and non-violent strategies that these groups are likely to use by changing the mix of fixed and variable costs.

Lichbach's model generates predictions about the conditions under which repression will cause escalation (e.g. if it is used indiscriminately, it will mobilize previously non-violent groups; if it is used selectively without accommodation, it will increase mobilization among previously mobilized groups). Lichbach's model also predicts that accommodating non-violent groups might prevent escalation—an idea that has received empirical support. Inclusion of previously marginalized ethnic groups has been shown to reduce the risk of ethnic war (Cederman, Wimmer and Min 2010), and concessions of autonomy have been found to expedite resolution of active conflicts (Cederman et al. 2015). Appeasement could fail, however, if concessions are perceived as a sign of state weakness that emboldens the opposition (cf. Walter 2006).[7]

These models suggest a variety of mechanisms through which repression and accommodation might cause conflict escalation or de-escalation, respectively. We do not adjudicate

[5]For example, the costs of maintaining the group, as well as costs incurred from the more general level of state repression in the country.

[6]For example, physical damage to the group that results from specific acts of state repression.

[7]We are not aware of an empirical study of appeasement that takes into account the confounding effects of mediating variables such as the parties' relative power. These mediating variables depend on the shifting balance of power between governments and rebels as conflict unfolds. The net effect of accommodation is complicated by the fact that some concessions may be seen as non-credible, especially in the absence of an external enforcer. These dynamics are hard to model and such an analysis is beyond the scope of this paper. We do, however, incorporate their basic logic by testing if actions that may be perceived as "accommodative" predict de-escalation.

between these mechanisms. Rather, we construct proxies for several equally plausible mechanisms and test their joint predictive power.

2.2 Process over structure

The logic of Lichbach's model is inherently dynamic: conflict escalates (or not) depending on how the state responds to group claims and, in turn, how groups react to state policies. Most models of civil war onset, however, do not attempt to capture this dynamic and ignore the pattern of escalation in the weeks or months preceding onset.[8] A typical empirical strategy is to place countries into one of two bins—those with a civil war and those without—and attempt to predict shifts from one bin to the other from changes in structural characteristics (e.g. GDP or regime type). Each bin includes countries (or country-years) with wildly different levels of conflict and violence, and in many cases the only thing separating positive from negative cases is that they happen to fall on one side or the other of a rather arbitrary threshold that marks the onset of civil war.

This approach can be useful for some applications—for example, it can highlight the large gap in overall civil war risk in liberal Western democracies on the one hand, and politically and economically underdeveloped countries on the other. It cannot, however, explain why or when countries cross the threshold from violence to civil war, or at what level of conflict the alarm bells should go off for those concerned with predicting onset.

Here we adopt a more process-focused approach. As long as our aim is prediction, we need not be concerned with the fact that violence or other indicators of conflict in the months prior to onset are likely correlated with some of the explanatory variables that are usually

[8]Again, there are exceptions. Efforts to model the transition between levels of ethnic conflict date at least to Gurr's (2000) analysis of ethnopolitical conflict, from which the Minorities at Risk project emerged. Gurr found that conflict between the state and organized, ethnic-based political movements tends to correlate with ethnic rebellion later on. This echoes lessons learned from the process-focused literature on contentious politics (McAdam, Tarrow and Tilly 2001). Our analysis is motivated by these approaches.

included in models of civil war; indeed, we expect this to be the case. Our goal is to test whether a model grounded in theories of escalation can produce more reliable predictions than structural or atheoretical alternatives. Demands by organized groups, non-violent repression and low-level violence all indicate the presence of policy disagreements between the state and opposition groups that could turn those groups into rebel organizations. The more demands these groups make; the more the government represses them and refuses to accommodate them; and the more incidents of low-level violence we observe, the more severe we expect the risk of civil war onset to be.

3 Data

Our predictors are drawn from two sources: ICEWS and the Political Instability Task Force (PITF). Our dependent variable is drawn from the Sambanis (2004) dataset on civil war onsets, updated through the end of 2015.[9]

We aggregate the ICEWS dataset into event counts at the country-month level, with actors classified into one of three categories: government,[10] rebels[11] or opposition.[12] We limit our analysis to events that occur between two actors within the same country, and to the period beginning January 1, 2001[13] and ending December 31, 2015. For the period

[9]Because South Sudan did not exist as a country until after the end of our test set, we exclude it from the analysis. We do, however, include Sudan.

[10]See the appendix for a list of all actors classified as government. These include "Executive Office," "Defense/Security Ministry," "Military" and "National/Supreme Court," among others.

[11]Actors classified as "rebels" include "Radicals/Extremists/Fundamentalists," "Organized Violent," "Rebel," "Insurgents" and "Separatists."

[12]Actors classified as "opposition" include "Dissident," "Protestors/Popular Opposition/Mobs," "Exiles," "Opposition Major Party (Out of Government)," "Opposition Minor Party (Out of Government)," "Opposition Provincial Party (Out of Government)," "Opposition Municipal Party (Out of Government)" and "Banned Parties."

[13]We opt not to use ICEWS data prior to January 1, 2001 based on advice from the PITF, as the actor and event dictionaries on which ICEWS is based were overhauled in 2000.

through 2013 we use the public version of ICEWS posted on March 18th, 2015. For 2014 and 2015 we use an embargoed version of the dataset made available to us through the PITF.[14]

The dependent variable throughout is civil war onset, coded 1 for each country-month in which a civil war began, 0 for each country-month of peace, and missing for each country-month of ongoing conflict (excluding the month of onset). Operationalized in this way, onsets are exceedingly rare events, occurring in just 20 of the 29,520 country-months in the dataset—an incidence of less than a tenth of a percent.

Civil wars are macro-level concepts defined in terms of cumulative levels of violence over a period of time. As such, predicting them is inherently problematic, particularly sub-annually, since we typically do not have access to high-frequency data on violence and there is often uncertainty about the precise point in time that violence crosses the threshold of civil war. Moreover, depending on the rules for coding termination, a given conflict with long pauses in violence could be coded as a single civil war or as several smaller, shorter ones, resulting in several different onsets. We have addressed this issue as best we can by using the dataset with the most explicit set of coding rules in the literature, and one which identifies the month of onset in most cases.[15]

[14]The 2014 and 2015 versions of the dataset include some back-coded observations for previous years as well. To ensure model stability, we exclude these observations from our analysis.

[15]As casualty data is now being collected more systematically and at higher frequency, future work might consider abandoning the concept of civil war altogether, focusing instead on forecasting spikes in the level of violence beyond some threshold set by the researcher. This would have the advantage of combining various forms of violence (civil wars, riots, mass atrocities, coups) that are for the most part treated as distinct from one another without proper justification or understanding of why violence is organized as a civil war as opposed to some other form. For more on the challenges of conceptual aggregation in forecasting, see Goldstone et al. (2010) and Montgomery, Hollenbach and Ward (2012).

3.1 Operationalizing escalation

We use ICEWS to construct four sets of predictors designed to capture the dynamics of escalation described above. Our predictors are operationalized as event counts at the country-month level: low-level violence; non-violent repression; demands; and accommodations. (See the appendix for the CAMEO codes comprising each predictor.)

The variable *low-level violence* includes all events in which (1) government is the source and opposition is the target; (2) opposition is the source and government is the target; (3) government is the source and rebels are the target; or (4) rebels are the source and government is the target. *Non-violent repression* and *accommodations* include only events in which government is the source and either opposition or rebels is the target, and *demands* includes only events in which either opposition or rebels is the source and government is the target.

Some caveats are warranted. First, CAMEO codes do not perfectly correspond to the concepts we wish to capture, and one could try different aggregations than the four we propose. Exhaustively comparing the performance of different permutations of these variables may prove a worthwhile exercise, but is beyond the scope of this paper. Second, there is enough overlap between categories that some events might be included as components of two or more variables simultaneously.[16] There are many variations on our approach that one might entertain, but we do not explore them here.

Third, some of the events included as components of the predictors above occur infrequently or not at all in the ICEWS dataset—at least not in the years or among the actors

[16] For example, CAMEO code 124 ("Refuse to yield, not specified below") might be classified as non-violent repression when the source is government and the target opposition, or perhaps as a demand when the source is opposition and the target is government.

we consider.[17] This is probably the result of a fourth concern about idiosyncrasies in the text-scraping algorithm used to populate the ICEWS dataset, which may misclassify actors, event types or both. We discuss this and related problems below.

3.2 Assessing data quality: Iraq 2006 as a case study

How serious are these concerns about data quality? To explore the nature of measurement error in the ICEWS dataset—and, in particular, to assess the dataset's accuracy and comprehensiveness in identifying events of low-level violence—we look more closely at the events coded for a single country (Iraq) in a single year (2006). This is an exploratory exercise, and is intended only to help fix ideas on the limitations and idiosyncrasies of the ICEWS dataset. We focus on the coding of events surrounding the bombing of the al-Askari Mosque in the city of Samarra on February 22, 2006. This event is often referred to as the start of Iraq's sectarian civil war.

ICEWS data for the first three months of 2006 show a pattern of escalation. Daily events coded range from a low of 25 on January 10th to 221 on February 22nd. The bombing is the first event coded on February 22nd: it is referred to indirectly in a BBC report of a demonstration called to condemn the bombing of the shrine. The next four rows in ICEWS are also about the same event, classified in two different ways: three are classified under "demonstrate or rally" (two of which are identical listings of the same headline) and one is classified under "make appeal or request." The bombing of the shrine is reported directly as an event of "use conventional violence" on the fifth row, though the target is listed as an "Islamic Preacher" rather than the broader Shiite community; and the same text report is repeated as a different form of violence ("conduct suicide") with a different target ("Iraq").

[17] For example, there is no case of rebel groups violently protesting for leadership change against government (CAMEO code 1451) in the dataset. Similarly, there is only one case of government easing a ban on political parties organized by the opposition, and no cases of government easing a ban on political parties formed by rebels (CAMEO code 0812).

About half of all events coded on that day are in some way related to the bombing, though many are duplicates of one another. Duplication takes several forms: the same event may be reported under different "event text" categories,[18] or the same (or similar) headlines may appear in two or more of the text sources from which the raw data are scraped. Casual inspection suggests that more than 75% of reports on this day are repeated 2 to 3 times and in some cases as many as 8 times.

Some events are also misclassified. For example, a headline about Muqtada al Sadr "cutting short a visit to Lebanon to return home after the bombing" is classified as an instance of "reducing or breaking diplomatic relations." There are 12 separate reports related to this event (e.g., reports of Muqtada crossing the border, or of the number of days his visit to Lebanon was supposed to last). In this case, the actions of a Shia leader are mischaracterized as reflecting the Iraqi government's diplomatic relations with its neighbors; further, that leader's response to a crisis at home is mischaracterized as a rupture in diplomatic relations abroad. Although duplication occurs across all "Event Text" categories, misclassification is more of a problem for some than for others. For example, "unconventional violence" (and other "low-level violence" categories that we use in our forecasts) are less likely than other categories to include content that is obviously unrelated to the escalation or de-escalation processes that they intend to capture.

Finally, one limitation of the data is that it cannot yet be used to code dynamic action-reaction sequences between pairs of actors. Actions and reactions are not tied together in a way that would allow focused analysis of tit-for-tat escalation processes. For example, the Samarra mosque bombing in Iraq provoked a number of reprisal killings of Sunni clerics, but these reactions cannot be connected to the bombing except by reading the relevant headlines

[18] For example, the Samarra mosque bombing is reported as "Use unconventional violence" (at least 5 separate reports); "Conduct suicide, car, or other non-military bombing" (at least 4 separate reports); "Make statement" (at least 3 separate reports). These counts do not include statements of blame attribution or other characterizations of the attack, nor do they include reports of actions that were a result of the bombing.

and hand-coding the connections—a prohibitively onerous task given the size of the dataset.

We discuss data issues further in section 6. But, overall, despite these limitations, a pattern of escalation is obvious in the Iraq data when one reads the headlines and event sentences scraped from the data sources. The bombing prompted diplomatic initiatives and statements, changes to planned meetings between political leaders, demonstrations and riots, violent retaliations, and increased security patrols to prevent further violence escalation. Although some events are duplicated and misclassified, the overall picture that emerges is of a country on the verge of escalating violent conflict.[19]

4 Empirical strategy

4.1 Models

Do we need theory to predict civil war? The question is by no means rhetorical. Even strong statistically significant correlates of civil war prove to be weak predictors, and good theories may nonetheless produce bad forecasts (Ward, Greenhill and Bakke 2010). It is not obvious that theoretical models will outperform atheoretical ones, especially given vast computational power and sophisticated machine learning techniques to inductively identify empirical patterns in "big" data. Moreover, the noise in datasets like ICEWS should privilege atheoretical over theoretical approaches: if events are misclassified or duplicated, then we should expect "made-to-order," theoretically-motivated predictors to perform poorly. We thus view ICEWS as a relatively hard test for the value added of theory for prediction. To our knowledge, ours is the first attempt to conduct such a test in political science.

Our analysis consists of the following five models:

1. The *escalation* model, composed of country-month event counts for (1) low-level vio-

[19]Instead of being perceived as a type of measurement error, it might be possible to use the amount of duplication of an event as an importance weight that could inform the analysis of conflict escalation. We leave this for future studies to explore.

lence, (2) non-violent repression, (3) demands and (4) accommodations, as defined in Section 3.1 above.

2. The *Quad* model, composed of country-month "Quad" counts for (1) material conflict, (2) material cooperation, (3) verbal conflict and (4) verbal cooperation (Leetaru and Schrodt 2014).

3. The *Goldstein* model, composed of country-month Goldstein scales that take values between -10 and 10 depending on how "cooperative" or "conflictive" the corresponding event is deemed to be, where -10 is the most conflictive and 10 is the most cooperative (Goldstein 1992).

4. The *CAMEO* model, composed of country-month counts for all events in the CAMEO codebook—1,159 predictors in all.

5. The *average* model, which generates predictions based on the unweighted average of the four models above.

The *Quad*, *Goldstein* and *CAMEO* models incorporate all event types in the ICEWS dataset (albeit in different ways), and thus constitute more atheoretical benchmarks against which to assess the *escalation* model. We include the *average* model as well because ensembles of predictions gleaned from multiple models often outperform those of a single model alone (Montgomery, Hollenbach and Ward 2012). For the *Quad*, *Goldstein* and *CAMEO* models we include all events in which (1) government is the source and opposition is the target; (2) opposition is the source and government is the target; (3) government is the source and rebels are the target; or (4) rebels are the source and the government is the target.

In addition to the five models listed above, we explore various approaches to combining ICEWS-based predictors with structural characteristics gleaned from the PITF dataset. We discuss these alternatives in Section 5.4.

4.2 Estimation

The number of techniques available for forecasting rare events is immense and growing. We opt for random forests. Random forests are ensembles of classification and regression trees (CART), and are now standard in the machine learning literature (though they have been slow to emerge in political science; see Muchlinski et al. 2015 for a recent exception). We offer an intuitive explanation of random forests here; for technical details, see Breiman (2001).

Intuitively, random forests allow researchers to interpret predictors in the same way doctors interpret symptoms and risk factors for disease: primary symptoms are used to make preliminary diagnoses, and secondary and tertiary symptoms are used to confirm hunches drawn from primary ones. For example, when emergency responders attempt to identify a heart attack victim, they may ask first about pain in the chest, then shortness of breath, then pain in other parts of the body. The primary symptom (pain in the chest) is the single best predictor of the underlying condition, allowing early responders to make as accurate a diagnosis as possible with as little information as possible. Secondary (shortness of breath) and tertiary (pain in other parts of the body) symptoms are used to confirm that diagnosis.

Classification and regression trees work in exactly this way. They search first for a "primary" risk factor that most effectively distinguishes positives (in our case, onsets of civil war) from negatives (no war). This first split partitions the data into two subsamples, called "nodes." At each node, CART then identifies secondary and tertiary risk factors to further improve the reliability of the "diagnosis." Each observation is passed down the tree until it reaches a terminal node, at which point a prediction is made. Random forests iterate this process over many random subsamples of the data and many random subsets of predictors, increasing stability and obviating the need for cross-validation (Siroky 2009).

Random forests have several additional properties that are especially attractive for our purposes. First, they can accommodate hundreds of highly collinear predictors (e.g. those in the *CAMEO* model described above). Second, they are inherently interactive, as the importance of any one predictor in any given tree depends on all the other predictors at higher nodes in the tree. Finally, random forests have the dual advantages of being well-established and widely-used in the machine learning literature but also intuitive and thus relatively easy to understand, even for non-specialists.

We use random forests with five terminal nodes and 100 observations (sampled without replacement) per tree, and with 100,000 trees per forest. We train each model on a subset of the data beginning January 1, 2001 and ending December 31, 2007, forecasting over intervals of either one or six months. We test the models on the remaining subset of the data. In the appendix we assess the robustness of our results to a variety of different values for these parameters.[20] While the absolute performance of our models is somewhat sensitive to parameter changes, their relative performance is not.

5 Results

5.1 Model performance

Table 1 provides the area under the curve (AUC) for each of our five models over 1-month (top panel) and 6-month (bottom panel) forecasting windows. AUC is the area under the Receiver Operating Characteristic (ROC) curve, which plots the true positive and false positive rates of a given model across many "discrimination thresholds" (the minimum predicted probability above which a country-month is classified as a 1). The higher the AUC, the better the model. Figure 1 presents the ROC curves themselves (smoothed for ease of interpretation).[21]

[20]We assess robustness to the number of terminal nodes in each tree (five or ten), the number of observations in each tree (100 or 500), the number of trees in each forest (100,000 or 1,000,000) and the date separating the training from the test sets (January 1, 2008 or January 1, 2010).

[21]Different studies use different metrics to evaluate predictive performance. Montgomery, Hollenbach and Ward (2012), for example, use Brier scores (in addition to AUCs), computed as the average squared deviation of all predicted probabilities from the observed values of the dependent variable (Brier 1950). In our case, because the dependent variable is so rare and the predicted probabilities so small, the Brier score is indistinguishably close to zero for all models. Goldstone et al. (2010) report results at the discrimination threshold that equalizes

Table 1: Out-of-sample AUCs

1-month forecasts

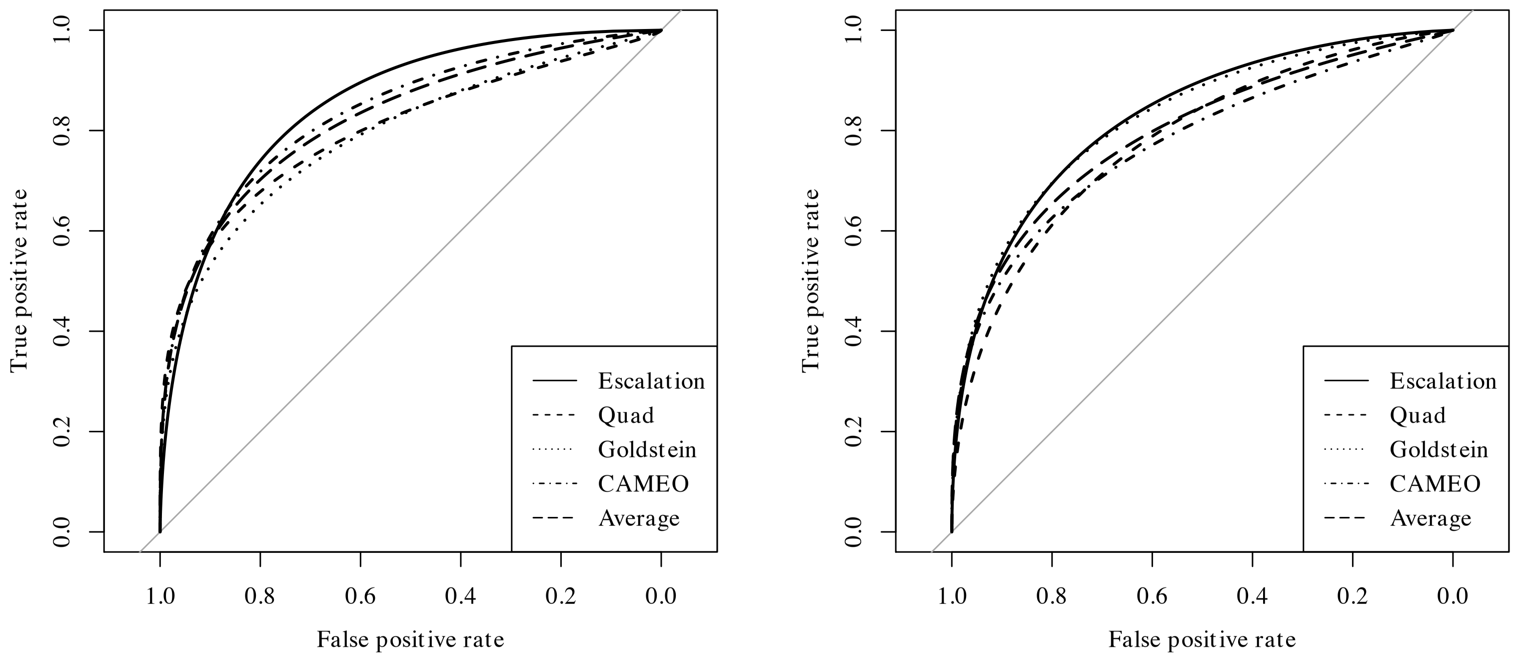
Escalation	*Quad*	*Goldstein*	*CAMEO*	*Avg.*
0.85	0.80	0.79	0.84	0.82

6-month forecasts

Escalation	*Quad*	*Goldstein*	*CAMEO*	*Avg.*
0.83	0.78	0.83	0.77	0.79

Notes: Performance of random forests models with five terminal nodes and 100 observations (sampled without replacement) per tree, and 100,000 trees per forest. The training set begins January 1, 2001; the test set begins January 1, 2008 and ends December 31, 2015. The top and bottom panels report results from 1-month and 6-month forecasts, respectively.

Figure 1: Out-of-sample ROC curves



Notes: Receiver operating characteristic (ROC) curves for random forests models with five terminal nodes and 100 observations (sampled without replacement) per tree, and 100,000 trees per forest. The training set begins January 1, 2001; the test set begins January 1, 2008 and ends December 31, 2015. The left and right plots show results from 1-month and 6-month forecasts, respectively.

The performance of the *escalation* model equals or exceeds that of the other models over both 1-month and 6-month forecasting windows. Over 1-month windows, the AUC for the *escalation* model is 0.85, compared to 0.80, 0.79, 0.84 and 0.82 for the *Quad*, *Goldstein*, *CAMEO* and *average* models, respectively. The discrepancy is especially marked at higher sensitivity rates (i.e. further to the right along the ROC curve's x-axis)—an important result in itself, given the low ratios of true positives to false positives that typically characterize forecasts of rare events, especially at high sensitivity. Moreover, while the differences in the 1-month AUCs are not dramatic, the rarity of the dependent variables means that achieving even small gains in sensitivity requires tolerating large losses in specificity and accuracy. Under these conditions, a gain of even a few points in AUC is an important improvement.

The *escalation* model outperforms the *Quad* and *CAMEO* models over 6-month windows as well, and performs about as well as the *Goldstein* model (with an AUC of 0.83). Importantly, while the *escalation* and *CAMEO* models perform similarly well over 1-month windows, the former outperforms the latter by six points over 6-month windows, and is thus the more reliable and versatile of the two. While the *average* model equals or exceeds the *Quad* and *CAMEO* models over both 1-month and 6-month windows (and exceeds the *Goldstein* model over 1-month windows as well), it too underperforms the *escalation* model.

The other models are also much less consistent than the *escalation* model in their performance relative to one another. Over 1-month windows, the *CAMEO* model outperforms both the *Quad* and *Goldstein* models; over 6-month windows, the opposite is true. Over 1-month windows the *Quad* model (slightly) outperforms the *Goldstein* model; over 6-month windows, the opposite is true. The *escalation* model, in contrast, performs as well as or better than the others regardless of the window we choose. And as we show in the appendix, the *escalation* model continues to outperform across a variety of robustness checks, while

accuracy, sensitivity and specificity. In our case, however, the rarity of the dependent variable ensures that these three metrics are almost never exactly equalized, complicating comparison across models. We therefore opt to report AUCs only.

the relative performance of the others remains idiosyncratic.

Nor are these results an artifact of parsimony alone. As forecasters have long recognized, simple models often outperform more complex ones in out-of-sample tests (Hastie, Tibshirani and Friedman 2001). This may explain why the *escalation* model outperforms the *CAMEO* model: the latter is composed of over 1,000 predictors, the former only 10. It is possible that the *CAMEO* model would have performed better with *ex ante* pruning. Parsimony cannot, however, explain why the *escalation* model outperforms the *Quad* and *Goldstein* models, since those models share the *escalation* model's virtue of simplicity (indeed, *Goldstein* is the most parsimonious of the models we tested). While the differences in performance are small, they suggest that predictive power improves not only with a more parsimonious selection of variables, but with a more theoretically coherent selection as well.

5.2 *Escalation* model performance in country-months of onset

While AUCs are useful for assessing the overall predictive performance of our models, they tell us little about the models' performance in the cases that matter most—those in which a civil war actually occurred. Here we consider two approaches to assessing model performance in the countries that suffered civil war outbreaks between January 1, 2008 and December 31, 2015 (the duration of the test set). For simplicity, and because it outperforms the others, we focus on the *escalation* model alone.

Because our dependent variable is so rare, the predicted probability of onset in any given country-month is always low—typically below 1%, even in countries and months in which onsets actually occurred. To help contextualize these point estimates, we rank the change in predicted probability in each country-month of onset relative to (1) the predicted probabilities in all other countries in the same month and (2) the predicted probabilities in all other months in the same country. These two comparisons address two distinct but related policy questions: (1) around the world, *where* is the risk of onset likely to be highest, and (2) in countries believed to be at high risk relative to the others, *when* is the risk of

Table 2: Test set prediction rankings for 1-month *escalation* model

				Within-country			Cross-country		
	Year	Month	Pred.	Rank	Rank change	% months with ≥ change	Rank	Rank change	% countries with ≥ change
Georgia	2008	8	0.000	16	+7	20%	22	+3	15%
Nigeria	2009	7	0.003	3	+1	41%	4	+2	73%
Yemen	2010	6	0.013	3	+1	50%	1	+1	78%
Libya	2011	2	0.000	8	0	50%	28	-5	81%
Cote d'Ivoire	2011	2	0.001	4	+6	12%	14	+5	9%
Syria	2011	3	0.000	5	0	84%	22	+6	72%
Sudan	2011	7	0.015	1	+24	2%	1	+4	8%
Mali	2012	1	0.000	5	0	91%	19	+6	74%
DRC	2012	5	0.003	2	+11	5%	3	+16	1%
Libya	2012	6	0.004	4	+4	20%	4	+18	1%
Ukraine	2014	4	0.001	10	-7	93%	15	-5	91%

Notes: Change in ranked predicted probabilities for test set country-months of civil war onset derived from an *escalation* model with five terminal nodes and 100 observations (sampled without replacement) per tree, and 100,000 trees per forest. The training set begins January 1, 2001; the test set begins January 1, 2008 and ends December 31, 2015.

onset likely to be highest?

Table 2 presents results for the *escalation* model; results for the other models are in the appendix. Columns 5 and 8 (labeled "rank") indicate the ranked predicted probability of onset relative to all other months in the same country (column 5) or all other countries in the same month (Column 8). Columns 6 and 9 (labeled "rank change") indicate the change in rank from the previous month. Columns 7 and 10 report the percentage of all months in the same country (column 7) or the percentage of all countries in the same month (column 10) with a rank change greater than or equal to the one that occurred in each country-month of onset. The smaller the percentage, the greater the risk.[22] Dramatic rank changes are relatively rare in the *escalation* model, which makes columns 7 and 10 especially informative, and which is arguably another virtue of the *escalation* model relative to the others. (Dramatic rank changes are much more common in the *Quad*, *Goldstein* and *CAMEO* models; see the appendix).

We view this as an intuitive approach to interpreting risk and evaluating model performance. In both Sudan and the DRC, for example, the predicted probability in the month of onset is abnormally high and increasing relative to previous months, and relative to other countries in the same month as well. In these cases the model would have provided a strong, reliable signal in favor of intervention. The same is true, albeit to a lesser extent, in Cote d'Ivoire and Libya (in the case of the second civil war). In Syria, Mali and Ukraine, in contrast, the predicted probability in the month of onset is unusually low and either stable or decreasing relative to prior months. In these cases the model would have been misleading, signaling strongly but erroneously against concern.

Why are some onsets so much harder to predict than others? Consider, for example,

[22]For example, the first row indicates that in August 2008—the month the Georgian civil war began—the predicted probability of onset increased 7 ranks relative to all other months in Georgia, and 3 ranks relative to all other countries in August 2008. Georgia witnessed a rank change at least that large in 20% of all months in the test set; 15% of all countries witnessed a rank change that large in August 2008.

Syria in March 2011. The predicted probability of onset is low and stable relative to previous months, and low (though increasing) relative to other countries in the same month as well. This may be a result of ICEWS's reliance on media coverage, which, judging by the number of events reported, appears to be limited in the periods prior to onset but extensive thereafter. A country or period neglected in the media will generally be neglected in our model as well. (As we show in the appendix, none of the other models performs especially well in the case of Syria either.) Moreover, few analysts anticipated such quick and disastrous escalation in Syria, and most agree that the Assad regime provoked escalation by over-reacting to fears generated by the Arab Spring—a unique event for that region.

These disparities in difficulty arise not just across but also within countries. For example, February 2011 in Libya is ranked equal to or lower than most other months in the country, and lower than most other countries in the same month as well. Moreover, Libya's predicted probability of onset in February 2011 is declining relative to other countries. In contrast, June 2012 in Libya is ranked higher than most other months, and higher than most other countries in the same month. This may be a result of the fact that the risk of escalation al in Libya was low until the international intervention—an observation consistent with case-study accounts of the conflict, which escalated after the NATO intervention in March 2011 (Kuperman 2013).

Overall, what should we conclude about the performance of our model in country-months of onset? If we use both within- and cross-country rank changes as indicators, our model would have provided strong signals of high and increasing risk in at least four cases—Cote d'Ivoire, Sudan, DRC and Libya in 2012—and a weaker (but still accurate) signal in Georgia as well. The model would have (misleadingly) signaled decreasing risk in Ukraine in 2014 and Libya in 2011, and low but steady or increasing risk in Nigeria, Syria, Yemen and Mali. These false negatives are, of course, disappointing. Nevertheless, given that civil war is both rare and costly in terms of lives lost and livelihoods destroyed, providing an accurate, unambiguous warning in even two or three countries strikes us an important achievement.

5.3 Ranking of importance scores for the *escalation* model

The *escalation* model is composed of four categories of events designed to capture escalation dynamics in the months preceding civil war: low-level violence, non-violent repression, demands and accommodations. It also includes three sets of actors: government, opposition and rebels. Which of these events, and between which sets of actors, are the most reliable predictors of onset?

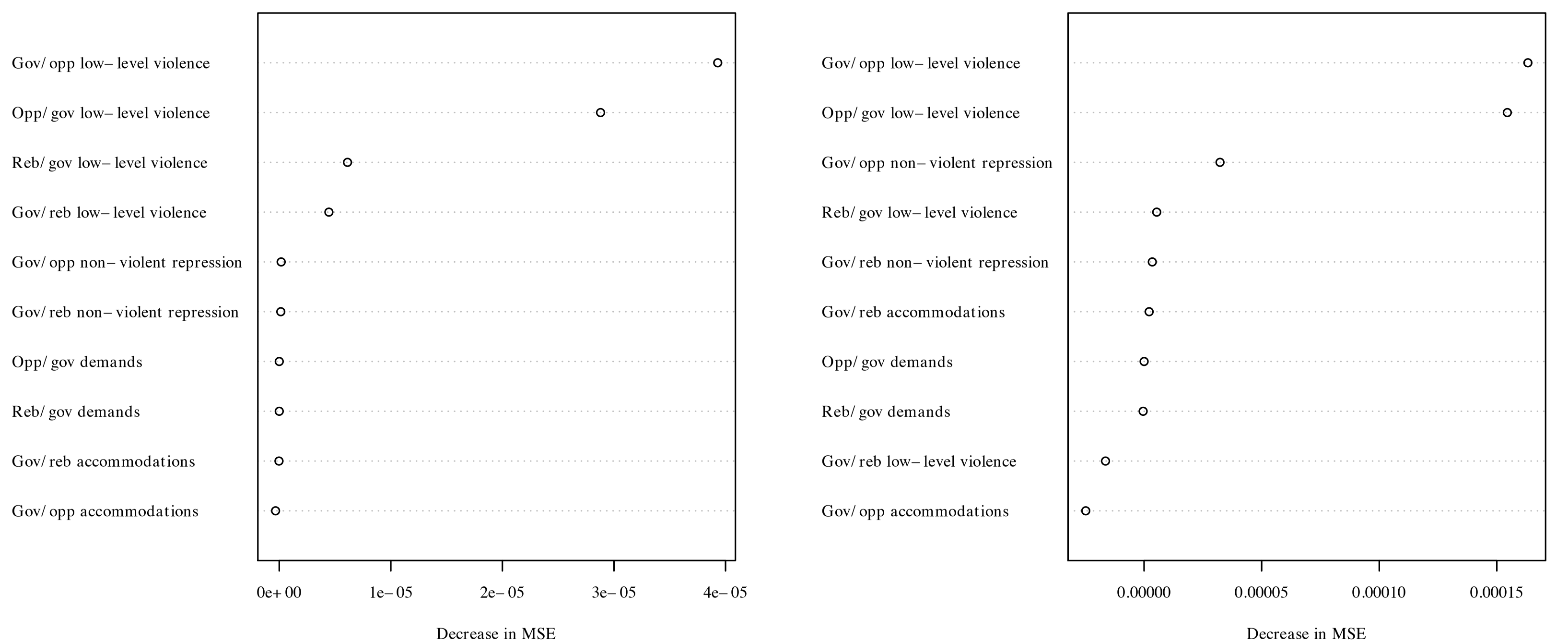
Figure 2 ranks the *escalation* model's predictors according by their "importance scores;" rankings for the other models are in the appendix. Importance scores are calculated by permuting the values of each predictor across observations in the training set, then estimating the impact of those permutations on "out-of-bag" accuracy over all trees in the forest (Breiman 2001).[23] Importance scores do not lend themselves to easy substantive interpretation, but their ordinal ranking and differences in magnitude are nonetheless informative signals of their relative predictive power.

Over both 1-month and 6-month windows, low-level violence between government and opposition groups is the most reliable predictor of civil war onset—far more important than any of the other predictors in our model. Low-level violence perpetrated by rebel groups against governments is a consistently important predictor as well, ranked third over 1-month windows and fourth over 6-month windows (though low-level violence perpetrated by governments against rebel groups is less important).

These patterns are consistent with the theories of escalation described in Section 2. That the importance of low-level violence relative to non-violent repression is more marked over 1-month rather than 6-month windows is consistent with the logic of escalation, as we would

[23]Following Strobl and Zeileis (2008), we report raw (unscaled) permutation-based scores rather than scaled permutation-based or Gini-based scores. Another option is to compute a modified version of the importance score that adjusts for random forests' tendency to prefer correlated predictors. We found this modification empirically intractable in a dataset this large (and, in the case of the *CAMEO* model, with this many predictors), and opt for the raw permutation-based score as the best available alternative.

Figure 2: Random forest importance score rankings for escalation model



Notes: Ranking of predictors from the *escalation* model based on random forests importance scores. Importance scores are estimated as the decrease in mean squared error for each predictor across 100,000 trees, with five terminal nodes and 100 observations (sampled without replacement) per tree. The training set begins January 1, 2001; the test set begins January 1, 2008 and ends December 31, 2015. The left and right plots show results from 1-month and 6-month forecasts, respectively.

expect low-level violence to become an increasingly reliable predictor of higher-level violence at increasingly shorter intervals of time, as onset draws near. Moreover, with the exception of demands, events occurring between government and opposition groups are consistently more important than those occurring between government and rebels. This, too, is consistent with the logic of escalation: conflict between government and opposition should be logically and empirically prior to conflict between government and rebels, which should occur primarily during (rather than before) civil war.

5.4 Combining process with structure

The models above rely on ICEWS alone, and thus do not incorporate any of the structural characteristics typically included in models of civil war onset, predictive or otherwise. While structural characteristics change too slowly to allow us to anticipate *when* conflict will occur (at least at the sub-annual level), they may nonetheless help us identify *where* it will occur in ways that ICEWS alone cannot. Opposition groups, for example, may make as many demands of government in high-risk countries as they do in low-risk ones, but low-risk countries may have structural characteristics more likely to result in accommodation, or may have conflict resolution mechanisms to prevent escalation. Identifying these low-risk countries *ex ante* could, in principle, improve the performance of our models.

To estimate structural risk we rely primarily on the PITF's "Internal Conflict Watch List." The Watch List provides predicted probabilities of civil war onset for all countries in the world, updated annually and derived from Goldstone et al's (2010) cross-national, case-control model. The model is composed of four structural characteristics: regime type,[24] infant mortality, prevalence of civil war in neighboring countries, and "state-led discrimina-

[24]Goldstone et al. use Polity scores to construct indicators for four regime types: partial autocracy, partial democracy with factionalism, partial democracy without factionalism, and full democracy.

Table 3: Out-of-sample AUCs for models including PITF

1-month forecasts

Escalation	With PITF predictors	With PITF predictions	Weighted by PITF	PITF high risk subset	PITF split population	PITF only
0.86	0.80	0.78	0.71	0.85	0.86	0.77

6-month forecasts

Escalation	With PITF predictors	With PITF predictions	Weighted by PITF	PITF high risk subset	PITF split population	PITF only
0.83	0.86	0.82	0.71	0.83	0.84	0.76

Notes: Performance of random forests models with five terminal nodes and 100 observations (sampled without replacement) per tree, and 100,000 trees per forest. The training set begins January 1, 2001; the test set begins January 1, 2008 and ends December 31, 2015. The top and bottom panels report results from 1-month and 6-month forecasts, respectively.

tion."[25] Goldstone et al. use conditional logistic regression to generate predicted probabilities from these structural characteristics. We use these predicted probabilities, as well as the structural characteristics themselves, as proxies for structural risk.

We consider five approaches to combining process (ICEWS) with structure (PITF). Table 3 reports our results, focusing, again, on the *escalation* model. Two related approaches, and perhaps the most obvious, are simply to include either the PITF predictors (column 2) or the Watch List predicted probabilities (column 3) as additional predictors in the *escalation* model. A third is to use the Watch List predicted probabilities to weight the predictions generated by the *escalation* model (column 4). This ensures that high-risk and low-risk countries that happen to take similar values on one or more of our ICEWS-based predictors will nonetheless receive different predicted probabilities of onset in most months. While the first and second approaches integrate PITF *during* the process of estimation, the third approach leverages information gleaned from PITF after estimation is complete.

The fourth and fifth approaches incorporate PITF before estimation begins, by dividing the countries on the Watch List into *ex ante* high risk and low risk categories. To do this, we first take the average PITF predicted probability for each country across all years in our training set. Comparing these averages across countries, we define those that fall in the bottom quartile of the distribution as "low risk," and the rest as "high risk." We then either run our model on the high risk subset alone (column 5) or on the high risk and low risk subsets separately, combining the results into a single forest composed of 200,000 (rather than 100,000) trees (column 6). For comparison Table 3 also reproduces the *escalation* model results from Table 1 (column 1) and includes results from a "dummy" model composed of the PITF predictors alone (column 7).

Of the approaches we test, the fifth—a random forests analog to split-population modeling—

[25]Goldstone et al. operationalize state-led discrimination as a score of 4 or higher on indices of political or economic discrimination for any racial, ethnic or religious group in the Minorities at Risk dataset.

appears to be the most reliable, with AUCs of 0.86 and 0.84 over 1-month and 6-month windows, respectively. These slightly exceed the AUCs of the *escalation* model alone. The first approach performs well over 6-month windows, but less so over 1-month windows. This makes sense, as slow-moving structural characteristics are likely to be more reliable predictors over longer rather than shorter intervals of time. The fourth approach appears promising as well, though fit statistics for this model are somewhat misleading, since countries deemed "low risk" in the training set are dropped from the analysis altogether, even if they suffer an outbreak of civil war in the test set. (This is the case for Libya.)

Overall, however, our results suggest that structural characteristics and related measures of structural risk do not reliably provide additional predictive power over that of the *escalation* model alone. Incorporating the PITF dataset either significantly reduces or only marginally improves the performance of the *escalation* model, regardless of the approach we take. This is counterintuitive, but may again be a result of the relatively short windows over which we generate our forecasts. The longer the window, the more informative slow-moving structural variables should be. Over shorter windows, process may substitute for structure, making structural variables superfluous for purposes of forecasting.

5.5 *Escalation* model forecasts for first half of 2016

Finally, Table 4 provides "true" prospective forecasts for the countries at highest risk of civil war in the first half of 2016, using predictors measured through December 31, 2015. For compactness and ease of exposition, we report results for the top 30 highest-risk countries only, based on cross-country rankings generated over 6-month windows using the *escalation* model alone. In raw terms, the risk of civil war in these countries ranges from a high of 7% (Nigeria) to a low of 1% (Ghana, Tajikistan and France, on which more below), following an approximately exponential rate of decay from the highest- to the lowest-ranked country on the list, making the top 30 an appropriate place to focus our attention. The predicted probabilities of onset become almost indistinguishably small lower down the list. Rankings

Table 4: 2016 prediction rankings for 6-month *escalation* model

Country	**Pred.**	**Country**	**Pred.**	**Country**	**Pred.**
Nigeria	0.073	Burundi	0.041	Iran	0.016
India	0.072	Egypt	0.039	Thailand	0.016
Iraq	0.071	Yemen	0.037	Myanmar	0.015
Somalia	0.071	Colombia	0.027	Montenegro	0.014
Syria	0.065	Mali	0.022	Niger	0.012
Pakistan	0.065	China	0.021	Bangladesh	0.012
Philippines	0.065	Indonesia	0.019	El Salvador	0.011
Turkey	0.064	Ukraine	0.019	Ghana	0.010
Afghanistan	0.064	Sudan	0.017	Tajikistan	0.010
Russia	0.047	Lebanon	0.017	France	0.010

Notes: Ordinal ranking of the top 30 countries at highest risk of civil war in the first half of 2016 based on predicted probabilities from the *escalation* model, using predictors measured through December 31, 2015. The list includes countries with ongoing civil wars in the second half of 2015; we interpret results for these countries as indicative of a high risk that civil war will persist.

from the other models are included in the appendix.

Importantly, we make no assumptions about the persistence of civil war between the second half of 2015 and the first half of 2016, meaning that all countries—even those with ongoing conflicts as of December 31, 2015—are included in the 2016 test set. As a result, 14 of the 30 countries on the list (and eight of the top ten, including numbers one through six) are cases of ongoing conflict. In fact, four of the countries on the list (India, Afghanistan, Russia and Colombia) are coded as ongoing conflicts throughout the duration of our dataset (which begins January 1, 2001), and so are excluded altogether from the analyses in Tables 1 through 3 above. These cases thus constitute a *de facto* out-of-sample test for our model, since none of the model's parameters were trained on data from these countries.

The results are encouraging. That even a single 6-month period of data (covering the second half of 2015) is sufficient to identify countries with ongoing conflicts as high risk suggests that the predictors we constructed indeed capture dynamics that are relevant to both the onset and persistence of civil war, and that random forests is indeed assigning appropriate weights to those predictors. Moreover, the fact that the list includes countries currently suffering severe, geographically-dispersed violence (such as Afghanistan and Syria) as well as those suffering milder, more geographically-isolated violence (such as India and Colombia) suggests that our predictors are capable of capturing a wide range of conflict dynamics as well.

There is much overlap between the list in Table 4 and the PITF Watch List for 2016. This, too, is encouraging. The PITF identifies 18 of the countries on our list as being "in crisis," including 16 of our top 20 and all of our top seven.[26] Of the remaining 12, the PITF classifies three (Indonesia, Iran and Bangladesh) as being "most" at risk, five (Turkey, Lebanon, Niger, Tajikistan and France) as being at "significant" risk, and three (Burundi, El Salvador and Ghana) as being at "some" risk. (The PITF classifies Montenegro as low

[26]The PITF uses a five-category classification system to characterize risk: (1) in crisis, (2) most risk, (3) significant risk, (4) some risk and (5) low risk.

risk.) These parallels are especially remarkable given that ICEWS is newer and noisier than the datasets on which the PITF relies (e.g. Minorities at Risk, which began in 1986).

Moreover, several of the countries that we rank higher than the PITF experienced episodes of instability in 2015 that might justify their inclusion on the list—episodes that the PITF model, by its nature, cannot capture. In Burundi, for example, President Pierre Nkurunziza incited a wave of popular protests and an attempted coup d'etat after announcing his intention to run for a third term in office, in defiance of constitutional term limits. In Montenegro, thousands of protestors demanded (and continue to demand) the resignation of the country's prime minister, resulting in repeated clashes with police in the capital city. And in Ghana, sporadic incidents of violence have occurred ahead of the general elections planned for later this year, including the killing of a member of the New Patriotic Party and a raid on the party's head office in Accra. These incidents are not indicative of impending civil war, and it is unlikely that any of these countries will experience an outbreak in the first half of 2016 (the predicted probabilities for Burundi, Montenegro and Ghana are .041, .014 and .010, respectively). Nonetheless, there are reasons to believe they belong on a list of countries to watch.

That leaves France as the most obviously inappropriate candidate for the top 30 list. While we cannot say for certain why our model assigns an unrealistically high predicted probability of onset to France, the text-scraping algorithm used to populate ICEWS does seem to record an abnormally high number of incidents of low-level violence there, especially relative to other rich industrialized nations.[27] Moreover, France ranks surprisingly high on the PITF Watch List as well, and is classified as being at "significant" risk of civil war. Whatever sources of risk our model is detecting, other models seem to be detecting them

[27]For example, ICEWS records 219 incidents of low-level violence inflicted by government on opposition groups in France between January 1, 2001 and December 31, 2015, and another 393 incidents of low-level violence perpetrated by opposition groups against government. In contrast, ICEWS records no such events in the United States over this period and, with the exception of Italy, considerably fewer in all of France's neighbors.

too—even if they are false alarms.

Together these results provide further evidence of the *escalation* model's value added over alternatives based on structural characteristics alone. If our results merely reproduced those of the PITF, or if the divergences between them were clear indicators of our model's inferiority, then even the possibility of higher-frequency forecasting would not suffice to justify the effort. This, however, is not the case. We view both the parallels and discrepancies between the two models as reason for confidence in ICEWS's promise.

6 Discussion

In this paper we assess whether ICEWS event data can be used to accurately forecast the onset of civil war, and whether ICEWS can surpass or improve the predictive power of existing models composed of structural characteristics alone. Civil wars are rare events, and we anticipated that accurate prediction would be difficult. We find, however, that ICEWS can indeed be used to generate relatively reliable forecasts, with AUCs ranging from 0.83 to 0.85 in our preferred model. We also find that ICEWS produces forecasts similar to those provided by the PITF, but over considerably shorter windows (1-month or 6-month). We view these results as promising, especially given the relative novelty of the ICEWS dataset and the rarity of the events we are attempting to predict.

We also find there is value-added in designing models guided by theory. Again, it is not obvious that this should be the case, especially given the noise in ICEWS, the availability of vast computing power and the proliferation of machine learning techniques for automated variable selection. Yet our preferred model, grounded in the theoretical literatures on conflict escalation, consistently outperforms more atheoretical alternatives. This finding is relevant not only for purposes of prediction, but also for the more general debate about the relationship between theory and forecasting in political science. Our findings are consistent with intuitions embedded in the contentious politics research program (McAdam, Tarrow and Tilly 2001), specifically the conclusion that process is more important than structure in certain settings. Though our approach is different, our conclusions about the importance of process over structure is consistent with Beissinger's (2002) classic study of nationalist mobilization in the Soviet Union that clearly demonstrated the path-dependence of protest activity. Our framework could be used to prospectively test propositions flowing from theories of conflict and repression (Davenport 2010; Lichbach 1987) and further the empirical foundations of the broader literature on contentious politics.

Of course, our models are not without limitations. One important shortcoming is our inability to link sequences of actions and reactions between actor dyads. This limitation arises from the structure of the ICEWS dataset, which does not record the possible connections between related events and thus cannot capture the tit-for-tat dynamics that typically characterize conflict escalation. Coding these connections would almost certainly require more intimate involvement of human coders at various stages of the data collection process. This may be a worthwhile investment for countries viewed as especially high priorities for monitoring and intervention. Our probe into the Iraq data, for example, suggests that important escalation and de-escalation dynamics are likely to be contained but not coded in ICEWS, and that there may even be a correlation between the severity of certain types of events (e.g. unconventional violence) and the amount of duplication in the data. ICEWS could be made suitable for dynamic modeling in a small selection of countries and/or years if sequences of related events could be connected in analytically tractable ways. The fact that many of these events already appear in the dataset suggests that such a proposition is at least feasible.

Finally, the history of the ICEWS project, as described briefly in the preface and introduction to the CAMEO codebook,[28] suggests inherent intellectual constraints that limit

[28]See https://urldefense.proofpoint.com/v2/url?u=http-3A__eventdata.parusanalytics.com_cameo.dir_CAMEO.Manual.1.1b3.pdf&d=AwIGaQ&c=-dg2m7zWuuDZ0MUcV7Sdqw&r=cqVPv_YiLydNJ3UKCqrhpBAJMnD7M0ODcVHHvOPt6e4&m=SoNhGu7CP7ihUerpEb-zHaJwJ2DNUP8dVxkmNUqoppk&s=02SRifGUHx-KkesNop2Gxu5VYH_

the dataset's usefulness for some applications. The CAMEO system reflects an interest in inter-state conflict and bargaining; as such, the dataset contains many variables related to mediation and negotiation, with the implicit assumption that these concepts apply equally to all parties involved in a conflict. This assumption, however, may not always be viable. For example, civil wars often feature a variety of informal or irregular forces that may not be "eligible" for inclusion in the CAMEO categories (which are premised on some degree of recognition), or that may defy easy classification. Furthermore, while in some cases these actors cohere into loose alliances that could be considered more-or-less unitary rebel or opposition groups, in others (perhaps the majority) they do not, making categorization difficult and fluid.

These are issues for ICEWS's designers to consider in future. In the meantime, our analysis suggests that ICEWS can be used to forecast civil wars with relatively high levels of accuracy, especially when informed by theories of conflict escalation. We view the combination of existing theories with relatively new event datasets like ICEWS as one of this literature's most interesting and promising frontiers.

`VT6S_LRF8hAPjZXE&e=`, accessed July 28, 2015.

References

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

Beissinger, Mark R. 2002. *Nationalist Mobilization and the Collapse of the Soviet State.* Cambridge: Cambridge University Press.

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

Brier, Glenn W. 1950. "Verification of Forecasts Expressed in Terms of Probability." *Monthly Weather Review* 78(1):1–3.

Brubaker, Rogers and David Laitin. 1998. "Ethnic and Nationalist Violence." *Annual Review of Sociology* 24:423–452.

Cederman, Lars-Erik, Andreas Wimmer and Brian Min. 2010. "Why Do Ethnic Groups Rebel? New Data and Analysis." *World Politics* 62(01):87–119.

Cederman, Lars-Erik, Simon Hug, Andreas Schädel and Julian Wucherpfennig. 2015. "Territorial Autonomy in the Shadow of Conflict: Too Little, Too Late?" *American Political Science Review* 109(02):354–370.

Davenport, Christian. 2010. *State Repression and the Domestic Democratic Peace.* Cambridge: Cambridge University Press.

Della Porta, Donatella. 1995. *Social Movements, Political Violence, and the State: A Comparative Analysis of Italy and Germany.* Cambridge: Cambridge University Press.

Eckstein, Harry. 1965. "On the etiology of internal wars." *History and Theory* 4(2):133–163.

Goldstein, Joshua S. 1992. "A Conflict-Cooperation Scale for WEIS Events Data." *Journal of Conflict Resolution* 36(2):369–385.

Goldstein, Robert J. 1983. *Political Repression in 19th Century Europe*. London: Croom Helm.

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

Gurr, Ted Robert. 1970. *Why Men Rebel*. Princeton: Princeton University Press.

Gurr, Ted Robert. 2000. *Peoples Versus States: Minorities at Risk in the New Century*. Washington, DC: United States Institute of Peace.

Gurr, Ted Robert and Will H. Moore. 1997. "Ethnopolitical Rebellion: A Cross-Sectional Analysis of the 1980s with Risk Assessments for the 1990s." *American Journal of Political Science* 41(4):1079–1103.

Hastie, Trevor, Robert Tibshirani and Jerome Friedman. 2001. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. New York: Springer.

Kitschelt, Herbert. 1985. "New Social Movements in West Germany and the United States." *Political Power and Social Theory* 5:273–324.

Kuperman, Alan J. 2013. "A Model Humanitarian Intervention?: Reassessing NATO's Libya Campaign." *International Security* 38(1):105–136.

Leetaru, Kalev and Philip A. Schrodt. 2014. "GDELT: Global Data on Events, Location and Tone, 1979-2012." Presented at the International Studies Association Annual Convention, San Francisco.

Lichbach, Mark. 1987. "Deterrence or Escalation? The Puzzle of Aggregate Studies of Repression and Dissent." *Journal of Conflict Resolution* 31(2):266–297.

McAdam, Doug, Sidney Tarrow and Charles Tilly. 2001. *Dynamics of Contention*. Cambridge: Cambridge University Press.

Montgomery, Jacob M., Florian M. Hollenbach and Michael D. Ward. 2012. "Improving Predictions Using Ensemble Bayesian Model Averaging." *Political Analysis* 20(3):271–291.

Muchlinski, David, David Siroky, Jingrui He and Matthew Kocher. 2015. "Comparing Random Forest with Logistic Regression for Predicting Class-Imbalanced Civil War Onset Data." *Political Analysis* 24:87–103.

Sambanis, Nicholas. 2004. "What Is Civil War? Conceptual and Empirical Complexities of an Operational Definition." *Journal of Conflict Resolution* 48(6):814–858.

Siroky, David S. 2009. "Navigating Random Forests and Related Advances in Algorithmic Modeling." *Statistics Surveys* 3:147–163.

Strobl, Carolin and Achim Zeileis. 2008. "Danger: High Power! – Exploring the Statistical Properties of a Test for Random Forest Variable Importance." University of Munich Department of Statistics Technical Report 017.

Tarrow, Sidney. 1989. *Democracy and Disorder: Social Conflict, Political Protest, and Democracy in Italy, 1965–1975.* New York: Oxford University Press.

Tarrow, Sidney. 1998. *Power in Movement: Social Movements and Contentious Politics.* Cambridge: Cambridge University Press.

Walter, Barbara F. 2006. "Information, Uncertainty, and the Decision to Secede." *International Organization* 60(1):105–135.

Ward, Michael D., Brian D. Greenhill and Kristin M. Bakke. 2010. "The Perils of Policy by P-Value: Predicting Civil Conflicts." *Journal of Peace Research* 47(4):363–375.

Wilson, John. 1976. "Social Protest and Social Control." *Social Problems* 24:469–481.

Appendix

CAMEO codes for predictors

We use the ICEWS dataset to construct four sets of predictors designed to capture the dynamics of escalation. These predictors are operationalized as event counts at the country-month level, and categorized as follows:

- ***Low-level violence***, comprising the following CAMEO events:

145:	Protest violently
1451:	Violent protest for leadership change
1452:	Violent protest for policy change
1453:	Violent protest to demand rights
1454:	Violent protest to demand change in institutions
170:	Residual violent repression category
180:	Use unconventional violence
183:	Non-military bombing; outside of military engagements
171:	Seize or damage property
175:	Use repression
186:	Assassinate
191:	Impose blockade; restrict movement
1952:	Remotely piloted aerial munitions
193:	Fight with small arms

- ***Non-violent repression***, comprising:

123:	Reject request or demand for political reform, not specified below

1232: Reject request to change policy

1233: Reject request for rights

1234: Reject request for change in institutions, regime

124: Refuse to yield, not specified below

1241: Refuse to ease administrative sanctions

1321: Threaten with restrictions on political freedoms

1322: Threaten to ban political parties or politicians

137: Threaten with repression

172: Impose administrative sanctions, not specified below

1721: Impose restrictions on political freedoms

- ***Demands***, comprising:

104: Demand political reform, not specified below

1042: Demand policy change

1043: Demand rights

1044: Demand change in institutions, regime

- ***Accommodations***, comprising:

035: Express intent to yield, not specified below

0342: Express intent to change policy

0343: Express intent to provide rights

0344: Express intent to change institutions, regime

034: Express intent to institute political reform, not specified below

080: Yield, not specified below

081:	Ease administrative sanctions, not specified below
0811:	Ease restrictions on political freedoms
083:	Accede to requests or demands for political reform, not specified below
0832:	Accede to demands for change in policy
0833:	Accede to demands for rights
0834:	Accede to demands for change in institutions, regime
0812:	Ease ban on political parties

Actors classified as "government"

Below is a list of all actors in the ICEWS dataset that we classify as "government:"

- Government
- Executive
- Executive Office
- Cabinet
- Foreign Ministry
- Interior/Home Ministry
- NGO Ministry
- Management/Budget/Planning/Organization Ministry
- Agriculture/Fishing/Forestry Ministry
- Finance/Economy/Commerce/Trade Ministry
- Industrial/Textiles/Mining Ministry

- Post/Telecoms Ministry
- Science/Tech Ministry
- State Owned Enterprises
- State-Owned Agricultural
- State-Owned Transportation
- State-Owned Utilities
- State-Owned Heavy Industrial/Chemical
- State-Owned Defense/Security
- State-Owned Durable Goods
- State-Owned Consumer Goods
- State-Owned Consumer Services
- State-Owned Consulting/Financial Services
- State-Owned Science/Tech/Knowledge/Innovation
- State-Owned Medical/Health/Pharmaceutical
- Women/Children/Social/Welfare/Development/Religion Ministry
- Education Ministry
- Energy Ministry
- Environment Ministry
- Transportation Ministry
- Food Ministry

- Disaster Ministry
- Health Ministry
- Science/Tech/Knowledge/Innovation Ministry
- Human Rights Ministry
- Elections Ministry
- Housing/Construction Ministry
- Justice/Law Ministry
- Tourism Ministry
- Drugs Ministry
- Labor Ministry
- Water Ministry
- Information/Communication/Transparency Ministry
- State Media
- Defense/Security Ministry
- Government Major Party (In Government)
- Government Minor Party (In Government)
- Government Provincial Party (In Government)
- Government Municipal Party (In Government)
- Government Religious
- Military

- Research and Design Wings
- Research and Design Wings Headquarters
- Research and Design Wings Education/Training
- Research and Design Wings Support
- Research and Design Wings Medical
- Army
- Army Headquarters
- Army Special Forces
- Army Infantry/Regular
- Army Mechanized (Ships, Tanks, Planes)
- Army Education/Training
- Army Support
- Army Medical
- Navy
- Navy Headquarters
- Navy Special Forces
- Navy Infantry/Regular
- Navy Mechanized (Ships, Tanks, Planes)
- Navy Education/Training
- Navy Support

- Navy Medical
- Air Force
- Air Force Headquarters
- Air Force Special Forces
- Air Force Infantry/Regular
- Air Force Mechanized (Ships, Tanks, Planes)
- Air Force Education/Training
- Air Force Support
- Air Force Medical
- Marines
- Marines Headquarters
- Marines Special Forces
- Marines Infantry/Regular
- Marines Mechanized (Ships, Tanks, Planes)
- Marines Education/Training
- Marines Support
- Marines Medical
- Coast Guard
- Coast Guard Headquarters
- Coast Guard Special Forces

- Coast Guard Infantry/Regular
- Coast Guard Mechanized (Ships, Tanks, Planes)
- Coast Guard Education/Training
- Coast Guard Support
- Coast Guard Medical
- Judicial
- National/Supreme Court
- Provincial Court
- Municipal/District Court
- Civil Court
- Religious Court
- Military/Tribunal
- Local
- Provincial
- Municipal
- Legal
- Parties (National)
- Major Party (National)
- Minor Party
- Provincial Party

- Municipal Party

Robustness checks

Table 5 reports robustness checks over 1-month (top panel) and 6-month (bottom panel) windows. The first, second and third rows report AUCs for random forests models with 10 (rather than five) terminal nodes, 500 (rather than 100) observations per tree, and 1,000,000 (rather than 100,000) trees per forest, respectively. The fourth rows report AUCs for models with a test set beginning January 1, 2010 (rather than January 1, 2007). This corresponds to a 3/4 (rather than 1/2) split between the training and test sets.

Model performance in country-months of onset for *Quad, Goldstein*, *CAMEO* and *Avg.* models

Tables 6, 7, 8 and 9 present ranked predicted probabilities in all country-months of onset from the *Quad*, *Goldstein*, *CAMEO* and *average* models. Columns 5 and 8 (labeled "rank") indicate the ranked predicted probability of onset relative to all other countries Columns 6 and 9 (labeled "rank change") indicate the change in rank from the previous month. Columns 7 and 10 report the percentage of all months in the same country or the percentage of all countries in the same month with a rank change greater than or equal to the one that occurred in the month of country-month of onset. The smaller the percentage, the greater the risk.

Ranking of importance scores for *Quad*, *Goldstein* and *CAMEO* models

Figures 3, 4 and 5 report rankings of importance scores for predictors in the *Quad*, *Goldstein* and *CAMEO* models, respectively. Importance scores are calculated by permuting the values of each predictor across observations in the training set, then estimating the impact of those

Table 5: Robustness checks using out-of-sample AUCs

	1-month forecasts				
	Escalation	*Quad*	*Goldstein*	*CAMEO*	*Avg.*
Terminal nodes	0.86	0.79	0.77	0.82	0.82
Sample size	0.88	0.81	0.71	0.86	0.84
Trees per forest	0.85	0.80	0.79	0.83	0.82
Training/test sets	0.79	0.80	0.69	0.75	0.75

	6-month forecasts				
	Escalation	*Quad*	*Goldstein*	*CAMEO*	*Avg.*
Terminal nodes	0.80	0.76	0.80	0.77	0.78
Sample size	0.83	0.78	0.79	0.79	0.79
Trees per forest	0.83	0.78	0.83	0.77	0.79
Training/test sets	0.88	0.72	0.82	0.69	0.79

Notes: Robustness of random forests model performance to changes in parameters. The top row in each panel reports AUCs for models with 10 rather than five terminal nodes. The second row reports AUCs for models with 500 rather than 100 observations (sampled without replacement) per tree. The third row reports AUCs for models with 1,000,000 rather than 100,000 trees per forest. The fourth row reports AUCs for models with a test set that begins January 1, 2010 and ends December 31, 2015.

Table 6: Test set prediction rankings for 1-month *Quad* model

				Within-country			Cross-country		
	Year	Month	Pred.	Rank	Rank change	% months with ≥ change	Rank	Rank change	% countries with ≥ change
Georgia	2008	8	0.015	2	+15	22%	1	+9	31%
Nigeria	2009	7	0.007	1	+10	12%	2	+3	62%
Yemen	2010	6	0.009	5	-1	75%	2	-1	66%
Libya	2011	2	0.000	12	+0	25%	28	+0	57%
Cote d'Ivoire	2011	2	0.000	29	-28	97%	66	-64	100%
Syria	2011	3	0.000	9	-3	76%	31	-3	65%
Sudan	2011	7	0.014	10	+26	10%	1	+0	66%
Mali	2012	1	0.000	11	+0	73%	24	+3	60%
DRC	2012	5	0.002	1	+10	21%	5	+20	22%
Libya	2012	6	0.002	4	+14	20%	5	+32	10%
Ukraine	2014	4	0.004	4	-3	59%	8	-6	72%

Notes: Change in ranked predicted probabilities for test set country-months of civil war onset derived from a *Quad* model with five terminal nodes and 100 observations (sampled without replacement) per tree, and 100,000 trees per forest. The training set begins January 1, 2001; the test set begins January 1, 2008 and ends December 31, 2015.

Table 7: Test set prediction rankings for 1-month *Goldstein* model

				Within-country			Cross-country		
	Year	Month	Pred.	Rank	Rank change	% months with ≥ change	Rank	Rank change	% countries with ≥ change
Georgia	2008	8	0.009	2	+2	49%	1	+1	63%
Nigeria	2009	7	0.004	7	-3	76%	5	-2	65%
Yemen	2010	6	0.010	5	-2	100%	2	-1	71%
Libya	2011	2	0.000	14	+0	50%	37	-6	64%
Cote d'Ivoire	2011	2	0.001	13	-10	69%	24	-15	77%
Syria	2011	3	0.000	12	-6	89%	48	-11	77%
Sudan	2011	7	0.011	13	+18	15%	1	+2	34%
Mali	2012	1	0.000	10	+0	75%	29	+11	56%
DRC	2012	5	0.001	2	+15	11%	9	+29	11%
Libya	2012	6	0.000	13	-5	80%	27	-16	80%
Ukraine	2014	4	0.002	3	-2	55%	8	-3	66%

Notes: Change in ranked predicted probabilities for test set country-months of civil war onset derived from a *Goldstein* model with five terminal nodes and 100 observations (sampled without replacement) per tree, and 100,000 trees per forest. The training set begins January 1, 2001; the test set begins January 1, 2008 and ends December 31, 2015.

Table 8: Test set prediction rankings for 1-month *CAMEO* model

				Within-country			Cross-country		
	Year	Month	Pred.	Rank	Rank change	% months with ≥ change	Rank	Rank change	% countries with ≥ change
Georgia	2008	8	0.004	7	+28	16%	4	+15	16%
Nigeria	2009	7	0.003	6	+2	47%	8	-2	52%
Yemen	2010	6	0.006	4	+1	25%	2	-1	64%
Libya	2011	2	0.000	13	+0	50%	25	+3	55%
Cote d'Ivoire	2011	2	0.000	35	-34	98%	41	-38	86%
Syria	2011	3	0.000	6	-4	76%	34	-9	72%
Sudan	2011	7	0.009	6	+27	10%	1	+0	36%
Mali	2012	1	0.000	11	+0	70%	24	+4	61%
DRC	2012	5	0.002	2	+5	45%	4	+21	22%
Libya	2012	6	0.002	4	+12	20%	3	+64	1%
Ukraine	2014	4	0.005	4	-3	58%	6	-5	72%

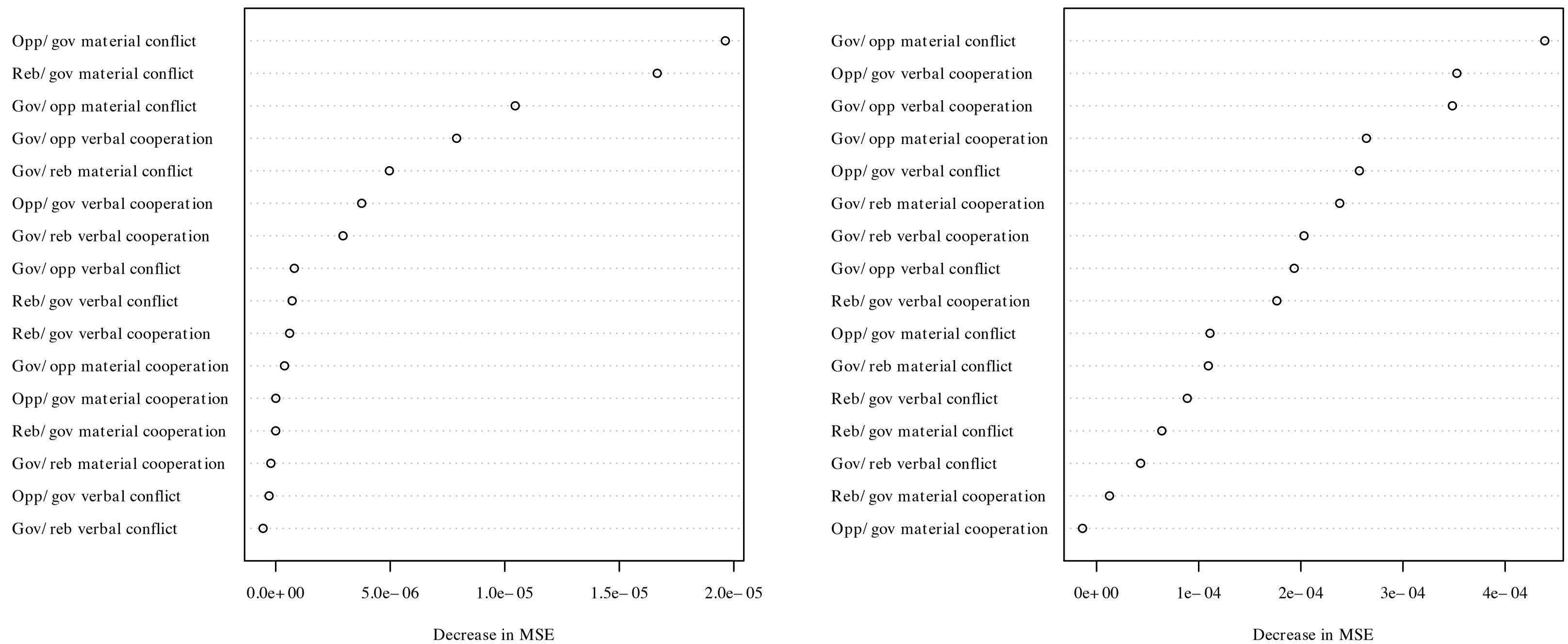
Notes: Change in ranked predicted probabilities for test set country-months of civil war onset derived from a *CAMEO* model with five terminal nodes and 100 observations (sampled without replacement) per tree, and 100,000 trees per forest. The training set begins January 1, 2001; the test set begins January 1, 2008 and ends December 31, 2015.

Table 9: Test set prediction rankings for 1-month *average* model

				Within-country			Cross-country		
	Year	Month	Pred.	Rank	Rank change	% months with ≥ change	Rank	Rank change	% countries with ≥ change
Georgia	2008	8	0.007	2	+10	34%	1	+7	50%
Nigeria	2009	7	0.004	6	+1	59%	4	+1	60%
Yemen	2010	6	0.009	5	-1	100%	2	-1	66%
Libya	2011	2	0.000	18	+0	50%	42	-3	59%
Cote d'Ivoire	2011	2	0.000	15	-13	71%	31	-27	81%
Syria	2011	3	0.000	10	-4	76%	46	-4	66%
Sudan	2011	7	0.012	3	+33	5%	1	+0	37%
Mali	2012	1	0.000	11	+0	68%	37	+6	57%
DRC	2012	5	0.002	2	+14	11%	3	+33	14%
Libya	2012	6	0.002	5	+7	20%	3	+23	16%
Ukraine	2014	4	0.003	4	-3	51%	9	-6	67%

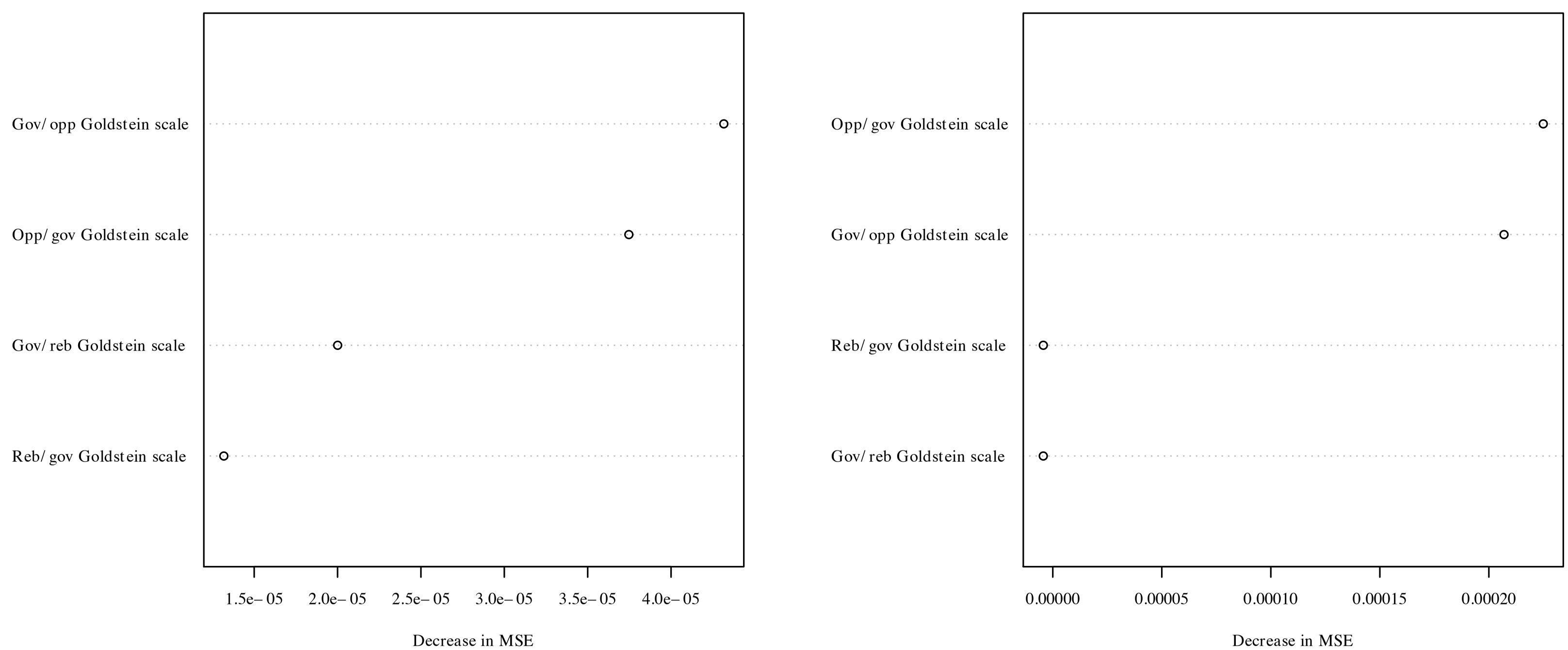
Notes: Change in ranked predicted probabilities for test set country-months of civil war onset derived from the average of the predicted probabilities generated by the *escalation*, *Quad*, *Goldstein* and *CAMEO* models. The training set begins January 1, 2001; the test set begins January 1, 2008 and ends December 31, 2015.

Figure 3: Random forest importance score rankings for *Quad* model



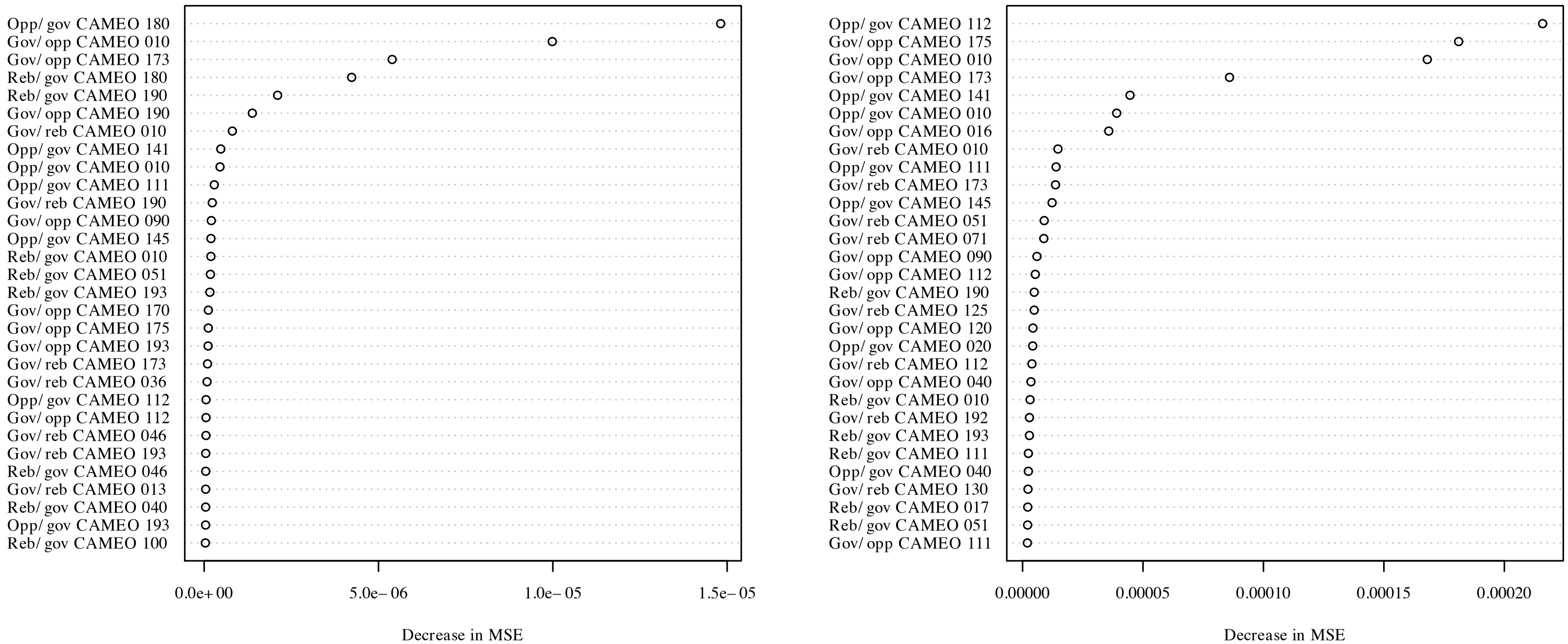
Notes: Ranking of predictors from the *Quad* model based on random forests importance scores. Importance scores are estimated as the decrease in mean squared error for each predictor across 100,000 trees, with five terminal nodes and 100 observations (sampled without replacement) per tree. The training set begins January 1, 2001; the test set begins January 1, 2008 and ends December 31, 2015. The left and right plots show results from 1-month and 6-month forecasts, respectively.

Figure 4: Random forest importance score rankings for *Goldstein* model



Notes: Ranking of predictors from the *Goldstein* model based on random forests importance scores. Importance scores are estimated as the decrease in mean squared error for each predictor across 100,000 trees, with five terminal nodes and 100 observations (sampled without replacement) per tree. The training set begins January 1, 2001; the test set begins January 1, 2008 and ends December 31, 2015. The left and right plots show results from 1-month and 6-month forecasts, respectively.

Figure 5: Random forest importance score rankings for *CAMEO* model (top 30 only)



Notes: Ranking of predictors from the *CAMEO* model based on random forests importance scores. Importance scores are estimated as the decrease in mean squared error for each predictor across 100,000 trees, with five terminal nodes and 100 observations (sampled without replacement) per tree. The training set begins January 1, 2001; the test set begins January 1, 2008 and ends December 31, 2015. The left and right plots show results from 1-month and 6-month forecasts, respectively.

permutations on “out-of-bag” accuracy over all trees in the forest (Breiman 2001). For ease of exposition we report only the top 30 most important predictors in the *CAMEO* model. Definitions of specific CAMEO codes can be found at `https://urldefense.proofpoint.com/v2/url?u=http-3A__data.gdeltproject.org_documentation_CAMEO.Manual.1.1b3.pdf&d=AwIGaQ&c=-dg2m7zWuuDZ0MUcV7Sdqw&r=cqVPv_YiLydNJ3UKCqrhpBAJMnD7M0ODcVHHvOPt6e4&m=SoNhGu7CP7ihUerpEb-zHaJwJ2DNUP8dVxkmNUqoppk&s=iXOUjXiHaG2B9Ehfprt_h1oqWlg_EhFoh68YE_8BpHU&e=`.

Forecasts for first half of 2016 from *Quad*, *Goldstein* and *CAMEO* models

Tables 10, 11, 12 and 13 list the countries with the highest risk of civil war onset in the first half of 2016 as estimated by the *Quad*, *Goldstein*, *CAMEO* and *average* models, respectively. For compactness and ease of exposition we report results for the top 30 highest-risk countries only, based on cross-country rankings generated over 6-month windows.

Table 10: 2016 prediction rankings for 6-month *Quad* model

Country	**Pred.**	**Country**	**Pred.**	**Country**	**Pred.**
India	0.134	Yemen	0.053	Kenya	0.018
Turkey	0.125	Lebanon	0.045	China	0.018
Afghanistan	0.119	Egypt	0.040	South Africa	0.017
Nigeria	0.116	Myanmar	0.036	Burundi	0.017
Somalia	0.115	Mozambique	0.035	Kyrgyzstan	0.016
Syria	0.094	Thailand	0.029	Mali	0.015
Philippines	0.093	Ukraine	0.028	Israel	0.015
Colombia	0.090	Russia	0.026	Indonesia	0.013
Pakistan	0.076	Sudan	0.025	France	0.013
Iraq	0.073	Burkina Faso	0.018	Ethiopia	0.012

Notes: Ordinal ranking of the top 30 countries at highest risk of civil war in the first half of 2016 based on predicted probabilities from the *Quad* model, using predictors measured through December 31, 2015. The list includes countries with ongoing civil wars in the second half of 2015; we interpret results for these countries as indicative of a high risk that civil war will persist.

Table 11: 2016 prediction rankings for 6-month *Goldstein* model

Country	**Prediction**	**Country**	**Prediction**	**Country**	**Prediction**
Iraq	0.079	Yemen	0.046	Romania	0.017
Philippines	0.077	Myanmar	0.044	Kenya	0.016
India	0.077	Ukraine	0.040	China	0.015
Afghanistan	0.077	Burundi	0.039	Belarus	0.015
Nigeria	0.077	Russia	0.037	Congo-Brazzaville	0.015
Turkey	0.077	Ethiopia	0.031	Indonesia	0.015
Somalia	0.075	Iran	0.025	Tunisia	0.014
Pakistan	0.075	Thailand	0.022	Mali	0.013
Egypt	0.062	Colombia	0.022	Montenegro	0.012
Syria	0.057	Germany	0.020	Sudan	0.010

Notes: Ordinal ranking of the top 30 countries at highest risk of civil war in the first half of 2016 based on predicted probabilities from the *escalation* model, using predictors measured through December 31, 2015. The list includes countries with ongoing civil wars in the second half of 2015; we interpret results for these countries as indicative of a high risk that civil war will persist.

Table 12: 2016 prediction rankings for 6-month *CAMEO* model

Country	**Pred.**	**Country**	**Pred.**	**Country**	**Pred.**
Afghanistan	0.109	Yemen	0.026	France	0.008
India	0.109	Myanmar	0.023	Mali	0.008
Nigeria	0.071	Mozambique	0.019	Georgia	0.008
Turkey	0.060	Israel	0.016	Australia	0.008
Colombia	0.055	Russia	0.015	Indonesia	0.007
Syria	0.052	Ukraine	0.013	Egypt	0.007
Somalia	0.052	China	0.013	Kenya	0.007
Philippines	0.052	Lebanon	0.013	South Africa	0.007
Pakistan	0.047	Thailand	0.009	Niger	0.006
Iraq	0.030	Sudan	0.009	Burundi	0.006

Notes: Ordinal ranking of the top 30 countries at highest risk of civil war in the first half of 2016 based on predicted probabilities from the *escalation* model, using predictors measured through December 31, 2015. The list includes countries with ongoing civil wars in the second half of 2015; we interpret results for these countries as indicative of a high risk that civil war will persist.

Table 13: 2016 prediction rankings for 6-month *average* model

Country	**Pred.**	**Country**	**Pred.**	**Country**	**Pred.**
India	0.098	Yemen	0.040	Sudan	0.015
Afghanistan	0.092	Egypt	0.037	Mali	0.015
Nigeria	0.084	Russia	0.031	Ethiopia	0.014
Turkey	0.082	Myanmar	0.030	Iran	0.014
Somalia	0.078	Burundi	0.026	Indonesia	0.014
Philippines	0.072	Ukraine	0.025	Kenya	0.012
Syria	0.067	Lebanon	0.021	Israel	0.012
Pakistan	0.066	Thailand	0.019	France	0.010
Iraq	0.063	Mozambique	0.019	South Africa	0.010
Colombia	0.049	China	0.017	Germany	0.010

Notes: Ordinal ranking of the top 30 countries at highest risk of civil war in the first half of 2016 based on predicted probabilities from the *average* model, using predictors measured through December 31, 2015. The list includes countries with ongoing civil wars in the second half of 2015; we interpret results for these countries as indicative of a high risk that civil war will persist.