

# The shape of things to come? Expanding the inequality and grievance model for civil war forecasts with event data

Daina Chiba, University of Essex

Kristian Skrede Gleditsch, University of Essex & Peace Research Institute Oslo (PRIO)

## **Abstract**

We examine if dynamic information from event data can help improve on a model attempting to forecast civil war using measures reflecting plausible motivation and grievances. Buhaug, Cederman and Gleditsch predict the risk of civil war using a horizontal inequality model with measures reflecting motivation and relevant group characteristics at the country level. The predictions from their model outperform in an out-of-sample forecast conventional country-level models of civil war, emphasizing vertical inequality and country characteristics. However, most grievance measures change little over time. We surmise that a model reflecting potential motivation for conflict can be improved with more dynamic information on mobilization and the behavior of actors. Our conjecture receives some support in the empirical analysis, where we consider both conflict onset and termination over territorial and governmental incompatibilities in the Uppsala/PRIO Armed Conflict Data, and find some evidence that event data can help improve forecasts. Moreover, models with the original grievance measures do better than purely event based models, supporting our claim that both structure and event based components can add value to conflict prediction models. However, the contribution of events to improving predictive power is modest and not entirely consistent, and some types of conflict events seem easier to forecast than others.

## Introduction

We examine if more dynamic information on interactions from event data can help improve the predictive ability of models focusing on grievances and inequality. Buhaug, Cederman & Gleditsch (2014) argue that efforts to predict civil wars can be improved by more attention to actors and their grievances. They propose better theoretically informed measures at the country-level, reflecting motivation and group characteristics relevant to civil war, and show that the predictions from the proposed model outperform a conventional country-level models of civil war, emphasizing country characteristics, both for in-sample classification as well as in an out-of-sample forecast. However, since most of the grievance measures change little over time, the model identifies primarily ‘structural risk’ or potential motivation and opportunity for conflict. We surmise that such a model can be improved with more dynamic information on mobilization and the behavior of actors. Stated more poetically, attention to actors and their grievances may help us appreciate the shapes of different types of pegs and holes when it comes to the motivation for conflict, but more attention to change and the behavior of actors may help identify the shape of things to come in their interaction. To anticipate, we will present results that partially support our conjecture that event data can improve out-of-sample forecasts of civil war events over governmental and territorial incompatibilities in the Uppsala/PRIO Armed Conflict Data. Moreover, models with the original grievance measures do better than purely event based models, supporting our claim that both structure and event based components can add value to conflict prediction models. However, the contribution of events to improving predictive power is modest and not entirely consistent across samples, and some types of conflict events seem easier to forecast than others.

## Structure and events in forecasting conflict

The ability to forecast conflict has long been an important aspiration (see, e.g. Choucri & Robinson, 1978). The last decades have seen many efforts to predict civil war or domestic instability and to develop more systematic global risk assessments, with the Political Instability Task Force as a prominent example (see Esty et al., 1995; Goldstone et al., 2010; King & Zeng, 2001). Many early efforts to predict civil war relied largely on off-the-shelf political and economic country characteristics such as GDP per capita, democracy, and income inequality (see Ward, Greenhill & Bakke, 2010). Drawing on previous disaggregated studies of civil war (see Cederman & Gleditsch, 2009), Buhaug, Cederman & Gleditsch (2014) argue that understanding and predicting civil war can be improved by a more actor oriented approach to motivation and opportunities, and focus on group-based (or ‘horizontal’) inequalities rather than individual (or ‘vertical’) inequalities or country average or aggregate measures. They train a model on observations for 1960 – 1999 and use the estimated probabilities of conflict for 1999 to create a forecast for civil war onset over the next decade, 2000 – 2009. They find that the suggested ‘horizontal’ inequalities model correctly identifies 8 out of the 26 conflicts over the period while a traditional model emphasizing only ‘vertical’ inequalities only identifies 4.

We extend the Buhaug et al. model by looking at how information on interactions between the government and dissidents in prior months influence our ability to predict civil war events. Whereas Buhaug et al. looked only at conflict onset we consider a more general transition model with both conflict onset and termination over territorial and governmental incompatibilities. We also extend the original Buhaug et al. data, both in terms of resolution as well as coverage.

Traditional episodic conflict data identify violence starting at one date  $t$  and ending at  $t+k$ . Disaggregated event data identify individual interactions between specific actors and targets (i.e., who did what and to whom). Event data have often been used as the response or to identify specific conflict events of interest. Here we consider whether event data can inform predictions of conflict episodes through information reflecting heightened tension between actors likely to precede a conflict outbreak.<sup>1</sup> For example, event data may help identify dissident mobilization prior to an outbreak of lethal conflict exceeding the conventional battledeaths threshold. Likewise, increasing repression by the government against non-state actors could reflect anticipated conflict or motivate armed dissident mobilization. Event data may also help predict the termination of ongoing armed conflict. For example, conciliatory acts by the government and rebel organizations may signal improved prospects for a peace agreement or augur an imminent end to armed hostilities.

There is a long tradition of event data models of political conflict (see, e.g., Freeman & Goldstein, 1989; Goldstein & Pevehouse, 1997). However, many studies are country or dyad specific, and focus on evaluating propositions rather than prediction *per se*.<sup>2</sup> Moreover, many traditional collections such as the Conflict and Peace Data Bank (COPDAB) and World Event Interaction Survey (WEIS) have not been updated in ways that allow generating or evaluating real-time forecasts. Many recent event data projects use automated coding from news media sources (see, e.g. Bond et al., 2003; Schrodt & Gerner, 1994). This bears promise of data on a near real-time basis,

---

<sup>1</sup>Ghodes & Carey (2017) use a similar approach, where they treat killing of journalists as a precursor to political repression.

<sup>2</sup>An important early exception is Pevehouse & Goldstein (1999), who use evidence from Bosnia-Herzegovina to predict *ex ante* the effects of NATO involvement on the Kosovo conflict.

and machine coding can help avoid common problems associated with human coders, who often classify events differently (see, e.g., King & Lowe, 2003; Ruggeri, Gizelis & Dorussen, 2011).

Following interest in the predictive power of existing civil war models (see Ward, Greenhill & Bakke, 2010; Weidmann & Ward, 2010), researchers have proposed global forecasting models of civil war integrating structural characteristics and behavior gleaned from events. Ward et al. (2013) develop a model relating the risk of civil wars in the Uppsala/PRIO data to conflictual events collected from news media sources, also considering a host of country characteristics as well as random effects and spatial clustering. They report high predictive accuracy, both in-sample and out-of-sample. In another important contribution, Chadeaux (2014) looks at whether prior news events can serve as an early warning indicator of wars within and between states. He finds that adding events substantially improves predictive accuracy over a purely structural model, drawing on studies of the liberal peace. Our approach here is conceptually similar, but also involves a number of important differences.

First, we consider a transition model of conflict and peace, where we look at predicting the onsets of new conflicts and the termination of existing conflict episodes separately. Ward et al. (2013) appear to model monthly conflict incidence rather than initial onset. Many predicted conflict months are thus likely to reflect ongoing conflict, which presumably is easier to predict than new onsets. It is not possible to evaluate from the published results how much of the reported predictive accuracy derives from correctly predicted new onsets and correctly predicted ongoing conflicts.

Second, while most focus on any civil war, we distinguish conflicts by their incompatibility, i.e., whether over territory or the government. There is a great deal of evidence suggesting that

these exhibit different risk factors and dynamics (Bartusevičius, 2014; Buhaug, 2006; Buhaug, Cederman & Gleditsch, 2014; Sobek & Payne, 2010). For example, separatist claims are unlikely in the absence of ethnic groups with plausible grievances and tend to take place in the periphery, unlike civil wars over the government emerging out of military coups or challenges from below.

Third, we consider a wider range of structured events specifications, focusing both on specific type and the actor-target direction. We consider propositions on the type of events that we believe are likely to be helpful for predicting conflict onset and termination and contrast these with alternative specifications.

Fourth, we engage in more explicit model comparisons, examining the possible contribution of events within a model with more plausible structural characteristics relevant to conflict. This is essential to evaluate the possible independent contributions of events. Chadeaux (2014) conducts a detailed comparison of models, indicating a substantial contribution from events. However, he considers a model based on work on the liberal peace, highlighting capabilities and regime type. Although this body of research sets forward plausible propositions on factors that may decrease the likelihood of conflict, liberal peace models essentially consider the risk of conflict as externally determined and contain little information on the plausible motivation (see Gleditsch & Ward, 2013). The only plausible motivation related factors in Ward et al. (2013) is democracy and a curvilinear specification of the percentage excluded population, and there is no explicit comparison of the contribution of ‘structural’ versus ‘events’. Following Buhaug, Cederman & Gleditsch (2014) we believe that a more plausible baseline model of structural influences is possible and essential for assessing the added value of event based information.

# Measuring directed dyadic interactions using the Integrated Conflict Early Warning System (ICEWS)

We first discuss the event data we use, and then turn to the specific directed dyadic interactions that we expect will be related to higher or lower probabilities of conflict onset and termination. Many existing global forecasts using event data counts all events or some smaller subset deemed conflictual. We believe one potentially can do better by distinguishing between the domestic events more relevant to conflict dynamics as well as the direction of interactions or who does what to whom.

We use event data from the Integrated Conflict Early Warning System (ICEWS, see Boschee et al., 2015).<sup>3</sup> Previously restricted, these data became publicly available in early 2015, with an embargo on the data for most recent twelve months.<sup>4</sup> The data are available from January 1996.

We extract domestic dyadic events reflecting government-dissident interactions that we believe are relevant to future conflict.<sup>5</sup> We identify events as domestic if both the actor and targets originate from the same country, and exclude all international events.

---

<sup>3</sup>Ward & Beger (2017) also use the ICEWS event data for forecasting irregular leader changes.

<sup>4</sup>ICEWS does not include domestic events for the USA, due to US restrictions on funding for research with potential impact on domestic affairs.

<sup>5</sup>We have also considered scaled measures, based on the sum of the Goldstein (1992) scores for all events over a time interval. However, we find little evidence that this improves on the simple event counts, and report only the latter here. Moreover, negative (conflictual) and positive (cooperative) scores can potentially average out for summed measures (see Schrodt, 2011, 6-7).



We identify directed government-dissident interactions based on the pairwise combinations of the Conflict and Mediation Event Observations (CAMEO) sector codes (see Schrod, 2012). More specifically, we consider the following combinations as government to a non-state actor  $G \rightarrow N$  events:

- Government to Civilians (GOV-CVL),
- Government to Opposition (GOV-OPP), and
- Government to Rebels (GOV-REB).

We consider the following combinations as non-state actor to the government  $N \rightarrow G$  events:

- Opposition to Government (OPP-GOV) and
- Rebels to Government (REB-GOV).

We do not include actions by civilians towards the government, since this is less likely to reflect the organized collective action necessary for civil war. However, we include actions by the governments against civilians since repression can motivate mobilization.

To classify events as cooperative or conflictual, we consider the 2x2 quad categories suggested by Schrod (2011, 7), where events are placed into four conceptually unique and mutually exclusive categories, based on their CAMEO codes:

- Q1 - Verbal Cooperation: ‘The occurrence of dialogue-based meetings (i.e. negotiations, peace talks), statements that express a desire to cooperate or appeal for assistance (other than material aid) from other actors. CAMEO categories 01 to 05’.

- Q2 - Material Cooperation: ‘Physical acts of collaboration or assistance, including receiving or sending aid, reducing bans and sentencing, etc. CAMEO categories 06 to 09’.
- Q3 - Verbal Conflict: ‘A spoken criticism, threat, or accusation, often related to past or future potential acts of material conflict. CAMEO categories 10 to 14’.
- Q4 - Material Conflict: ‘Physical acts of a conflictual nature, including armed attacks, destruction of property, assassination, etc. CAMEO categories 15 to 20’.

While previous global studies have used past conflictual events to predict civil war onset, we consider how events predict to both conflict onset and conflict termination. We expect that conflict events will tend to precede conflict onset, while cooperation can help anticipate conflict termination. Moreover, we expect material events to be more important than verbal events. In sum, we expect material conflict events ( $Q4$ ) to be potentially useful leading indicators of civil war onset, while material cooperation ( $Q2$ ) should lead conflict termination.

We also believe that the direction or the specific actors and targets will be important. Governments  $G$  consistently respond to severe threats such as violent dissent with force (see, e.g., Davenport, 2007). However, whether state repression generates a reaction depends on non-state actors  $N$ , who normally face a host of barriers impeding collective action and organized or coordinated responses (see, e.g. Lichbach, 1995). Civil war presumes organized armed non-state actor mobilization. Thus, we conjecture that higher  $N \rightarrow G$  Q4 (material conflict) counts will have a larger effect than higher  $G \rightarrow N$  Q4 counts.

For conflict termination it is commonly assumed that rebels are generally open to negotiations,

since these transmit some legitimacy and recognition of their claims, while the government is more likely to resist (see, e.g., Clayton & Gleditsch, 2014). If so, then overtures by the government are likely to be more influential for termination prospects than gestures by rebels, and higher  $G \rightarrow N$  Q2 (material cooperation) counts should carry more weight than higher  $N \rightarrow G$  Q2 counts.

While many analyses use raw event data, we believe that trending can make this problematic over longer periods. The ICEWS data and many other event data projects have a strong positive trend in recorded events over time, possibly due to increasing news coverage. Figure 1 shows a clear increasing trend since 1996 in the count of domestic events by country month for all quads in the ICEWS data. This is difficult to reconcile with the emerging scholarly consensus that violent conflict has declined since the mid-1990s (see Pinker, 2011; Gleditsch & Pickering, 2014; Cederman, Gleditsch & Wucherpfennig, 2017). Moreover, it makes it less plausible to assume that some count value  $c$  will have a uniform impact over long periods of time.

Some have suggested addressing the issue of trending data through normalization, for example by dividing conflictual events to all reported events to identify deviations from the baseline.<sup>6</sup> However, relative rates can be affected by numerous things affecting the numerator, much as summed conflict and cooperation measures. Moreover, normalization makes it more difficult to interpret the coefficients or derive direct predictions from the raw data. Instead, we propose as an alternative to add a time indicator ( $T$ ) and interact this with the event counts. This has the advantage of generating simple results to directly assess the degree of trending as well as making predictions

---

<sup>6</sup>See, e.g., the discussion at <https://dartthrowingchimp.wordpress.com/2014/06/06/another-note-on-the-limitations-of-event-data/>.

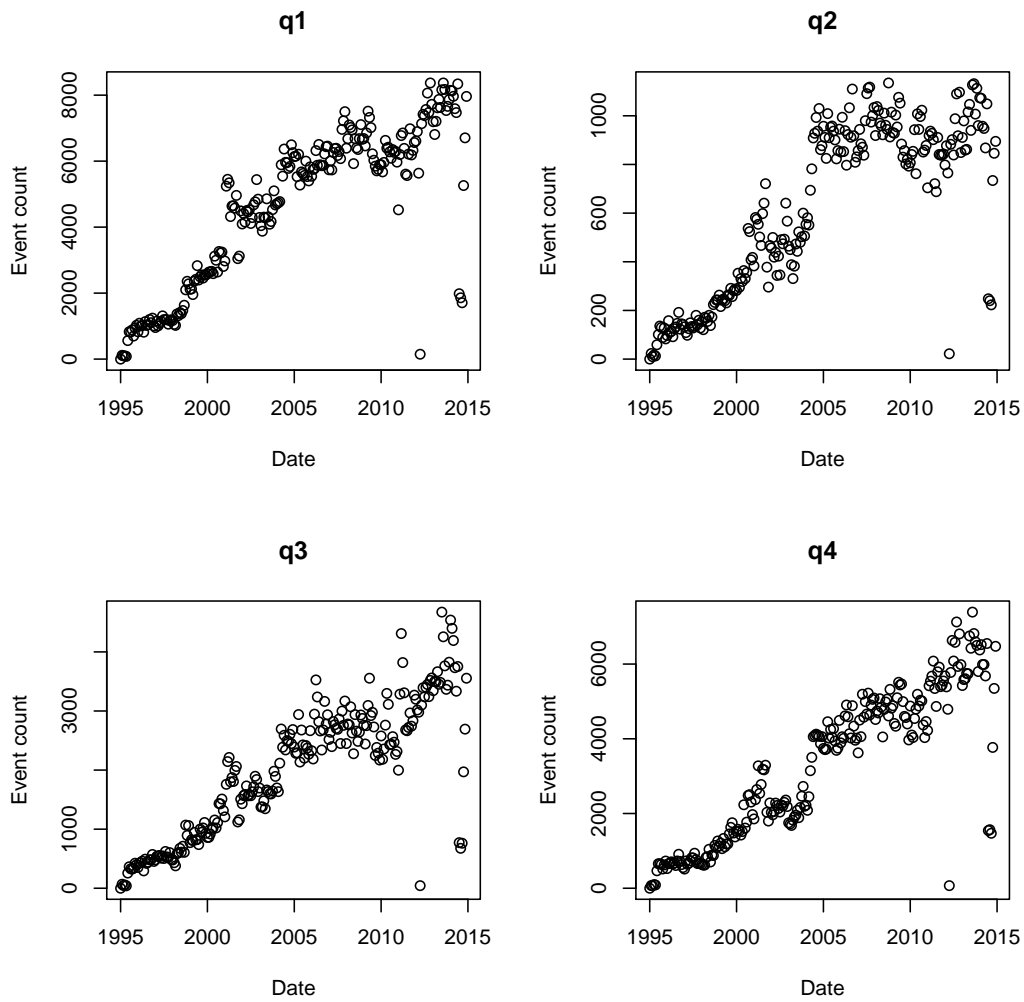


Figure 1. Quad counts by time

directly from individual raw data points.<sup>7</sup>

---

<sup>7</sup>Likewise, studies can control for either time or total/other reported events on the right hand side, e.g., Witmer et al. (2017).

## A transition model of civil war and conflict termination

In this section we outline the transition model framework as well as the specific structural covariates that we consider. We extend the Buhaug et al. model of civil war onset to a two-way transition model. We consider separate equations for a) transitions from peace to war, i.e.,  $y_{t-1} = 0 \rightarrow y_t = 1$  and b) transitions from war to peace  $y_{t-1} = 1 \rightarrow y_t = 0$ . We estimate the equations using logit models. Note that estimating the equations separately is equivalent to estimating a joint transition model here, since the variance for logit is set to 1 for identification. Although other estimation approaches from machine learning conceivably could do much better for rare events (see, e.g., Kennedy, 2015; Muchlinski et al., 2016), we chose a simple logit model for comparability to previous studies and to make it easier to evaluate the impact of incremental model changes.

Whereas Buhaug et al. used country years or annual observations we here look at monthly observations. Although many covariates in the Buhaug et al. model change slowly over time or are available only for annual data, others such as Polity can change within years in ways that may be associated with conflict events.<sup>8</sup> Moreover, we expect the lead time of events to be relatively short, and early warning conflict using leading indicators from events is both more useful and potentially realistic using time periods shorter than a year. We follow previous research on prediction such as Ward et al. (2013) in using monthly observations.<sup>9</sup>

Our measures are generally similar to those used by Buhaug et al., but we recollect a new data

---

<sup>8</sup>See, e.g., Gleditsch & Ruggeri (2010) on opportunity structures and civil war onset.

<sup>9</sup>Differences in temporal aggregation can of course also affect the results (see Shellman, 2004), but space precludes us from examining alternative time units here.

set from scratch using the most recent original sources. The original Buhaug et al. data are now dated and end in 2009, and we use updated data through 2015 for a complete sample of all states in the Gleditsch & Ward (1999) list of independent states since 1 January 1996.<sup>10</sup> The original Buhaug et al. data also contained missing values, either due to missing data in the sources or populations that diverge from the Gleditsch and Ward population of states. For example, income inequality data is missing for many countries in standard sources, especially in the Middle East. Moreover, the Ethnic Power Relations (EPR) data do not report data for smaller countries with a population below 500,000 in 1990 (Vogt et al., 2015, 1330), including some that experience civil war such as Comoros and Suriname.<sup>11</sup> A global forecasting model should be able to derive risk forecasts for all countries in the UCDP/PRIO universe (i.e., the Gleditsch & Ward list). An estimate based on the best available sources is likely to be more useful than simply assuming full ignorance, or that conflict cannot occur in small countries.

We refer to Buhaug et al. for a more detailed theoretical rationale for the structural covariates, but briefly summarize the data and core expectations here. With regards to our conflict data and measures of past conflict history, we consider:

- Civil war (territorial or governmental) onset/termination: an outbreak or end of an UCPD/PRIO conflict episode involving more than 25+ deaths over the specific incompatibility in the specific month, see Gleditsch et al. (2002). We use most recent version as of the time of writing,

---

<sup>10</sup>For sources not available beyond 2013 we use the last available values.

<sup>11</sup>The population of Suriname exceeds 500,000 from 2006 according to World Bank estimates

<http://www.worldbank.org/en/country/suriname>.

i.e., 4-2015 (Melander et al., 2016). We disregard ongoing conflict periods in the onset equation and peace periods in the termination equation. Buhaug et al. distinguish between ethnic and non-ethnic conflicts, based on the ACD2EPR data linking the actors in the Armed Conflict Data to individual ethnic groups in the EPR data through explicit claims (Wucherpfennig et al., 2012). The ACD2EPR data are not available for the most recent conflicts, and we here opt for the simpler alternative of distinguishing between territorial and governmental incompatibilities. This can be classified with less information and subjective judgments, and easily compared to any UCDP data. We expect territorial conflict to be driven primarily by the characteristics of ethnic peripheral groups and have very different risk factors from governmental conflicts.

- Lagged peace/war months: a count of consecutive time in months of peace/conflict. For peace durations we count from 1 January 1946 or the end of previous conflict. For war durations we count from the start of ongoing conflict, using the full data from after 1946. We expect longer peace durations to make new conflict onset less likely. Longer conflict duration reflects persistent conflict, less likely to end in any given time periods, as opposed to coups that often last less than a month. We use logged values after adding one to the base since we expect increases to matter increasingly less the longer the spells.

Following Buhaug et al. we consider a series of standard indicators for vertical inequalities reflecting potential motivation and factors affecting the opportunities for conflict, similar to the first global forecasting models in the State Failure Task Force (Esty et al., 1995), including Ethno-Linguistic Fractionalization, Democracy, GINI coefficient, Population, and GDP per capita. Be-

yond the standard measures of vertical inequalities or opportunities for conflict we introduce a series of measures reflecting horizontal inequalities or political and economic group characteristics that we expect to be relevant for the risk of conflict, including Largest Discriminated Group, Downgrade/upgrade dummy (whether any ethnic group has experienced a loss/gain in access to state power), Power sharing dummy, and Negative/Positive Horizontal Inequality. The supplementary appendix includes a full description of these covariates.

## **Empirical analysis**

We first replicate the Buhaug et al. model for monthly data over the period with event data (from 1 January 1996) up through 31 December 2012. We then add event count measures to assess the degree of improvement in-sample. In the subsequent section we consider out-of-sample performance through cross-validation and comparisons of forecasts beyond 2012.

### **In-sample estimates**

The first two columns in Table I provide a vertical inequality model and an expanded model with indicators of horizontal inequality similar to Buhaug et al. for all conflict onsets. To preserve space we report the full classification matrices in the supplementary appendix, but report as a summary measure the share correctly predicted for  $y = 1$  at the threshold 0.01. There is no evidence for any added value of horizontal inequality indicators in the general conflict model. None of the individual terms are significant, and there is no improvement in model fit. However, the added



value of the horizontal inequality indicators is much clearer when we separate between territorial onsets and governmental conflict onsets in columns 3-4 and 5-6 respectively. For territorial onsets, we find a substantial increase in the risk of conflicts with countries with large excluded groups (an increase in the odds by factor of almost 8 for the largest excluded group in our data) as well as higher ratio of the maximum low horizontal income inequality measure. The findings for territorial conflict onset are generally similar to the findings of Buhaug et al. on ethnic territorial conflicts. As expected we find that powersharing increases the likelihood of conflict onset consistent with coups or infighting, and we find a somewhat surprisingly a negative impact of the max low ratio. Moreover, although there is no general relationship between a higher GINI and the risk of civil war, we are more likely to see government conflict onsets in countries with higher individual income inequality. Overall, the results are generally consistent with prior expectations.<sup>12</sup>

We now proceed to examine conflict termination. Table II indicate a fair amount of parallelism, or negative influences on termination from factors that increase the risk of onset. There is little evidence in the first two columns that a model with horizontal inequality measures improves much on a model of termination when we lump together all conflicts. However, columns 3 and 4 indicate a clearer improvement for territorial conflict termination. In particular, powersharing between groups and a groups upgrade makes territorial conflict termination more likely, consistent

---

<sup>12</sup>Save for the first column for model with all conflicts and no horizontal inequality indicators we find no significant impact of Polity. We have also considered a dichotomous measure and a possible curvilinear relationship, argued to be important by some studies (Goldstone et al., 2010), but find no evidence that alternative specifications fit the data better (see appendix).

Table I. Conflict onset estimates

	<i>Dependent variable:</i>					
	All CW		Terr		Gov	
	(1)	(2)	(3)	(4)	(5)	(6)
Ln population	0.436** (0.060)	0.428** (0.064)	0.776** (0.090)	0.786** (0.094)	0.408** (0.081)	0.519** (0.086)
Ln GDP p.c.	-0.154 (0.095)	-0.193 <sup>†</sup> (0.103)	-0.251* (0.121)	-0.328* (0.129)	-0.501** (0.119)	-0.320* (0.129)
ELF	1.305** (0.415)	1.721** (0.560)	0.977 <sup>†</sup> (0.568)	1.302 <sup>†</sup> (0.693)	1.349* (0.534)	0.668 (0.714)
Democracy	-0.026 <sup>†</sup> (0.014)	-0.016 (0.015)	0.022 (0.020)	0.035 (0.021)	-0.010 (0.019)	-0.014 (0.020)
GINI	-0.003 (0.013)	0.001 (0.014)	-0.001 (0.018)	-0.003 (0.019)	0.042** (0.016)	0.064** (0.018)
Powersharing		-0.239 (0.278)		-0.129 (0.329)		0.770* (0.378)
Max R disc.		1.223 (0.781)		2.473** (0.906)		1.214 (1.224)
Downgrade		-0.586 (0.531)		-1.678 (1.020)		0.399 (0.532)
Max low		0.154 (0.123)		0.385** (0.127)		-0.778 <sup>†</sup> (0.418)
Max high		-0.193 (0.149)		-0.248 (0.198)		-0.128 (0.179)
Ln peace months	-0.684** (0.050)	-0.674** (0.050)	-0.380** (0.052)	-0.373** (0.054)	-0.294** (0.055)	-0.314** (0.055)
Constant	-5.851** (1.076)	-5.711** (1.211)	-10.962** (1.458)	-10.792** (1.585)	-7.143** (1.421)	-9.448** (1.581)
Observations	32,037	32,037	34,978	34,978	33,616	33,616
Log Likelihood	-617.529	-613.271	-418.386	-409.252	-439.923	-431.817
Akaike Inf. Crit.	1,249.057	1,250.542	850.772	842.503	893.845	887.633
% Correctly Pred. for $y = 1$ <sup>[a]</sup>	52%	51.2%	51.35%	55.41%	34.25%	35.62%

<sup>†</sup>p<0.1; \*p<0.05; \*\*p<0.01

[a] Classification threshold is  $\Pr(y = 1) > 0.01$ .

with a positive contribution to conflict resolution. Larger remaining excluded groups makes termination less likely, consistent with the idea of ethnic exclusion making conflicts more persistent (Wucherpfennig et al., 2012). For governmental conflicts, however, there is no evidence that the horizontal inequality measure contribute much, and the AIC is higher.

We now proceed to add event counts to the horizontal inequality model to see if information on events in the prior months can help improve the classification of conflict onset. Since the above analysis suggests that conflicts with different incompatibilities have different risk factors we examine territorial and governmental conflicts separately. Table III presents the results for four models of territorial onsets with the different quad count combinations. There are a number of important results to highlight. We find strong evidence of a time trend in the event counts, and models without the time trend fit noticeably worse. As expected cooperative events (Q1 and Q2) do little to help improve the classification of territorial civil war onset, but there is more evidence that conflictual events appear to lead territorial conflict. However, our expectations about the difference between verbal and material conflict are not fully borne out, as the model with verbal conflict Q3 fits as well as that with material conflict Q4. Our expectation about the influence of non-state actor to government events exceeding that of state to non-state actors is borne out by the model with verbal conflict, but not the material conflict model. The final column in Table III includes a purely event based model, which we find fits the data noticeably worse than models with structural covariates. In Table IV we conduct a similar analysis for governmental conflict onsets. This results here are generally consistent with our expectations in that the model with the Q4 events fits best overall, and better than the Q3 model by some margin, and that  $N \rightarrow G$  events are most

Table II. Conflict termination estimates

	<i>Dependent variable:</i>					
	All terminations		Terr		Gov	
	(1)	(2)	(3)	(4)	(5)	(6)
Ln population	-0.375** (0.098)	-0.372** (0.107)	-0.475** (0.142)	-0.457** (0.171)	-0.467** (0.153)	-0.478** (0.159)
Ln GDP p.c.	0.027 (0.118)	0.186 (0.133)	-0.164 (0.170)	0.339 (0.269)	0.129 (0.175)	0.182 (0.186)
ELF	-0.835 <sup>†</sup> (0.462)	-1.273* (0.643)	-0.387 (0.709)	-2.131 <sup>†</sup> (1.144)	-1.062 <sup>†</sup> (0.589)	-1.361 <sup>†</sup> (0.764)
Democracy	0.013 (0.019)	0.003 (0.020)	-0.028 (0.027)	-0.101** (0.035)	0.037 (0.026)	0.029 (0.028)
GINI	0.009 (0.017)	0.022 (0.018)	0.035 (0.027)	0.096* (0.039)	0.002 (0.020)	0.009 (0.022)
Powersharing		0.430 (0.333)		1.836** (0.663)		0.267 (0.395)
Max R disc.		-1.997 <sup>†</sup> (1.202)		-3.823 <sup>†</sup> (2.128)		-0.279 (0.964)
Upgrade		-0.420 (0.500)		2.536** (0.788)		-0.635 (0.627)
Max low		-0.296 (0.194)		-0.243 (0.198)		-0.154 (0.521)
Max high		0.303 (0.216)		0.899 <sup>†</sup> (0.465)		0.179 (0.291)
Ln conflict months	-0.601** (0.073)	-0.572** (0.073)	-0.716** (0.119)	-0.610** (0.128)	-0.389** (0.091)	-0.372** (0.092)
Constant	1.578 (1.430)	-0.142 (1.604)	3.551 (2.360)	-3.965 (3.612)	1.503 (1.752)	0.856 (1.907)
Observations	5,069	5,069	2,092	2,092	3,444	3,444
Log Likelihood	-389.461	-384.199	-158.343	-146.026	-265.540	-264.525
Akaike Inf. Crit.	792.922	792.398	330.685	316.052	545.079	553.050
% Correctly Pred. for $y = 1^{[a]}$	85.39%	84.27%	87.8%	87.8%	80.7%	77.19%

<sup>†</sup>p<0.1; \*p<0.05; \*\*p<0.01

[a] Classification threshold is  $\Pr(y = 1) > 0.01$ .

informative.

Overall, these results suggest some support for our conjectures that events can help classify onsets, although our expectations that material conflict should be more important than verbal conflict events is not fully borne out for territorial conflict onsets. For simplicity, and based on our prior expectations, we consider primarily the Q4 model in the following comparisons. We report a number of additional results in the supplementary appendix, including model with both Q3 and Q4, the sum of all events, and the conflict/(conflict+cooperation) balance (see Dewal, Goldstone & Volpe, 2013; Rummel, 1963). In general, we find that most of these specifications fit notably worse, and the increase in added complexity in a model with both 3 and 4 is generally not matched by a sufficient increase in model fit. We also report estimates limited to ICEWS data after 2001 in the supplementary appendix, as the pre-2001 data may have been collected using different methods and differ in important ways. However, we find only limited differences and no evidence for a larger predictive power from the event data in the latter period.

We now turn to the contribution of the event counts to identifying conflict termination. Table V reports the results for the termination of territorial conflict. These results support our argument that cooperative events often precede territorial conflict termination, and there is little evidence that conflictual events provide much help as a leading indicator. In this case we find that both verbal and material cooperation appear helpful to predict to termination, and that a model combining the two does considerably better than a model considering only material cooperation events by the AIC. However, we do not find support for our claim that cooperative gestures from the government are more influential than cooperative events by non-state actors, as both appear to have positive

Table III. Territorial conflict onset, horizontal inequality model with events

	<i>Dependent variable:</i>				
	(1)	(2)	Territorial		(5)
			(3)	(4)	
Ln population	0.695** (0.105)	0.719** (0.106)	0.660** (0.106)	0.650** (0.107)	
Ln GDP p.c.	-0.078 (0.155)	-0.092 (0.154)	-0.118 (0.157)	-0.068 (0.154)	
ELF	1.972** (0.762)	1.622* (0.733)	1.661* (0.740)	1.649* (0.748)	
Democracy	0.038 <sup>f</sup> (0.022)	0.032 (0.022)	0.030 (0.023)	0.032 (0.022)	
GINI	-0.014 (0.018)	-0.011 (0.018)	-0.014 (0.018)	-0.014 (0.018)	
Powersharing	-0.271 (0.342)	-0.207 (0.338)	-0.339 (0.339)	-0.221 (0.339)	
Max R disc.	1.955* (0.970)	2.020* (0.932)	2.052* (0.923)	2.080* (0.950)	
Downgrade	-1.410 (1.023)	-1.602 (1.046)	-1.365 (1.022)	-1.329 (1.023)	
Max low	0.054 (0.168)	0.207 (0.140)	0.283* (0.139)	0.250 <sup>f</sup> (0.141)	
Max high	-0.401 (0.264)	-0.227 (0.203)	-0.358 (0.247)	-0.243 (0.205)	
Ln peace months	-0.369** (0.057)	-0.398** (0.057)	-0.372** (0.057)	-0.356** (0.059)	
T	-0.006* (0.002)	-0.005* (0.002)	-0.006* (0.002)	-0.006* (0.002)	-0.007** (0.002)
Q1, G → N	0.008 (0.015)				
Q1, N → G	0.015 (0.018)				
T: Q1, G → N	<0.001 (<0.001)				
T: Q1, N → G	> -0.001 (<0.001)				
Q2, G → N		0.038 (0.029)			
Q2, N → G		0.172 (0.116)			
T: Q2, G → N		> -0.001 (<0.001)			
T: Q2, N → G		-0.002 <sup>f</sup> (0.001)			
Q3, G → N			-0.001 (0.025)		
Q3, N → G			0.038** (0.015)		
T: Q3, G → N			<0.001 (<0.001)		
T: Q3, N → G			> -0.001* (<0.001)		
Q4, G → N				0.017* (0.008)	0.035** (0.007)
Q4, N → G				0.014 (0.020)	0.041* (0.019)
T: Q4, G → N				> -0.001 (<0.001)	> -0.001** (<0.001)
T: Q4, N → G				> -0.001 (<0.001)	> -0.001* (<0.001)
Constant	-10.629** (1.874)	-11.008** (1.871)	-10.105** (1.888)	-10.617** (1.854)	-5.870** (0.237)
Observations	34,598	34,598	34,598	34,598	34,598
Log Likelihood	-389.640	-391.615	-386.659	-386.744	-457.776
Akaike Inf. Crit.	813.280	817.230	807.319	807.488	927.552
% Correctly Pred. for $y = 1$ <sup>[a]</sup>	55.41%	58.11%	56.76%	58.11%	20.27%

<sup>f</sup>p<0.1; \*p<0.05; \*\*p<0.01  
[a] Classification threshold is Pr(y = 1) > 0.01.

Table IV. Governmental conflict onset with events

	<i>Dependent variable:</i>				
	Government				
	(1)	(2)	(3)	(4)	(5)
Ln population	0.257** (0.096)	0.324** (0.094)	0.284** (0.099)	0.269** (0.097)	
Ln GDP p.c.	-0.578** (0.144)	-0.510** (0.139)	-0.575** (0.144)	-0.575** (0.142)	
ELF	0.448 (0.722)	0.287 (0.712)	0.141 (0.729)	0.251 (0.727)	
Democracy	-0.029 (0.021)	-0.026 (0.020)	-0.038 <sup>†</sup> (0.021)	-0.032 (0.021)	
GINI	0.050** (0.019)	0.055** (0.019)	0.050** (0.019)	0.051** (0.019)	
Powersharing	0.811* (0.387)	0.943* (0.385)	0.924* (0.392)	0.979* (0.397)	
Max R disc.	1.545 (1.145)	1.644 (1.156)	1.613 (1.164)	1.620 (1.190)	
Downgrade	0.303 (0.540)	0.395 (0.540)	0.279 (0.543)	0.263 (0.565)	
Max low	-0.788* (0.337)	-0.950* (0.427)	-0.565 (0.359)	-0.691* (0.345)	
Max high	-0.284 (0.258)	-0.070 (0.184)	-0.233 (0.231)	-0.090 (0.181)	
Ln peace months	-0.287** (0.059)	-0.299** (0.059)	-0.272** (0.060)	-0.251** (0.060)	
<i>T</i>	-0.002 (0.002)	-0.002 (0.002)	-0.002 (0.002)	-0.002 (0.002)	-0.004 <sup>†</sup> (0.002)
Q1, <i>G</i> → <i>N</i>	0.012 (0.018)				
Q1, <i>N</i> → <i>G</i>	0.019 (0.024)				
<i>T</i> : Q1, <i>G</i> → <i>N</i>	<0.001 (<0.001)				
<i>T</i> : Q1, <i>N</i> → <i>G</i>	> -0.001 (<0.001)				
Q2, <i>G</i> → <i>N</i>		0.079** (0.031)			
Q2, <i>N</i> → <i>G</i>		0.165 (0.119)			
<i>T</i> : Q2, <i>G</i> → <i>N</i>		> -0.001 (<0.001)			
<i>T</i> : Q2, <i>N</i> → <i>G</i>		-0.001 (0.001)			
Q3, <i>G</i> → <i>N</i>			0.050 (0.031)		
Q3, <i>N</i> → <i>G</i>			0.017 (0.014)		
<i>T</i> : Q3, <i>G</i> → <i>N</i>			> -0.001 (<0.001)		
<i>T</i> : Q3, <i>N</i> → <i>G</i>			> -0.001 (<0.001)		
Q4, <i>G</i> → <i>N</i>				0.012 (0.010)	0.023** (0.008)
Q4, <i>N</i> → <i>G</i>				0.050 <sup>†</sup> (0.026)	0.021 (0.022)
<i>T</i> : Q4, <i>G</i> → <i>N</i>				<0.001 (<0.001)	> -0.001 (<0.001)
<i>T</i> : Q4, <i>N</i> → <i>G</i>				> -0.001* (<0.001)	> -0.001 (<0.001)
Constant	-4.287* (1.790)	-5.644** (1.755)	-4.842** (1.832)	-4.966** (1.783)	-6.095** (0.255)
Observations	33,378	33,378	33,378	33,378	33,378
Log Likelihood	-405.662	-409.881	-403.168	-397.424	-465.474
Akaike Inf. Crit.	845.325	853.761	840.337	828.848	942.949
% Correctly Pred. for $y = 1^{[a]}$	36.11%	34.72%	34.72%	36.11%	15.28%

<sup>†</sup>p<0.1; \*p<0.05; \*\*p<0.01

[a] Classification threshold is  $\Pr(y = 1) > 0.01$ .

influences for material cooperation and the coefficient for verbal cooperation by the government is significantly negative. The model with Q2 events has the best AIC for governmental conflict termination, but here the contribution of events seems negligible.

We report in Table VII alternative measures of model performance including Area Under the Curve (AUC) for the Receiver Operating Characteristic (ROC) plot and Brier scores for all models discussed. Figures 2 and Figure 3 display the Receiver Operation Characteristics (ROC) curves for the final models with and without events and structural characteristics respectively. The ROC curve identifies the performance of a binary prediction for different classification thresholds by plotting the true positive rate against the false positive rate. A curve further to the upper left corner indicates more accurate predictions and better model performance. The Area under the Curve is simply the area under the ROC curve. It is clear that the purely event based models do relatively poorly compared to models with the structural characteristics. The models with both structural characteristics and events generally do best, although the degree of improvement is more pronounced for governmental conflict onset than territorial conflict onset, even if the latter has a generally higher AUC. The ROC/AUC for the proposed termination models also indicate some modest improvements for the model with events. The Brier score is a measure comparing the accuracy of probabilistic predictions against the observed outcomes by taking the average of the squared deviations, with lower scores indicating more accurate predictions. However, the Brier score is not well suited for distinguishing between alternative specifications for rare events model, and in many instances here the Brier scores are near identical when rounded to three digits.



Table V. Territorial termination with events

	<i>Dependent variable:</i>					
	Termination: Territory					
	(1)	(2)	(3)	(4)	(5)	
Ln population	-0.826** (0.238)	-0.928** (0.243)	-0.647** (0.247)	-0.776** (0.251)	-0.969** (0.256)	
Ln GDP p.c.	0.295 (0.294)	0.213 (0.290)	0.419 (0.288)	0.254 (0.285)	0.237 (0.301)	
ELF	-1.450 (1.191)	-1.743 (1.194)	-1.804 (1.150)	-1.525 (1.191)	-1.748 (1.219)	
Democracy	-0.146** (0.040)	-0.158** (0.041)	-0.136** (0.041)	-0.143** (0.041)	-0.164** (0.042)	
GINI	0.114** (0.044)	0.115** (0.044)	0.109* (0.044)	0.104* (0.045)	0.124** (0.045)	
Powersharing	1.830** (0.710)	1.946** (0.715)	1.892** (0.715)	1.768* (0.724)	2.044** (0.730)	
Max R disc.	-4.245 <sup>†</sup> (2.306)	-3.711 <sup>†</sup> (2.255)	-4.609* (2.279)	-3.832 <sup>†</sup> (2.182)	-3.840 (2.338)	
Upgrade	2.814** (0.772)	2.976** (0.806)	2.774** (0.785)	3.024** (0.799)	2.810** (0.810)	
Max low	-0.394 (0.281)	-0.196 (0.219)	-0.156 (0.218)	-0.198 (0.219)	-0.331 (0.275)	
Max high	0.850 (0.528)	0.906 <sup>†</sup> (0.507)	1.231* (0.547)	0.917 <sup>†</sup> (0.504)	0.980 <sup>†</sup> (0.543)	
Ln war months	-0.560** (0.138)	-0.592** (0.137)	-0.531** (0.135)	-0.596** (0.132)	-0.582** (0.144)	
T	0.013** (0.004)	0.014** (0.004)	0.012** (0.004)	0.012** (0.004)	0.015** (0.004)	
Q1, G → N	-0.041 <sup>†</sup> (0.021)				-0.087** (0.033)	
Q1, N → G	0.081** (0.025)			0.095** (0.028)	0.063** (0.021)	
T: Q1, G → N	<0.001* (<0.001)			0.001** (<0.001)	0.001** (<0.001)	
T: Q1, N → G	-0.001** (<0.001)			-0.001** (<0.001)	-0.001** (<0.001)	
Q2, G → N		0.127* (0.057)		0.186* (0.078)	0.066 (0.060)	
Q2, N → G		0.221 (0.137)		0.287* (0.120)	0.098 (0.148)	
T: Q2, G → N		-0.001 <sup>†</sup> (<0.001)		-0.001* (0.001)	> -0.001 (<0.001)	
T: Q2, N → G		-0.002 (0.001)		-0.002 <sup>†</sup> (0.001)	-0.001 (0.001)	
Q3, G → N			-0.029 (0.072)			
Q3, N → G			0.048 (0.031)			
T: Q3, G → N			<0.001 (0.001)			
T: Q3, N → G			> -0.001 (<0.001)			
Q4, G → N				-0.007 (0.019)		
Q4, N → G				0.052 (0.033)		
T: Q4, G → N				<0.001 (<0.001)		
T: Q4, N → G				> -0.001 (<0.001)		
Constant	-2.193 (4.225)	-0.835 (4.369)	-5.010 (4.364)	-1.946 (4.391)	-1.207 (4.503)	-3.779** (0.349)
Observations	2,092	2,092	2,092	2,092	2,092	
Log Likelihood	-136.184	-137.544	-139.022	-140.441	-131.424	
Akaike Inf. Crit.	306.367	309.088	312.043	314.882	304.847	
% Correctly Pred. for $y = 1^{[a]}$	85.37%	85.37%	87.8%	87.8%	90.24%	

<sup>†</sup>p<0.1; \*p<0.05; \*\*p<0.01  
[a] Classification threshold is  $\Pr(y = 1) > 0.01$ .

Table VI. Governmental termination with events

	<i>Dependent variable:</i>				
	Termination: Government				
	(1)	(2)	(3)	(4)	(5)
Ln population	-0.358* (0.167)	-0.328 <sup>†</sup> (0.171)	-0.311 <sup>†</sup> (0.168)	-0.374* (0.175)	
Ln GDP p.c.	0.292 (0.194)	0.273 (0.190)	0.323 <sup>†</sup> (0.193)	0.325 <sup>†</sup> (0.192)	
ELF	-1.618* (0.777)	-1.536* (0.770)	-1.547* (0.776)	-1.553* (0.770)	
Democracy	0.035 (0.028)	0.037 (0.028)	0.031 (0.028)	0.035 (0.029)	
GINI	0.010 (0.022)	0.002 (0.022)	0.004 (0.022)	> -0.001 (0.023)	
Powersharing	0.365 (0.401)	0.283 (0.389)	0.349 (0.396)	0.324 (0.389)	
Max R disc.	-0.382 (0.977)	-0.383 (0.977)	-0.321 (0.981)	-0.445 (0.987)	
Upgrade	-0.543 (0.636)	-0.440 (0.640)	-0.593 (0.637)	-0.420 (0.641)	
Max low	-0.229 (0.544)	-0.259 (0.530)	-0.281 (0.538)	-0.230 (0.536)	
Max high	0.267 (0.306)	0.257 (0.305)	0.322 (0.291)	0.306 (0.322)	
Ln war months	-0.392** (0.093)	-0.371** (0.094)	-0.380** (0.093)	-0.368** (0.094)	
T	-0.002 (0.002)	> -0.001 (0.002)	> -0.001 (0.003)	-0.001 (0.003)	0.002 (0.002)
Q1, G → N	-0.061 <sup>†</sup> (0.036)				
Q1, N → G	0.069 (0.042)				
T: Q1, G → N	<0.001 (<0.001)				
T: Q1, N → G	> -0.001 (<0.001)				
Q2, G → N		-0.102 (0.113)			-0.171 (0.117)
Q2, N → G		0.213 (0.183)			0.179 (0.183)
T: Q2, G → N		0.001 (0.001)			0.001 (0.001)
T: Q2, N → G		-0.003 <sup>†</sup> (0.002)			-0.004 <sup>†</sup> (0.002)
Q3, G → N			-0.067 (0.083)		
Q3, N → G			0.032 (0.039)		
T: Q3, G → N			<0.001 (0.001)		
T: Q3, N → G			> -0.001 (<0.001)		
Q4, G → N				0.007 (0.023)	
Q4, N → G				-0.047 (0.051)	
T: Q4, G → N				> -0.001 (<0.001)	
T: Q4, N → G				<0.001 (<0.001)	
Constant	-0.642 (2.038)	-0.575 (1.964)	-1.223 (2.002)	-0.443 (1.954)	-3.822** (0.267)
Observations	3,309	3,309	3,309	3,309	3,309
Log Likelihood	-260.319	-257.673	-258.584	-258.780	-275.298
Akaike Inf. Crit.	554.639	549.346	551.168	551.561	562.596
% Correctly Pred. for $y = 1^{[a]}$	80.7%	87.72%	84.21%	85.96%	91.23%

<sup>†</sup>p<0.1; \*p<0.05; \*\*p<0.01

[a] Classification threshold is  $\Pr(y = 1) > 0.01$ .

Table VII. In-sample (1996-2012) accuracy measures

	<b>BCG</b>	<b>BCG + events</b>	<b>Events</b>
Territorial onset AUC	0.882	0.890	0.758
Territorial onset Brier score	0.002	0.002	0.002
Governmental onset AUC	0.847	0.886	0.669
Governmental onset Brier score	0.002	0.002	0.002
Territorial termination AUC	0.881	0.902	0.677
Territorial termination Brier score	0.016	0.014	0.019
Governmental termination AUC	0.761	0.860	0.834
Governmental termination Brier score	0.006	0.006	0.006

Our results indicate more generally that models based on structural covariates do better for territorial than governmental conflict onset, likely reflecting that the concept of horizontal inequalities was developed to account for the former rather than the later type of conflict. Adding events can improve the performance of the model more when the structural model fits less well, as demonstrated by the larger AUC for the territorial termination model with events. However, large subsequent improvements will be more difficult the better the fit of the prior baseline.

It is instructive to look at some of cases where model predictions diverge to understand the sources. Pure event based models generally do much worse than the models with structural risk based indicators. The lack of information relevant to prior grievances make these unable to distinguish between unorganized riots and organized dissent under situations with plausible grievances.

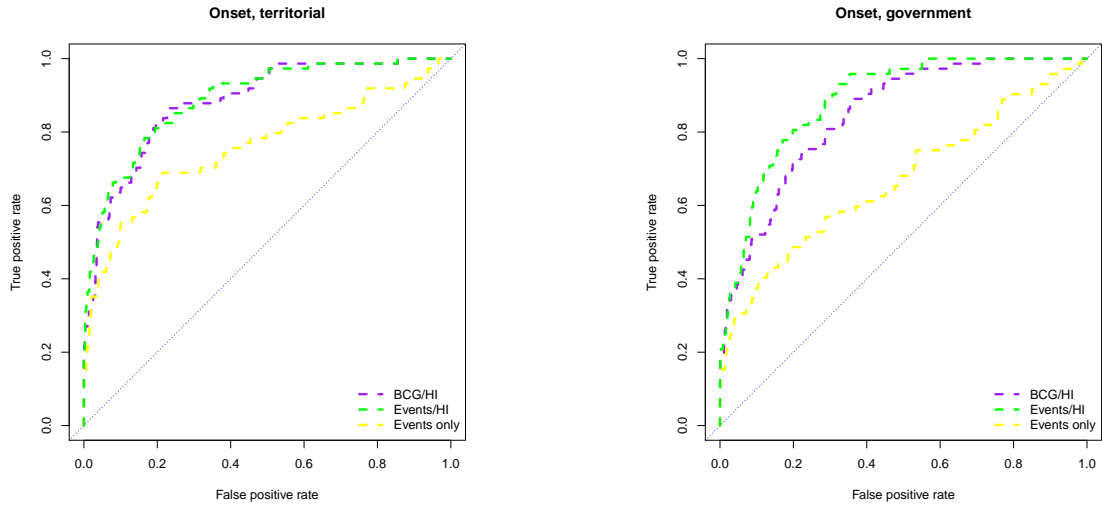


Figure 2. ROC onset, by type

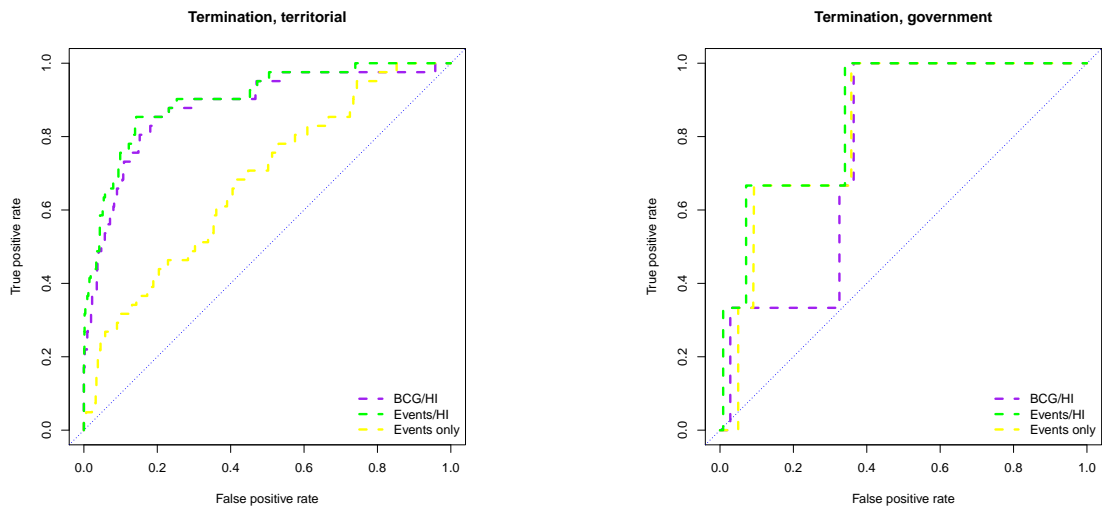


Figure 3. ROC termination, by type

For example, Australia has a high count of conflict events in various months in 2005, and thus a relatively high predicted risk of conflict according to pure event based models. These events seem to reflect the Cronulla riots, where surfers and youth of Middle Eastern origin clashed in a Sydney beachside suburb, leading to large police operations to control the violence (see Poynting, 2006). Violent riots such as these are unlikely to escalate to civil wars, given the lack of organizations and

clear incompatibilities over government or territory.

## **Out-of-sample results**

A good in-sample fit does not imply good out-of-sample performance on other data, given the problem of potential overfitting to the idiosyncracies of a specific sample (see, e.g. Morgan & Henrion, 1990; Ward, Greenhill & Bakke, 2010). Measures that trade off fit and complexity such as AIC scores can help remedy some of the problems from overfitting, but may not fully overcome them.

We first check the out-of-sample predictive ability of our three models of onset and termination through cross-validation analyses, where we train the models on one subset on the data and then compare how well they do in predicting to other subsets of the data. We randomly split the data into four slices, and then train each model using three-fourths of the data and use these estimates to predict to the remaining one-fourth of the data. We repeat the process partition process 10 times, and report the AUCs in the ROC curves over the replications in Table VIII. Consistent with our in-sample results we find that the model with structural covariates fit much better out-of-sample than the pure event model. However, adding the events to the structural model also provide an improvement over the original model without the events for all cases (albeit the AUCs are almost identical for territorial onset). Moreover, the AUCs are generally higher for conflict onset than conflict termination. On balance, we see this as evidence that event information in general can contribute a structural model of conflict onset and termination can vary by sample, although the improvement is often modest and vary considerably across samples.

Table VIII. Cross validation results

	<b>Onset</b>		<b>Termination</b>	
	Territorial	Government	Territorial	Government
BCG/HI	0.864	0.819	0.810	0.672
BCG/HI + events	0.865	0.863	0.821	0.695
Events only	0.693	0.660	0.631	0.651

Cell entries are AUC scores.

## Out-of-sample conflict forecast

All our models have been estimated on data up to 2012. In this section we look at how the model could generate cumulative forecasts of risk over the subsequent two years for which we have UCDP data, i.e., 1 January 2013 through 2014. We aggregate the predicted probabilities  $\hat{p}_{i,t}$  to measure of predicted risk over the overall forecast period, i.e.,  $\hat{p}_i^\dagger = 1 - \prod_{t=\tau}^{t=\tau+23} (1 - \hat{p}_{i,t})$ , where  $\tau$  is an indicator for January 2013. Figure 4 displays ROC plots for the two onset predictions and the two termination predictions over the period, while Table IX reports the AUC and Brier scores. The plot for territorial conflict suggests some improvement in the out-of-sample model for the model with events in this period. However, this does not extend to governmental conflict onsets, where all the models do worse in absolute terms and the model with events does worse than the purely structural model. For conflict termination we only have three territorial and three governmental events over the 2013 and 2014 periods, and the ROC curves look distinctively jagged.

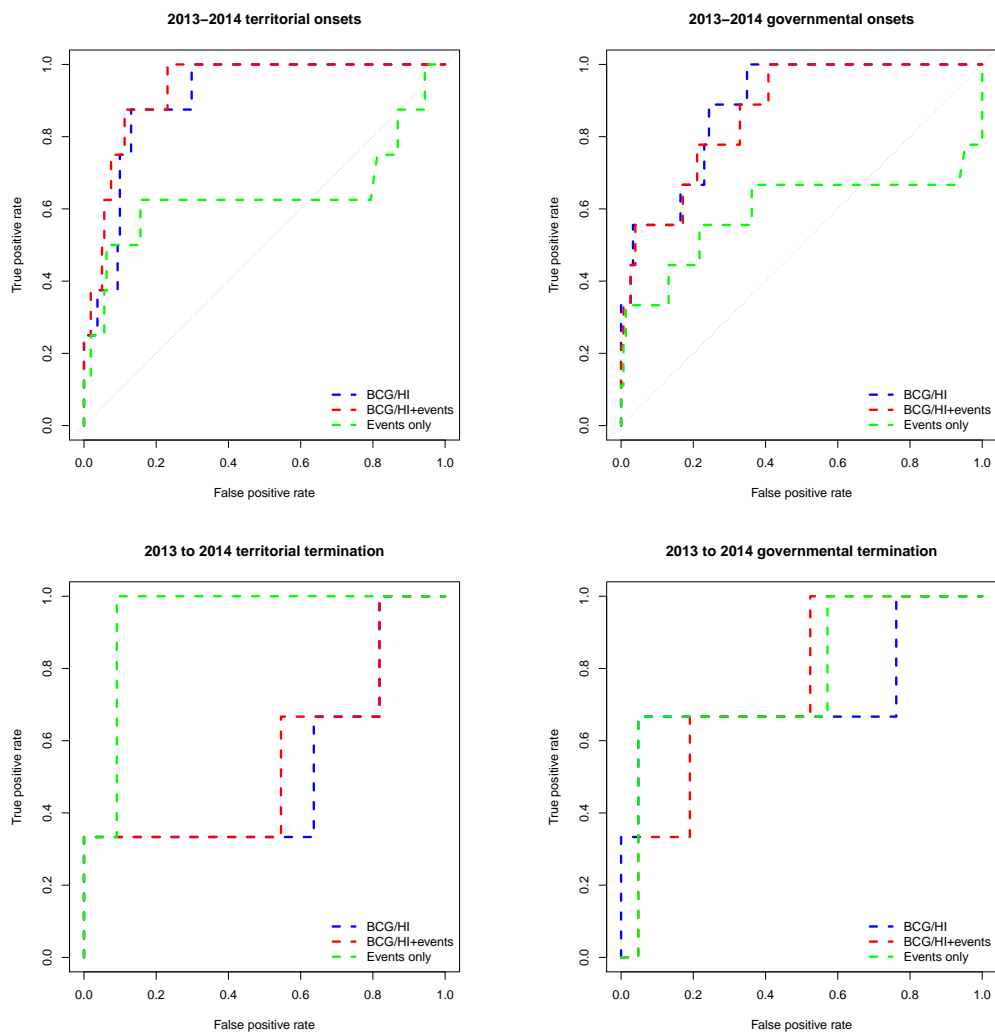


Figure 4. ROC predictions, 2013 to 2014

Table IX. AUC + Brier for out-of-sample predictions for 2013-2014

		<b>BCG</b>	<b>BCG + events</b>	<b>Events</b>
Territorial onset AUC		0.905	0.932	0.636
Territorial onset Brier score		0.038	0.036	0.041
Governmental onset AUC		0.884	0.868	0.592
Governmental onset Brier score		0.041	0.038	0.040
Territorial termination AUC	31	0.515	0.545	0.939
Territorial termination Brier score		0.515	0.411	0.294
Governmental termination AUC		0.730	0.746	0.778

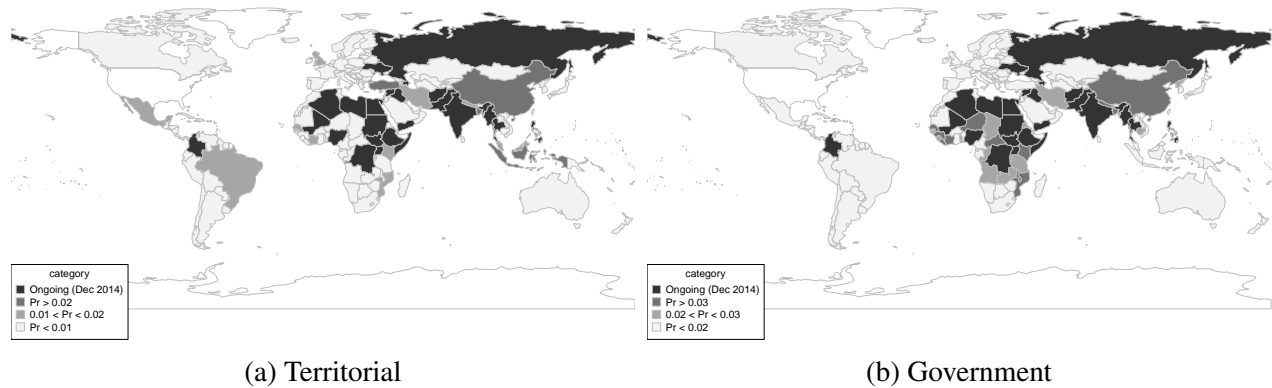


Figure 5. Forecast for January – December 2015

We then use our results to develop a conflict forecast for the January to December 2015 period beyond the most recent UCDP data. We generate the predictions using the model estimates and event data available up through December 2015. This exercise is helpful as a supplement more in the spirit of a true out-of-sample forecast, given that the response is already observed in our other results, even if not included in the training data.

We plot the predicted probabilities of conflict onset in Figure 5 (countries with ongoing conflict are indicated in black, and note that we do not have event data for the USA). Predicted probabilities of conflict over governmental incompatibilities are generally higher than those over territorial incompatibilities. Although the predicted risks of new conflicts are generally low, we see relatively high predicted values of territorial conflict (left panel) for Indonesia (0.07), China (0.05), and Turkey (0.03), and governmental conflict (right panel) for Central African Republic (0.10), China (0.07), Liberia (0.06), and Mozambique (0.06). The higher values for Mozambique reflects a higher number of  $G \rightarrow N$  material conflict events from April 2015 through December 2015 for a country that normally has few events.



## Discussion and conclusions

We have examined the predictive ability of a model reflecting horizontal inequalities on the onset and termination of territorial and governmental civil conflict and to what extent this can be improved by lagged directed dyad event data, both in-sample and out-of-sample. We find that anchoring prediction models in theories—and recognizing that different types of conflict have different risk factors and potential actors—can yield models with reasonable predictive ability, even for rather rare conflict onset and termination events. Adding event counts distinguishing between types and specific interactions can provide some additional contribution, and different types of events are helpful for assessing the likelihood of onset and termination. However, the improvement is modest and varies a great deal across conflict outcome and samples. In general, it appears as if the value of the simple event counts used here is higher the less well-suited the structural predictors. In sum, simple event counts can help provide leading indicators, but they do not provide a magic solution for predicting new conflict onsets or terminations.

There are a number of possible limitations of our analysis and plausible avenues for improving predictive power. Our research and that of Buhaug et al. demonstrate how theoretical attention to relevant actors for particular types of conflict can help generate improved empirical measures. But the efforts presented here are primarily helpful to identify motives and opportunities for territorial incompatibilities as well as how accommodation and conflict management can predict to termination. Our current thinking is much less advanced for governmental conflict, and better attention to issues such as measures of mobilization and security force organization may help improve predictive ability here.

With regards to the event data we may be able to do better by distinguishing between more specific event codes and whether actions are carried out by actors that have previously been involved in conflict for onset (or are currently in conflict, for termination). Moreover, there may be gains from trying to identify whether events pertain to specific incompatibilities, and whether verbal events are ‘retrospective’ comments on things that have already happened or prospective statements or threats, where we would expect the latter to be more helpful for forecasting. The new Phoenix project (<http://openeventdata.org/>) will provide freely available daily coded data using open-source software, which may help overcome the restrictions on the ICEWS data.

Finally, absolute predictive power can also be improved by alternative estimation approaches. We have deliberately kept our models simple to make it easier to assess individual contributions from the results and allow comparisons with existing work. That said, there are likely to be possible advances in predictive ability from models from machine learning often shown to have advantages for prediction, taking into account spatial dependencies and spill-over mechanisms, non-parametric approaches and random effects models, as well as model pruning to remove covariates unhelpful for forecasting. We do not think, however, that pure changes in methods or the baseline model are likely to change our main conclusions about the relative contribution of the current event measures.

In sum, although prediction is difficult—especially about the future—our results support the value of more theoretically grounded measures of structural characteristics reflecting grievances and opportunities, and provide a proof of concept that more more dynamic behavioral information can help forecasting conflict onset and termination.

## **Replication data**

The dataset and code to replicate the empirical analysis reported in this article can be found at

<http://www.prio.org/jpr/datasets>.

## **Acknowledgments**

This research was conducted for the Political Instability Task Force, which is funded by the Central Intelligence Agency. The views expressed herein are the authors' alone and do not necessarily represent the views of the Task Force or the U.S. Government. Gleditsch is also grateful for support from the Research Council of Norway (213535/F10) and European Research Council (313373).

The authors are listed in alphabetical order, equal authorship implied. We are grateful for helpful comments from Patrick Brandt, Robert Blair, Erica Chenoweth, Michael Colaresi, James Fearon, Andrew Halterman, Håvard Hegre, Grace Scarborough, Michael Ward, as well as the editors and anonymous reviewers.

## References

- Bartusevičius, Henrikas (2014) The inequality-conflict nexus reexamined: Income, education and popular rebellions. *Journal of Peace Research* 51(1): 35–50.
- Bond, Doug; Joe Bond, Churl Oh, J Craig Jenkins & Charles Lewis Taylor (2003) Integrated data for events analysis (IDEA): An event form typology for automated events data development. *Journal of Peace Research* 40(6): 733–745.
- Boschee, Elizabeth; Jennifer Lautenschlager, Sean O’Brien, Steve Shellman & James Starz and Michael Ward (2015) ICEWS coded event data. <http://dx.doi.org/10.7910/DVN/28075>, Harvard Dataverse, V4.
- Buhaug, Halvard (2006) Relative capability and rebel objective in civil war. *Journal of Peace Research* 43(6): 691–708.
- Buhaug, Halvard; Lars-Erik Cederman & Kristian Skrede Gleditsch (2014) Square pegs in round holes: Inequalities, grievances, and civil war. *International Studies Quarterly* 58(2): 418–431.
- Cederman, Lars-Erik & Kristian Skrede Gleditsch (2009) Special issue on ‘disaggregating civil war’. *Journal of Conflict Resolution* 53(4): 487–495.
- Cederman, Lars-Erik; Kristian Skrede Gleditsch & Julian Wucherpfennig (2017) Predicting the decline of ethnic civil war: Was Gurr right and for the right reasons? *Journal of Peace Research* 54(X): XXX–XXX.

- Chadefaux, Thomas (2014) Early warning signals for war in the news. *Journal of Peace Research* 51(1): 5–18.
- Choucri, Nazli & Thomas W Robinson, eds (1978) *Forecasting in International Relations: Theory, Methods, Problems, Prospects*. San Francisco, CA: W.H. Freeman.
- Clayton, Govinda & Kristian Skrede Gleditsch (2014) Will we see helping hands? Predicting civil war mediation and likely success. *Conflict Management and Peace Science* 31(3): 265–284.
- Davenport, Christian (2007) State repression and political order. *Annual Review of Political Science* 10: 1–23.
- Dewal, Snigdha; Jack Goldstone & Michael Volpe (2013) Forecasting stability or retreat in emerging democratic regimes. *Politics and Governance* 1(1): 32–47.
- Esty, Daniel C.; Jack Goldstone, Ted Robert Gurr, Pamela T. Surko & Alan N. Unger (1995) *The State Failure Task Force Report*. McLean, VA: Science Applications International Corporation.
- Freeman, John R & Joshua S Goldstein (1989) US-Soviet-Chinese relations: Routine, reciprocity, or rational expectations? *American Political Science Review* 85(1): 17–35.
- Ghodes, Anita R & Sabine Carey (2017) Canaries in a coal-mine? The killing of journalists as a precursor to increased repression. *Journal of Peace Research* 54(X): XXX–XXX.
- Gleditsch, Kristian Skrede & Steve Pickering (2014) Wars are becoming less frequent: A reply to Harrison and Wolf. *Economic History Review* 67(1): 214–230.

- Gleditsch, Kristian Skrede & Andrea Ruggeri (2010) Political opportunity structures, democracy, and civil war. *Journal of Peace Research* 47(3): 299–310.
- Gleditsch, Kristian Skrede & Michael D Ward (1999) Interstate system membership: A revised list of independent states since 1816. *International Interactions* 25: 393–341.
- Gleditsch, Kristian Skrede & Michael D Ward (2013) Forecasting is difficult, especially about the future: Using contentious issues to forecast interstate disputes. *Journal of Peace Research* 50(1): 17–31.
- Gleditsch, Nils Petter; Peter Wallensteen, Mikael Erikson, Margareta Sollenberg & Håvard Strand (2002) Armed conflict, 1945-99: A new dataset. *Journal of Peace Research* 39(5): 615–637.
- Goldstein, Joshua S (1992) A conflict-cooperation scale for WEIS events data. *Journal of Conflict Resolution* 36(2): 369–385.
- Goldstone, Jack A; Robert H Bates, David L Epstein, Ted R Gurr, Michael B Lustik, Monty G Marshall, Jay Ulfelder & Mark Woodward (2010) A global model for forecasting political instability. *American Journal of Political Science* 54(1): 190–208.
- Kennedy, Ryan (2015) Making useful conflict predictions: Methods for addressing skewed classes and implementing cost-sensitive learning in the study of state failure. *Journal of Peace Research* 52(5): 649–664.
- King, Gary & Will Lowe (2003) An automated information extraction tool for international conflict data with performance as good as human coders: A rare events evaluation design. *International Organization* 57(3): 617–642.

- King, Gary & Langche Zeng (2001) Improving forecasts of state failure. *World Politics* 53(4): 623–658.
- Lichbach, Mark I (1995) *The Rebel's Dilemma*. Ann Arbor, MI: Michigan University Press.
- Melander, Erik; Therése Pettersson & Lotta Themnér (2016) Organized violence, 1989-2015. *Journal of Peace Research* 53(5): 727–742.
- Morgan, M Granger & Max Henrion (1990) *Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis*. Cambridge: Cambridge University Press.
- Muchlinski, David; David Siroky, Jingrui He & Matthew Kocher (2016) Comparing random forest with logistic regression for predicting class-imbalanced civil war onset data. *Political Analysis* 24(1): 87–103.
- Pevehouse, Jon C & Joshua S Goldstein (1999) Serbian compliance or defiance in Kosovo? Statistical analysis and real-time predictions. *Journal of Conflict Resolution* 43(4): 538–546.
- Goldstein, Joshua S & John C Pevehouse (1997) Reciprocity, bullying, and international cooperation: Time-series analysis of the Bosnia conflict. *American Political Science Review* 91(3): 515–530.
- Pinker, Steven (2011) *The Better Angels of Our Nature: Why Violence Has Declined*. New York: Viking.
- Poynting, Scott (2006) What caused the Cronulla riot? *Race & Class* 48(1): 85–92.

- Ruggeri, Andrea; Theodora-Ismene Gizelis & Han Dorussen (2011) Events data as Bismarck's sausages? Intercoder reliability, coders' selection, and data quality. *International In* 37(3): 340–361.
- Rummel, Rudolph J (1963) Dimensions of conflict behavior within and between nations. *General Systems* 8: 1–50.
- Schrodt, Philip A (2011) Forecasting political conflict in Asia and the Middle East using latent Dirichlet allocation models. Presented at 'New Horizons in Conflict System Analysis: Applications to the Middle East', University of South Carolina, 28-30 October.
- Schrodt, Philip A (2012) CAMEO conflict and mediation event observations event and actor code-book version 1.1b3.
- Schrodt, Philip A & Deborah J Gerner (1994) Validity assessment of a machine-coded event data set for the Middle East, 1982-92. *American Journal of Political Science* 38: 825–854.
- Shellman, Stephen (2004) Time series intervals and statistical inference: The effects of temporal aggregation on event data analysis. *Political Analysis* 12(1): 97–104.
- Sobek, David & Caroline Payne (2010) A tale of two types: Rebel goals and the onset of civil wars. *International Studies Quarterly* 54(1): 213–240.
- Vogt, Manuel; Nils-Christian Bormann, Seraina Rüegger, Lars-Erik Cederman, Philipp Hunziker & Luc Girardin (2015) Integrating data on ethnicity, geography, and conflict: The ethnic power relations data set family. *Journal of Conflict Resolution* October 59(7): 1327–1342.



- Ward, Michael D; Brian D Greenhill & Kristin M Bakke (2010) Perils of policy by p-value: Predicting civil conflicts. *Journal of Peace Research* 47(4): 363–375.
- Ward, Michael D & Andreas Beger (2017) Lessons from near real-time forecasting of irregular leadership changes. *Journal of Peace Research* 54(X): XXX–XXX.
- Ward, Michael D; Nils W Metternich, Cassy L Dorff, Max Gallop, Florian M Hollenbach, Anna Schultz & Simon Weschle (2013) Learning from the past and stepping into the future: Toward a new generation of conflict prediction. *International Studies Review* 15: 473–490.
- Weidmann, Nils B & Michael D Ward (2010) Predicting conflict in space and time. *Journal of Conflict Resolution* 54(6): 883–901.
- Witmer, Frank; Andrew Linke, John O’Loughlin, Andrew Gettelman & Arlene Laing (2017) Climate-sensitive violent conflict forecasts for sub-Saharan Africa, 2015-2065. *Journal of Peace Research* 54(X): XXX–XXX.
- Wucherpfennig, Julian; Nils Metternich, Lars-Erik Cederman & Kristian Skrede Gleditsch (2012) Ethnicity, the state, and the duration of civil wars. *World Politics* 64(1): 79–115.

## **Biographies**

DAINA CHIBA, b. 1977, PhD in Political Science (Rice University, 2012); Lecturer, Department of Government, University of Essex (2013- ); Current research interests: violent conflict, international institutions, and political methodology.

KRISTIAN SKREDE GLEDITSCH, b. 1971, PhD in Political Science (University of Colorado, Boulder, 1999); Professor, Department of Government, University of Essex (2005- ); Research Associate, Peace Research Institute Oslo (2003- ). Current research interests: violent and non-violent conflict, democratization, and political violence. Most recent book: *Inequality, Grievances, and Civil War* (Cambridge University Press 2013, with Lars-Erik Cederman and Halvard Buhaug).