LEARNING FROM THE PAST AND STEPPING INTO THE FUTURE: THE NEXT GENERATION OF CRISIS PREDICTION

MICHAEL D. WARD, NILS W. METTERNICH, CASSY DORFF, MAX GALLOP, FLORIAN M. HOLLENBACH, ANNA SCHULTZ, AND SIMON WESCHLE

Abstract. Developing political forecasting models not only increases the ability of political scientists to inform public policy decisions, but is also relevant for scientific advancement. As part of a larger project, a team at Duke University created a series of geographically informed statistical models for conflict prediction. The generated predictions have been highly accurate, with few false negative and positive categorizations. Predictions are made at the monthly level for six months periods into the future, taking into account the social-spatial context of each individual country. The model has a high degree of accuracy in reproducing historical data measured monthly over the past 10 years, and is approximately equally accurate in making forecasts. This paper surveys the notion of forecasting and demonstrates the utility of creating forecasting models for predicting political conflicts in a diverse range of country settings. Apart from the benefit of making actual predictions, we argue that predictive heuristics are one gold standard of model development in the field of conflict studies and that the predictive heuristics shed light on an array of important components of the political science literature on conflict dynamics.

1. Introduction

"Ethnicity, Insurgency and Civil War" (Fearon and Laitin, 2003*a*) is one of the most venerated and cited articles about the onset of civil wars. The article has over 3000 citations in scholar.google.com and 755 citations in the Web of Science (as of August, 2012). It has been cited prominently in virtually every social science discipline, ranging from *Acta Sociologica* to *World Politics*; and it is the most downloaded of published articles from the *American Political Science Review*.[1] Clearly, this article is regarded as an important, even foundational piece of scholarship. In summer of 2012, this article was used by Jacqueline Stevens in an New York Times op-ed as evidence that political scientists are bad forecasters. That claim was wildly off the mark, in that Fearon and Laitin do not deal with forecasting at all, and Stevens ignored actual forecasting efforts in political science. We argue that conflict research in political science can be improved by more, not less, attention to predictions. The increasing availability of disaggregated data and advanced estimation techniques are making forecasts of conflict more accurate and precise. In addition, we argue that forecasting helps to prevent over-fitting, and can be used both to validate models, and inform policy makers.

Although the Fearon and Laitin article is not about out-sample prediction, it uses in-sample prediction to tell its main story. Figure 2 in Fearon & Laitin (p. 82), presents the probability of civil war onsets over a five-year period, conditional on ethnic homogeneity and GDP per capita, showing that the latter has a more profound effect than the former on the 5-year probability of civil war onset. Given the prominence of this research, and the transparency of the research program, we use their framework to predict civil war onsets in the period from 2000 to 2009.[1] We use their research to illustrate how prediction can inform us about the validity of models. We undertake a replication Fearon & Laitin's Model (1) for the year 1999,[2] the last year in the Fearon & Laitin dataset, and use the laws of probability to calculate the cumulative probability of civil wars in each country for each of the ten subsequent years

Date: December 5, 2012 Version 1.0.
This project was partially supported by the Defense Advanced Research Projects Agency and subsequently the Office of Naval Research (holding grants to the Lockheed Martin Corporation, Contracts FA8650-07-C-7749 and N00014-12-C-0066, respectively). For helpful insights we thank Philippe Loustaunau, Mark Hoffman, Scott de Marchi and Will Moore, as well as our colleagues at Duke. All the bad ideas and mistakes are our own. An earlier version of the paper was presented at the 2nd International Conference on Cross-Cultural Decision Making. Corresponding author: `Michael D. Ward`).

[1] `http://www.apsanet.org/content_30489.cfm`; August 8, 2012

Table 1. Top 10 cumulative predicted probabilities for countries with at least one Civil War onset between 2000 and 2009. Probabilities in years with onsets are **bold**.

Country	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009
Nigeria	.15	.27	.38	.47	**.54**	.15	.27	.38	.47	**.54**
Iran	.07	.13	.18	.24	.29	**.33**	.07	.13	.18	.24
Angola	.06	.12	.17	.22	**.27**	.06	.12	**.17**	.06	**.12**
Niger	.04	.07	.10	.14	.17	.20	.23	**.25**	.03	.07
Azerbaijan	.05	.10	.15	.20	.24	**.28**	.05	.10	.15	.20
Burundi	.02	.04	.07	.09	.11	.13	.15	.17	**.18**	.02
Dem. Rep. Congo	.03	.05	.08	.10	.13	.15	**.17**	.03	.05	.08
Peru	.02	.04	.06	.07	.09	.11	.13	**.14**	.02	.04
Mali	.02	.03	.05	.07	.08	.10	.11	**.13**	.02	.03
Georgia	.02	.05	.07	.09	**.11**	.02	.05	.07	**.09**	.02

(2000-2009).[3] We call this approach cumulative, because the probability mounts up over time when there is no onset. Table 1 shows the yearly cumulative predicted probabilities of civil war onset for the ten highest-probability countries that had an onset.[4] Probabilities are in **bold** when an onset actually occurred. This model does point to civil war onsets, such as Nigeria in 2004 and 2009 or Iran in 2005. Since the probabilities increase if onset does not occur, there are also a large number of countries that have high cumulative predicted probabilities, but where no onset occurred between 2000 and 2009. Examples are Kenya, Romania, Tajikistan, Tanzania, or Guatemala. They all have predicted cumulative probabilities similar to or even higher than those for the countries in Table 1.

Table 2 gives the number of correctly predicted civil wars as well as the number of false positives, taking different thresholds for the cumulative predicted probabilities, above which the model is considered to indicate a civil war onset. Using a probability threshold of 0.5, the model predicts only two out of the 33 onsets, but has no false positives. As the threshold is lowered, the number of correct predictions goes up, and so does the number of false positives. When using 0.1 as the cutoff, 18 onsets are correctly predicted, but at the same time the model forecasts 244 onsets that did not happen. Despite the fact that many of the variables used in the Fearon & Laitin study are statistically significant, the predictive accuracy of the model out-or-sample is not large.

Table 2. Number of correctly predicted onsets and false positives for cumulative predicted probabilities at varying cut-points for cumulative predicted probabilities.

Threshold	**Correctly Predicted**	**False Positives**
0.5	2/33	0
0.3	3/33	26
0.1	15/33	256
0.05	21/33	508

Table 3 gives common performance statistics for each year, using a threshold value of 0.1 where necessary. Both the proportion of correctly identified positives and negatives (sensitivity and specificity) are about 60%, but this percentage varies greatly by year. Accuracy measures the proportion of observations that are correctly classified (true positives and true negatives), which is again about 60%. Precision is the proportion of true positives out of all predicted positives. This is approximately 5%, meaning that if the model predicts an onset, it is wrong 95% of the time. Another way to evaluate predictions, which is employed here, is the Brier score (Brier, 1950), defined as the average squared deviation of the predicted probability from the true event. The Brier score is one of the few strictly proper scoring rules for predictions with binary outcomes (Gneiting and Raftery, 2007). Brier scores closer to zero indicate better predictive performance. This model's Brier scores are low, because the model predicts *peace* for many countries that indeed do not have an onset of civil war, and are thus in peace. But predicting the non-onset of civil war is not the main point of the endeavor. The AUC scores simply underscore that the model is not very accurate at predicting *conflict onsets*. The last row in the table provides an estimate of the cumulative number of global onsets predicted by the model in each year (20 estimates of 0.05 should produce one onset). The model vastly over-predicts civil war onsets.

Table 3. Performance statistics by year for cumulative predicted probabilities approach for a threshold of 0.1 (where necessary).

Statistic	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	Overall
Sensitivity	0	0	0	0	.571	.500	.500	1	.500	.750	.455
Specificity	.992	.953	.891	.859	.823	.780	.721	.693	.659	.622	.800
Accuracy	.969	.939	.878	.840	.809	.771	.718	.702	.656	.626	.791
Precision	0	0	0	0	.154	.067	.027	.093	.022	.059	.055
Brier	.023	.017	.018	.029	.046	.030	.023	.035	.029	.039	.029
AUC	.346	.426	.585	.367	.730	.693	.806	.854	.748	.689	.651
Nr. Civil War Onsets	3	2	2	3	7	4	2	4	2	4	33
Predicted Onsets	2	4	5	7	8	9	10	11	12	13	81

The most problematic aspect of the Fearon & Laitin model, however, is that many country-years that have a high predicted probability of conflict actually have no onset. It does well at predicting *peace*, but poorly at predicting *conflict*. Consider the following medical analogy. Say a test was developed to detect some disease that about 35 out of $1,000$ people in the population have. If everybody is tested, for the 35 who have the disease, the test will tell 5 of them that they are infected, while the other 30 are told they are healthy. Out of the 965 without the disease, 37 are diagnosed and receive treatment. This means that only about 12% of the treated are actually sick and 85% of the sick are left untreated. These are the performance statistics of the Fearon & Laitin model.[5]

Using the seminal work of Fearon & Laitin as an example, we have shown that a model with a number of statistically significant variables may nevertheless be poorly-equipped to tell us what we (and policy makers) are really interested in: when and where do civil wars occur? The fact that, for example, the coefficient for oil exporter is positive and significant in a regression does *not* necessarily mean that focusing on this variable (among others) will equip us better to explain why civil war broke out (or will break out) in a certain country but not in another one. The poor predictive performance is not an indictment of Fearon & Laitin's contribution, nor is it evidence that prediction is too treacherous to attempt. Rather, it points to an opening for social scientists, as to the benefits of adopting a predictive framework. First, it establishes a framework for rigorous and ongoing cross-validation of our models. Second, this cross-validation offers us the opportunity to test our theories, their scope, and their portability, which can provide valuable input in the theory-building process. Third, generating predictions makes the implications of our research more accessible to the the policy community and the general public.

1.1. **Model validation.** Although cross-validation techniques are well known in political science, they are not frequently used. Cross-validation is useful to minimize over-fitting and maximize predictive power. The prediction framework incorporates two types of cross-validation. Initially, it subsets the data into a training set and a test set. Coefficients for independent variables are estimated on the training set, and then used to predict outcomes in the test set. This provides an immediate source of cross-validation. Second, prediction invites the collection of new data and reevaluation of the model. Though the data was partitioned for in-sample versus out-of-sample predictions, there is a certain risk of what one might want to call "second-order over-fitting," in which models are optimized to make good out-of-sample predictions on a particular test set. From our experience this kind of over-fitting is especially likely in models with a small number of cases, a large number of variables, and short time series. Hence, we believe model evaluation by out-of sample prediction is especially powerful if the out-of sample data was not available to the researcher during the modeling phase. This is especially important in an arena wherein most analyses are done with observational, as opposed to experimental, data.

1.2. **Theory Testing.** An oft-heard claim, especially in the various responses to Stevens's recent New York Times Op-Ed on political science,is that political science is more interested in explanation rather than prediction. But given what we have shown above, we should be wary of saying that we can *explain* what causes civil war, even though we have a number of statistically significant variables. Assessing the quality of statistical models using prediction can tell us how well we understand social phenomena, and help us improve our explanations of where and when they occur. Building theory-driven forecast models, with the concomitant continual availability of new out-of-sample data for testing, will lead social science research toward better theory and explanations, not just better predictions.

Prediction also creates particular incentives which are may be useful for political science as a whole. Because true out-of-sample prediction could involve a shifting context, it requires integrated theories to be successful. When conducting all analyses *ex post*, the researcher will choose the theory she thinks is most likely to apply, often focusing on the novel or unusual. For example, parts of the Arab Spring were organized by social media, hence many commentators focused on this novel tool as an explanation for the success of the revolution. However, there were a host of other factors that made revolutions in Arab world highly likely. High unemployment, low growth rates, aging dictators, and religious divisions are long-standing explanations for popular uprisings that were all present in this context. When creating predictions, it becomes much harder to inadvertently select advantageous models for a particular context. The choices between competing models have to be "endogenized" in the model in order to have a durable and portable tool for prediction. This requirement–that we delimit the context in which our theory applies–makes our understanding and use of theory more precise. Making models portable across time also brings our focus to what societies have in common, leading us toward a more systematic understanding of social processes.

1.3. **Policy relevance.** Developing theoretically motivated, cross-validated models can make our findings more accessible to a wider audience. Sharing the knowledge generated by our research is an important part of the enterprise. In addition, the practice of generating these predictions, for researchers, gives us a reminder of and an answer to the often embarrassing question, "so what?" Predictions bind our independent variables to outcomes in a concrete way, bringing clarity, both to ourselves and to others, as to the mechanisms in our models. Current events make predictions of civil conflict even more desirable. The Arab Spring and its aftermath, continuing violence in Afghanistan, and sudden agreements between Georgia and Russia always raise the same question: could we have predicted these events to prepare for or even influence their emergence? Our profession struggles to do so. There are, of course, many who are interested in generating a predictive, analytical social science in the policy realm. During the Vietnam war, Jeffry S. Milstein, a new Ph.D. from Stanford University, conducted quantitative and simulation studies that were among the first-ever, real predictions in the discipline of international relations (Milstein and Mitchell, 1968; Milstein, 1974). These were aimed at elucidating the dynamics of the then ongoing conflict in South East Asia. Robert McNamara, the US Secretary of Defense, was promoting a systems analytic perspective which suggested that a war of attrition would allow the US to outlast the Viet Cong. Milstein's analysis was the first (outside of the DoD) to illustrate that the dynamics were likely to turn out differently. Since Milstein's early work, there have been a variety of efforts within the policy community to craft accurate predictive models that can guide policy.[6]

This article is a reminder that despite every crisis's unique features, we, as a discipline, should strive for the main prize: the identification of general mechanisms that allow us to make predictions about future events. In fact, the ability to predict future crises can be understood as the gold standard to scientifically advance the study of conflict, peace, and crises. The goal is to have something as theoretically sound as the Fearon & Laitin models that can *also* actually explain the data that have not yet been collected; namely, the future.

2. A Brief Review of Some Forgotten Literature

2.1. Advances in data.

2.1.1. *Where it all began: Richardson's scientific definition of war and the war-list format.*

> **Astronomer:** ... My own interest is not in "wars and how to win them" after the fashion of military history, but instead in "wars and how to prevent their occurence."

> **Pacifist:** But wouldn't it be more fruitful to begin your inquiry from the dates, not of wars, but of great co-operations, such as the Postal Union, the sale of Alaska, the regulation of radio-wavelengths, the foundation of the International Labor Office, and the beginnings of international language?
> A great mistake has been made by some peace societies which spend their time in discussing war. They should take note of Coué's maxim that when the will and the imagination act oppositely the imagination wins. If you intend peace but imagine wars, it is war that you will get. The peace societies should forget wars and think about the hundreds of forms in which nations actually do co-operate. The League of Nations has happily published a list of them: the *Répertoire des Organisations Internationales*; and Cruttwell has published *A History of Peaceful Change in the Modern World* (1937).
> **Author:** Splendid! Then I need not attempt anything of that kind.

This imagined dialogue ends the introduction to the founding book of the quantitative study of conflict: *The Statistics of Deadly Quarrels* by Richardson (1960). It is the start of the systematic collection of wars with the intent to explain and predict their occurrence, escalation, duration, termination, and spread. Richardson's analytic and systematic exploration of war is very much in the spirit of Karl Deutsch's introduction to Quincy Wright's *Study of War*, when he writes that "war, to be abolished, must be understood. To be understood, it must be studied." In *The Statistics of Deadly Quarrels* Richardson presents a list of more than 300 wars, between 1820 to 1949. Interestingly, he does not restrict his collection to a particular type of war but simply introduces a threshold for a violent conflict to be included. Defining wars by an intensity threshold has been one of the most important trademarks of the quantitative conflict literature until recently. In fact, the "objective" definition of war is what makes Richardson's war list the first scientific collection of conflict rather than the law inspired war list of Wright (1942) or Sorokin's (1937) list that selects on "greater" nations.

Richardson chooses to represent his data in a war-list format where each war is represented by separate tables and the names of the contestants are in the row and columns. For example, contestants that fought on side A would be in the rows and contestants that fought on side B would be in the columns. In the theoretical case that everybody fights everybody, the table would turn into a square matrix to represent all possible violent interactions. In the cells, Richardson codes when within the war the two contestants fought each other, the state of pre-war relations, war aims, and outcomes. But even though this format may strike the contemporary researcher as somewhat odd, it is to a large extent a tribute to Richardson not having a computer. In fact, he lays out a number of alternative formats that hold the same information and are now used commonly (or at least hybrids) in the study of conflict. (1) Geocoding wars by location or mapping them, (2) representing belligerent-dyads in one big matrix, (3) by date, (4) by alphabetical order, (5) or by importance.

2.1.2. *Advancing the quantitive study of conflict: The country-year format.* Inspired by the work of Pitirim Sorokin, Lewis Frye Richardson, and Quincy Wright, J. David Singer founded the Correlates of War Project in 1963 at the University of Michigan with the goal to systematically accumulate scientific knowledge about war. The impact of COW on the discipline cannot be overstated. It formed the gold-standard of quantitative conflict research in the twentieth century. The collaborative effort of Melvin Small and J. David Singer collecting information on inter-state and extra-state war lead to the publication of *The Wages of War* in 1972. It marks the beginning of the country-year format that has dominated conflict datasets and how the disciple thinks about conflict processes.

The peak of COW-inspired research and its intellectual advances is probably best documented in *The Process of War* (1995) edited by Stuart A. Bremer and Thomas R. Cusack, which concentrates on Militarized Interstate Disputes that formed another hallmark of COW enabled research. However, as the market leader of conflict data the COW dataset missed an important development in conflict studies: The heightened interest in civil conflicts and non-state actors. The need to analyze civil conflicts at lower escalation levels and understand sub-national level dynamics was driven by an increasing number of intra-state conflicts in the early nineties that did not fit COW's typology of intra-state wars, and the end of the dominating inter-state conflict: the cold-war. While leading COW scholars identified this problem early on (Sarkees et al.), the data project itself was slow to adjust to a changing scientific landscape. Hence, conflict data in the beginning of the 21st century was suddenly dominated by Scandinavians.

The UCDP was properly established at the Department in the mid-1980s under the name Conflict Data Project. It continuously collects data on armed conflicts. The definitions have gradually been refined primarily to fit scholarly requirements of global comparability. The definitions are designed so as to pick up the same phenomenon across time as well as across space. This makes the data useful for systematic studies of the origins of conflict, conflict dynamics and conflict resolution. With the creation of the global conflict database (UCDP Conflict Encyclopedia), this information is now available to anyone interested in the world. However, any user should keep in mind that the ambition of a systematic data collection means that the coding rules are very strict, and that the standards are set very high for inclusion of information. The data, furthermore, is collected on an annual basis, so that information is related to activities during one calendar year. This is so even if conflicts may start and/or end at dates that do not fit this pattern. Still, the emphasis is on the year as the basis for comparison and computation.

The Uppsala conflict data has been published, with respect to major armed conflicts, in the SIPRI Yearbook since 1988. As of 1993, a list of all armed conflicts appears in Journal of Peace Research (JPR). The time series are slightly different. The SIPRI Yearbook now starts at 1990. In JPR, after a joint effort with the International Institute of Peace Research, Oslo (PRIO) on backdating information, there are now data covering the period since 1946. Information on non-state conflicts appears in the Human Security Report in 2004. The global, online conflict database (UCDP Conflict Encyclopedia), which has the most detailed and reliable information available, covers conflicts since 1975 on a large number of variables relevant for conflict analysis. It is the broadest of all forms of publications from the UCDP. This means that there now is a dataset with data from 1946 on a limited number of variables (UCDP/PRIO dataset) and the UCDP online conflict database (UCDP Conflict Encyclopedia) with information from 1975 on a large number of variables and with detailed descriptions of the conflicts included.

2.1.3. *Event Data.* While the country-year and conflict-year format pushed the discipline forward and allowed for a subfield that has gained great recognition within political science, scholars were increasingly skeptical about the ability of highly aggregated to data to capture conflict dynamics. In fact, as theories about conflict became ever more precise in defining the micro-mechanisms of conflict, the data used to test these advancing theories stayed at the macro-level. Even thought there has been a surge of event data in the recent years, a number of early efforts have been made in the seventies and eighties. Edward Azar's COPDAB and Charles McClelland's WEIS can be seen as the front runners in this trend, but many smaller and specialized efforts followed (e.g Leng's BCOW). The NSF-funded Data Development in International Relations project (McGowan, Starr, Merritt, and Zinnes, 1988) stated that event data had become the second most common form of data, behind the dominating non-event COW datasets.

The rise of event data is in the early nineties can easily be interpreted as function of faster and cheaper computing power. Automated event coding suddenly become more feasibly and the access to electronically available news sources increased exponentially. The Kansas Event Data System (KEDS)–now known as the Penn State Event Data Project. used automated coding of English-language news reports to generate political event data focusing on the Middle East, Balkans, and West Africa. These data were used in statistical early warning models to predict political change. In addition, the proprietary VRA-Reader was developed to process increasing volume of international news reports with more precision and extensibility than earlier sparse parsers. As part of these project political event coding taxonomies were developed to deal with actors, actions, and locations associated with individual events. The best known are CAMEO (Gerner, Schrodt and Yilmaz, 2009) and IDEA (Bond, Bond, Oh, Jenkins and Taylor 2003).

However, the real success story of event data began to play out only recently. From the scientific side there was a growing demand to analyze conflict processes on the micro-level and at the same time government sponsors became interested in forecasting of political violence and other political events that demanded disaggregated data. Especially, the integration of Geographical Information Systems and event data changed the way conflict data could be used and analyzed. An early leader in this effort was the ACLED project (Raleigh et al., 2010), but by now most event data collections do disambiguation for geography. In fact, it lead to an explosion of conflict related datasets and an exciting and new way to test dynamic theories across time and space. See for example the recently released data assembled by Kalev Leetaru based on real time web scraping combined with event ontologies based on the Schrodt approach known as "TABARI."

2.2. **Advances in Estimation and forecasting.** Even though Richardson's work clearly has forecasting and prediction of conflict in mind, the first reflection of conflict prediction can be found in Quincy Wright's Study of War. Interestingly, he believes short-term forecasting should be based on public opinions rather than economic and political indices. However, if up-to-data and comprehensive predictions based on indices, "should have a value to for statesmen, similar to that of weather maps for farmers or of business indices for businessmen" (p. 1270). He goes on to say: "Such indices could be used not only for studying the probability of war between particular pairs of states but also for ascertaining the changes in the general tension level within a state of throughout the world" (p. 1271).

In political science, prediction is typically conceptualized as a conditional exercise, in which values on a dependent variable are calculated based on some estimated, or conditional, statistical model, and then compared with the actual observed values (Hildebrand, Laing and Rosenthal, 1976). But a recent tradition makes political predictions about things that have not yet occurred, in the sense that the *Old Farmer's Almanac* predicts the weather for the coming year. In 1978, a volume edited by Nazli Choucri and Thomas Robinson (Choucri and Robinson, 1978) provided an overview of the then current work in forecasting in international relations, much of which was done in the context of policy oriented research for the U.S. government during the Vietnam War. There were a variety of efforts to forecast or evaluate forecasting efforts(Freeman and Job, 1979; Singer and Wallace, 1979; Vincent, 1980) and a few efforts began to forecast internal conflict (Gurr and Lichbach, 1986), but the median empirical article in political science (as well as sociology and economics) used predictions only in the sense of in-sample observational studies.[2] Doran (1999) and others provided some criticism but most scholars avoided making predictions, perhaps because their models had enough difficulty in describing well what *had* happened.

Still there were a few scholars that continued to make predictions (yes, about the future), including Gurr and Harff (1996), Krause (1997), Davies and Gurr (1998), Pevehouse and Goldstein (1999), Schrodt and Gerner (2000), King and Zeng (2001), O'Brien (2002), Bueno de Mesquita (2002), Fearon and Laitin (2003*b*), de Marchi, Gelpi and Grynaviski (2004), Enders and Sandler (2005), Leblang and Satyanath (2006), Ward, Siverson and Cao (2007), Brandt, Colaresi and Freeman (2008), Bennett and Stam (2009), and Gleditsch and Ward (2010),among a few others.[3] However, just in the last years the field of conflict forecasting has tremendously expanded. The surge of prediction research in conflict and peace studies can be attributed to the new availability of spatio-temporal disaggregated data and the application of new estimation strategies. Both developments are a result of increasing computational power that allow access to large data sources and the implementation of complex statistical tools. Including our own, there are a number of political science projects that focus on conflict forecasting. For example, the Predictive Societal Indicators of Radicalism is a forecast model for Domestic Political Violence, created by Amanda Murdie. It forecasts political violence levels 5 years into the future. As mentioned before the Penn State Event Data Project heavily invests in forecasting and there are several smaller efforts such as a research effort led by Håvard Hegre at the Peace Research Institute Oslo. In addition to forecasting the future of armed conflict, the project aims at increasing our understanding of the causes of conflict by utilizing simulations and predictions to forecast the onset and spread conflict and of democracy.

3. Explaining Escalation

Strikingly, most civil conflict predictions are based on structural factors such as political institutions and economic indices. Behavior is largely absent, or rather, behavior is not explicitly modeled in the empirical models. We argue that political institutions matter both in how they condition the behavior of their constituents, and in the institutional capacity to respond to constituent behavior. More precisely, we stress that whether a civil conflict escalates to civil war depends on how the government addresses citizen demands and grievances. If the government has institutional restraints that prevent accommodation or repression of civil unrest, violence is likely to escalate and turn into a civil war.

[2]In the late 1990s, scholars of American electoral politics began making predictions of voting patterns in presidential elections (Campbell, 1992).

[3]A summary of classified efforts was declassified and reported in Feder (1995). A nice overview of some of the historical efforts along with a description of current thinking about forecasting and decision-support is given by O'Brien (2010), a former program manager at DARPA who has conceptualized and supported the ICEWS project under which we conducted some of the research reported in this manuscript.

While escalation is a well studied subject in international conflicts, only few scholars explicitly think of civil war as an escalation of smaller scale violence and civil unrest. This is surprising as internal conflicts are typically classified as civil wars when they reach 1000 battle deaths and, for example, UCDP codes much lower level of violence that only in few cases escalate to the 1000 battle deaths threshold. In fact, not modeling the escalation of conflict given the existence of risk factors, might help to explain the high rate of false positives in our Fearon & Laitin model predictions. When does the scope of violence expand such that it becomes a civil war? We outline a framework that highlights how institutions impact the way the government can respond to demands of the population and increase or decrease the risk of a potent rebel organization that can actually fight at intensity levels that we call civil wars.

3.1. **Governments, People, and Rebels.** Let us assume that the process of conflict escalation involves three actors: the government, the citizens, and a (possibly latent) rebel organization. The process of governance is an ongoing bargain between the government and the citizens. Citizens can make demands towards the government regarding their political or economic status. If the demands cannot be adequately addressed by the government, the citizens ignore this, or plausibly can turn to (violent forms of) protest and even attack the government. This process can be seen as the first step of escalating conflict and signaling the citizens' resolve towards the government. The government can ignore these demands, accommodate them, or repress them.

In the following, we argue that the institutional framework conditions the response of the government and whether it can or wants to repress or accommodate the citizens' demands. If the demands of the citizens are not met, they have the opportunity to support an organized rebel organization that actually has the capacity to further escalate the violence. We can think of the rebels as a latent organization in peace times, emerging only if supported by a sufficient number of unsatisfied citizens who have solved their collective action problem. This support brings the rebel organization into the fore, which is the second step in escalation. Whether this support makes the rebels capable of the level of violence associated with civil war depends on what proportion of the population unsatisfied with the current policies of the government. Political institutions affect the channel by which citizens make demands from the government and the ability of the government to respond in the first place. In a regime that is exclusive, many citizens do not have a political or institutional channel for accomplishing their goals. In an inclusive regime, citizens are more likely to be relatively satisfied with institutional channels as a means to accomplish their goals, and will be less likely to resort to protest, violence, or support of rebels.

If support for rebels is strong, the conflict is much more likely to reach civil war status. Even so, some rebellions do not escalate. The state's institutions affect its repressive apparatus, intelligence-gathering, and fighting capabilities, which curtail the effectiveness of the rebel organization. The state's institutions also affect the level of general support for the rebels among the populace. If the state's institutions are broadly inclusive, except for one small group, then the rebellion may not find support among the general population. This makes it unlikely to escalate into a civil war. Thus escalation from protest to civil war depends first on whether a rebel group forms, and second, on the relative efficacies of the rebel group and the government, and the general popular support for the rebels.

3.2. **Conflictual Behavior and Institutions.** Political institutions affect whether the government is representative, whether is adaptable, responding to citizen demands, and whether it can repress. The representativeness of the government determines whether the citizens have a grievance that must be expressed in protest rather than through more institutionalized political channels. The flexibility of the regime determines whether the government can make civil-war-preventing accommodations. When considering the effect of institutions on conflict, these roles are often collapsed into one. This is, in a sense, a reasonable simplification: a forward-looking citizen will consider the institutional response when deciding how to behave. If institutions and structural factors jointly determine incentives for conflict, they serve as a seemingly-reasonable proxy for behavior in the prediction of civil war. Yet, the fact that using this construct leads to such poor predictions calls it into question.

There are three ways that institutions affect the probability of civil war. Each of these depends on the behavior of citizens and rebel groups. First, institutions affect how responsive the government is to its constituents. This affects the probability citizens will have a grievance. Second, institutions affect the government's response to protest and low-level violence from the citizens. This affects the probability that citizens will continue to be frustrated, potentially turning to rebel groups to accomplish their goals. Third, institutions condition the government's response to rebel violence. This affects the scale of internal conflict, including whether it escalates to civil war status.

This proxy breaks down when we consider how little we know about human conflictual behavior. Conflictual behavior is itself an object of study, particularly in comparative politics. Unfortunately, for some motivations for conflictual behavior, data on the motivation is harder to collect than data on the behavior itself. Incorporating behavior is a way to account for the motivations, constraints, and facilitating structures that we have missed. One benefit is that it allows us to make better predictions about the outbreak of civil war. Our goal, however, is not only to predict civil war, but to understand its causes. Incorporating behavior into our theory helps. It divides civil war prediction into two parts: a) whether citizens are expressing grievance at all, and b) whether this conflict will escalate to civil war. We then focus on the latter, taking behavior as partially exogenous.

3.3. **Bringing behavior back in.** The addition of behavior to theories that predict conflict is important for two reasons. First, behavior is implicit in the institutional and structural variables in typical conflict prediction models. Theories of conflict are theories of behavior. However, structural and institutional variables do not always behave the way that we expect. The Laitin & Fearon Random Narratives project,[7] for instance, conducts a series of analyses of individual false predictions their model makes. They conclude, in the case of Burkina Faso, that the behavioral mechanisms they had proposed do take place, but conflict did not escalate. They take this as a validation of their theory, although the prediction was incorrect. We contend that the difference between discontent and civil war is substantially larger than the effort political scientists have dedicated to explaining it. While individual explorations of the behavioral elements of conflict are valuable for theory-building, an explicit theoretical integration of behavior makes it possible to trace the causal channels for civil war–from discontent through escalation–in a systematic way.

Second, incorporating behavior can allow us to be more specific about the contexts in which discontent leads to civil war. Behavior complicates civil war, both in theory and in practice. However, we believe that it complicates the theory in predictable ways. Rather than simply claiming that, for instance, protest signifies a heightened possibility of rebellion, these event data allow us to understand which interactions are likely to precipitate which kind of crisis. Including proximal causes does not interfere with our ability to incorporate structural causes in the model. That is, event data permits us to test theories on the social processes set in motion by structural conditions, allowing us to create a picture at once more detailed and more integrated than before.

4. Empirics: New Civil War Model

In this section we present a predictive model of conflict that incorporates behavioral and institutional variables to predict the escalation of conflict.[8] Our methodological approach is different from the standard estimators used in the civil conflict literature. Most commonly scholars estimate logit regression models assuming that all countries have the same baseline risk and are affected similarly by a set of covariates. In contrast, we estimate hierarchical models with random intercepts, a type of mixed effects model. They have the ability to provide a general framework for understanding a phenomenon, without requiring that the coefficients be exactly the same for each and every case. This makes consummate sense in a world in which there are lots of unmodeled aspects and in which there is a heterogeneity to the objects studied. These kinds of models have facets that operate with groupings, typically at different levels, such as the level of a country, but maybe also at a level of democracy as well. Hierarchical models also keep track of the variation between the groupings. This approach allows us a) to learn about processes that may vary slightly from one place or time to another, b) use all the data while compromising between within-group estimates that are highly uncertain because they are based on averages, and the more precise individual estimates that plausibly ignore influences that occur at the level of a group, as well as c) keep track of the uncertainty and co-variation across the different levels. As an example, it may well be that accumulated inequalities tend to be associated with rebellious onsets in a fairly predictable way, but that this relationship is perhaps slightly different for autocracies than it is in democracies. One simple way to model this is with an interaction term, but that ignores the variation that may occur among the groupings (dictators may get a lot of foreign aid) and the individual effects within each country.[4] We could model civil war separately in each country, and then average over all the coefficients we obtained. Hierarchical models are a compromise between this pooled approach and one in which there are groups of countries that are modeled together.

[4]See Gelman and Hill (2007); Pinheiro and Bates (2009) for a more complete statement of the benefits of this approach.

In the most general case, we can model civil war (and other crisis events) using hierarchical models in which both the intercept and slope vary. Simply stated, this means that we group the data along an indicator, such as level of executive constraints, creating a different intercept for each group. Thus, the varying intercepts correspond to group indicators and the varying slopes represent an interaction between predictor variables x and the group indicators:

$$\begin{aligned} \Pr(y_{it} = 1) &= \text{logit}^{-1}(\alpha_{j[it]} + \beta_{j[it]} x_{it}) \\ \begin{pmatrix} \alpha_j \\ \beta_j \end{pmatrix} &\sim N\left(\begin{matrix} \mu_\alpha \\ \mu_\beta \end{matrix}, \Sigma \right) \end{aligned}$$

where i denotes the countries, t the month and j the grouping variable, α_j are the grouping variable's random intercepts. x^G_{it} and x^O_{it} are predictor variables; β^G_j and β^O_j are the associated random coefficients; γ is a vector of fixed effects associated with Z_{it}.

However, in this paper we only allow for varying intercepts for each country. In addition to dealing with country-specific effects, we model civil conflict as a function of lagged event data that are gleaned from natural language processing of a continuously updated harvest of news stories, primarily taken from Factiva, an open source, proprietary repository of news stories from over 200 sources around the world. The baseline event coder is called JABARI, a java variant of TABARI (Text Analysis By Augmented Replacement Instructions), the former developed by Philip Schrodt and colleagues.[9] From this event data we construct explanatory variables that indicate whether high intensity events (e.g. protests, fighting, killings) or low intensity events (e.g. demands or threats) are taking place between the government and opposition groups.

These data are augmented with a variety of other attribute and network data. We use country attributes, coded on a monthly or yearly basis from the Polity (Democ and Autoc), World Bank (GDP per capita), and Excluded Population (Cederman, Buhaug and Rød (2009)) databases. In addition, we use information about relations among the countries, including geography to access the extent of conflictual events in the surrounding countries. We use this data to predict occurrences of civil conflict based on UCDP data. The results from our empirical model, measured at the monthly level from 1997-2011 are given in Table 4.

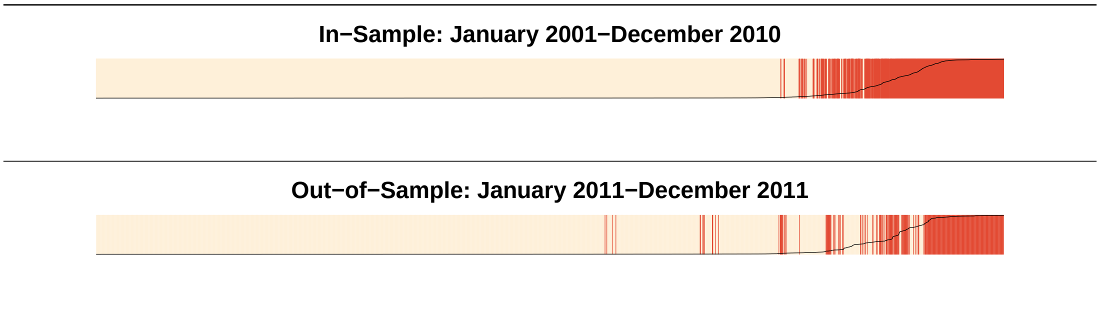
Table 4. Hierarchical model estimates are in line with our substantive expectations and are precisely estimated. The fit of the model, described below, is substantial.

	$\hat{\beta}$	$\sigma_{\hat{\beta}}$	Z-score
(Intercept)	0.48	1.09	0.44
High intensity conflictual events$_{t-1}$	1.94	0.13	14.45
Low intensity conflictual events$_{t-1}$	1.01	0.19	5.43
Excluded population$_{t-1}$	7.01	1.12	6.27
Excluded population$^2_{t-1}$	-7.79	1.40	-5.57
log GDP/capita$_{t-1}$	-0.62	0.15	-4.11
Democ$_{t-1}$	-0.26	0.05	-5.13
Autoc$_{t-1}$	-0.12	0.06	-2.14
Spatial low intensity conflictual events$_{t-1}$	-0.68	0.46	-1.48

Both high and low intensity conflictual events are associated with higher probabilities of civil war, but high intensity events are about twice as powerful on average as those classified as low intensity. At the same time civil wars are more likely to exist in societies which have large slices of the population that are excluded from the political access to the state. However, this relationship is non-linear which is line with previous findings in the literature. Also in line with existing research, richer countries are less likely to have civil wars and both very democratic and very autocratic countries have a reduced risk of civil war. Interestingly, conflict seems to follow a checkerboard pattern wherein regional conflicts, tend to suppress civil wars in neighboring countries.

Figure 1 illustrates the fit of the model both in and out of sample using separation plots. These plots provide a summary of the fit for each model by demonstrating the range and degree of variation among the predicted

Figure 1. Separation plots for CRISP Model Prediction of UCDP Data



probabilities and the degree to which predicted probabilities correspond to actual instances of the event. They are created by ordering all country-years according to their predicted probability of civil war onset, from lowest on the left to highest on the right. The black line through the center of the plot represents this probability. Countries where a civil war actually occurred are red, while those where no civil war occurred are white (Greenhill, Ward and Sacks, 2011). Red lines on the left show false negatives, while white events on the right show false positives. A good fit would be visualized with more red panels (events occurring) stacked at the right end of the plot. These plots show that: 1) the actual civil wars are among those cases with higher predicted probabilities and 2) as expected, the out-of-sample fit of the model is slightly worse than the in-sample fit. However, even out-of-sample, the model fits the data very well.

Table 5 makes these comparisons more standard way by reporting the standard performance statistics for fit in binary models (assuming a classification cutoff of 0.5). Again, the out-of-sample performance is a bit worse than in-sample, on all of these single number estimates, but the out-of-sample evaluation fair very well, except that it over-predicts the actual number of civil wars. The in-sample performance is pretty much exactly on-target.

Table 5. Performance statistics, in-sample and out-of-sample for civil war model

Statistic	In-Sample	Out-of-Sample
Sensitivity	0.83	0.70
Specificity	0.98	0.98
Accuracy	0.97	0.94
Precision	0.88	0.86
Brier	0.03	0.04
AUC	0.99	0.97
N. Civil Wars	2468	286
Predicted N. Civil Wars	2461	255

In line with our expectations, Table 6 demonstrates that as the classification threshold is reduced, the proportion of correctly predicted civil wars goes *up*, but the number of false positives also increases.

Table 7 lists the five actual civil war onsets during the out-of-sample period, the month during which the conflict is deemed to have begun, and the predicted probability as given by the model for that month, as well as for the immediate prior and subsequent months. Arguably, Côte d'Ivoire and Senegal rise to probability levels that are heuristic, but the three other cases (Nigeria, Syria, and Libya) do not. This suggests that the model itself is not performing as well as suggested by the in-sample results, let alone the statistical results.

Finally, we extend our cross-validation strategy by making forecasts for data that were neither used in the in-sample nor out-of-sample predictions. This reduces the dependency between the modeling and forecasting process. As outlined in the beginning of this paper we argue that such practices will improve the discipline's efforts to

Table 6. Number of correctly predicted conflicts and false positives for CRISP models on UCDP data

Threshold	**Correctly Predicted**	**False Positives**
0.50	199/286	33
0.30	235/286	62
0.10	249/286	131
0.05	261/286	154

Table 7. Predicted Probability for Out-of-Sample Onsets in UCDP Data

	Start of Civil War	$\hat{p}$ of CW	Prior Month	Next Month
Côte d'Ivoire	March, 2012	0.271	0.257	0.286
Senegal	December, 2011	0.260	0.725	0.245
Nigeria	March, 2011	0.115	0.111	0.111
Syria	October, 2011	0.005	0.005	0.005
Libya	March, 2011	0.004	0.004	0.004

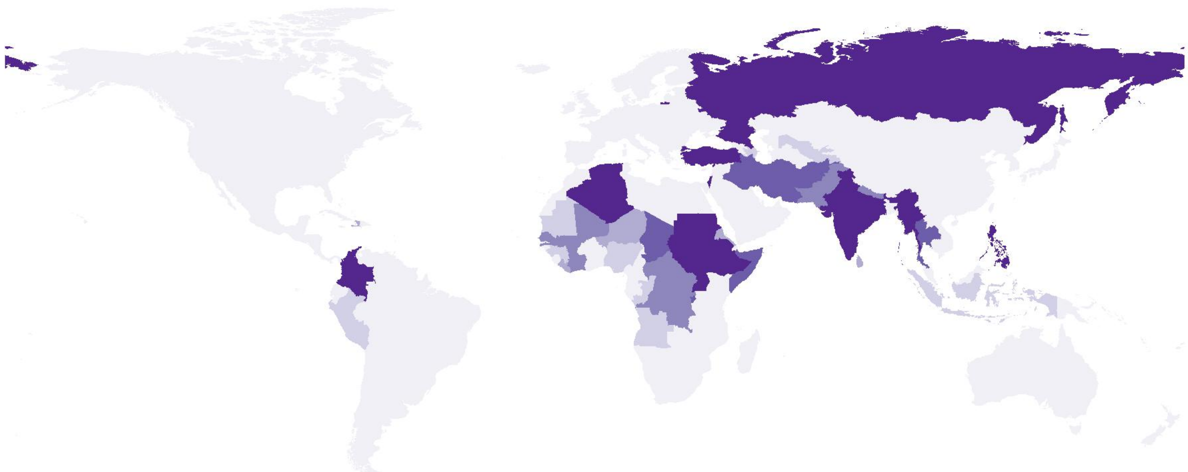
explain and predict violent conflicts. In our case, we make predictions for the first six months of 2013 (the future at the time of writing) based on our model estimates and available data.

Table 8 provides the ten countries which have the highest predicted probabilities during the 2012 (out-of-sample) period. The country with the highest probability is Sudan, which is indeed experiencing a rebellion. Similarly, Ethiopia is in the throes of an Islamic rebellion. Similarly, Algeria is experiencing a rebellion in the Tuareg region, related in large part to difficulties in neighboring Mali. The rebellion in the Philippines is ongoing, as are rebellions in Columbia and Uganda. Russia has a rebellion in parts of the federation as does India. Turkey and Israel are also experiencing substantial domestic conflicts even though they entail very large transnational components. Indeed all of these top ten countries are shown to have civil wars according to the UCDP, as of their latest data. The bottom of Table 8 shows the five *new* civil wars, according to UCDP, along with the attendant probabilities which are assigned by our statistical model. As evident, the model is reasonably able to identify onsets as well as ongoing civil wars, though in our CRISP project we use slightly different models for onsets.[10]

Table 8. Predictions for 2012: 10 Countries with Highest Average Probability of Continuing Conflict and 5 With Highest Average Probability of New Conflict

	01/2012	02/2012	03/2012	04/2012	05/2012	06/2012	07/2012	08/2012	09/2012	**Forecast** 10/2012 - 03/2013
Sudan	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99
India	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.94	0.99
Ethiopia	0.98	0.99	0.99	0.99	0.97	0.96	0.99	0.99	0.97	0.99
Russia	0.99	0.96	0.96	0.99	0.99	0.96	0.99	0.99	0.96	0.99
Israel	0.98	0.98	0.98	0.97	0.98	0.98	0.98	0.98	0.98	0.99
Turkey	0.99	0.97	0.99	0.99	0.99	0.99	0.99	0.99	0.92	0.99
Algeria	0.99	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.99
Philippines	0.98	0.99	0.99	0.98	0.98	0.98	0.98	0.98	0.89	0.99
Colombia	0.99	0.97	0.97	0.97	0.98	0.97	0.98	0.99	0.88	0.99
Uganda	0.93	0.95	0.95	0.99	0.99	0.99	0.95	0.94	0.96	0.99
Chad	0.67	0.64	0.64	0.60	0.66	0.66	0.67	0.67	0.67	0.99
DR Congo	0.29	0.33	0.37	0.33	0.35	0.38	0.36	0.35	0.41	0.91
Nepal	0.30	0.33	0.32	0.32	0.33	0.33	0.34	0.34	0.37	0.90
Mali	0.27	0.26	0.71	0.86	0.89	0.24	0.28	0.28	0.27	0.83
Sri Lanka	0.16	0.14	0.13	0.51	0.14	0.56	0.55	0.54	0.16	0.64

Figure 2. Map for CRISP Model Predictions of Civil War (October 2012-March 2013), with darker colors presenting higher probabilities.



These countries shown in Table 8 are also the countries with the highest probability of civil war during the six-months for which are truly in the future at the time of writing, with the last column represented the chance of civil war prior to March 2013. On average, we expect to see 168 country-months of civil wars over the next six months. Will all of the attendant civil wars occur or persist in the next six months? If they do, it will be further evidence that our model is on the right track, and give some additional evidence that it identifies important features that help to explain civil wars. If they don't, it becomes evident that the model is flawed and less reliable than the statistical evidence suggested. Where it does not play out as predicted, there will be information about things that we ignored or inadequately captured, measured poorly, or didn't understand. In particular, this will provide heuristic information that in principle should help us to create better understanding as well as better models of civil war onsets. This doesn't mean, however, that we should expect probabilistic models such as that developed here to be deterministically accurate. But the more accurate they are, the higher our confidence in the effort. We show in Figure 2 a heat-map illustrating where in the world we probabilistically expect to find civil wars. This may seem unsurprising to some, but it is nonetheless developed by using theoretically-based statistical models that have been annealed with cross-validation and subject to actually being wrong. For that alone, it represents a step toward better understanding.

5. Conclusion

In this article we have demonstrated the utility of creating forecasting models for predicting political conflicts in a diverse range of country settings. We have shown that statistical models of civil war can be highly accurate, generating few false negative and positive predictions, both in and out-of-sample. These models can shed light on an array of important components of the political science literature on conflict dynamics.

One frequent, and quickly surfaced, criticism of predictions in the social sciences is that social phenomena such as international crises are simply too complicated to predict by any means. Precisely because political conflicts are quite complicated we should expand reasoning into mechanisms that can support the complications. Indeed, complex systems involve a wide variety of mechanisms and phenomena that are not easily described, let alone understood in isolation. A good example is meteorology, wherein we each receive a variety of forecasts every day. These forecasts are typically generated by combining a large number of forecasts that are based on meteorological models of weather that are based on the physics and chemistry of what is governing the various interacting systems. These systems each use a vast amount of measured data on the stocks and flows of various physical characteristics. These systems permit heterogeneity, so the predictions are not the same everywhere. Also, they permit an increasingly accurate scale of prediction.

Indeed, the first attempt at weather prediction comes from Richardson over a century ago, when he used his mathematical approach to predict (retrospectively) the weather for 20 May 1910 by hand, using data from that date to predict the weather six hours subsequently. When corrected by modern smoothing techniques, Richardson's predictions were quite accurate, although he did not perceive them to be adequate at the time (Lynch, 2006). To Richardson a global model of weather forecasts would have taken tens of thousands of human calculators, which from his perspective seemed impossible. However, today with the available computational power and exact global data, most of the global weather models are based on the Richardson's equations.

Interestingly, Richardson turned away from weather predictions and wrote one of the first books on the statistical analysis of war: *Statistics of Deadly Quarrels*. After his experiences in the First World War, he thus focused on another complicated phenomenon to predict. Given advancements in theory, data collection, statistics, and computational power, we might be at an important point to push the boundaries of predicting political phenomenon beyond what we believed was possible only a few years ago. To preemptively declare defeat at the forecasting task seems foolish.

Finally, though some see the main benefit of prediction as creating a kind of social radar,[5] the real benefit of prediction may actually be as a heuristic allowing further probing of the empirical validity of specific models. Political science–especially where samples and experiments are not feasible–has an enormous vulnerability to over-reliance on the available data. In a statistical sense this is often seen as over-fitting: we use all the data to generate models that are dependent on all the data. That no longer seems like a very good research design. Being able to use our models to describe data we haven't seen before should be one gold standard criterion for model evaluation. In the face of a torrent of new data about the world, we can do this in almost real time. This permits the possibility of generating predictions about the future that may be useful, not just toward validating our theories.

The sabermetric revolution in sports portrayed in the movie *Moneyball* pitted geeky statistical modelers against wizened scouts. Some will see the same dynamics between those with "Big Data" and "Big Algorithms" and the subject matter experts with even more detailed knowledge of single cases. Tetlock has shown that some subject matter experts are especially bad (Tetlock, 2005). The same is true of some models, but without using prediction as a heuristic, we won't necessarily find this out. But even among those who are convinced the old scouts are the best scouts, there is renewed attention to how to evaluate political judgement such that it uses more data, keeps better track of mistakes, and integrates multiple estimates of outcomes.[6] Many others have begun researching ways to bring together better ideas in the generation of forecasts, better ideas that incorporate accountability and multiple methodologies for making predictions Bueno de Mesquita 2010; Tetlock 2010.

Notes

[1]The replication data, plus myriad reanalyses, are available at `http://www.stanford.edu/~jfearon/`.

[2]Model (1) uses the following independent variables: prior war, lagged per capita income, lagged log of the population, log of the percentage of mountainous terrain, noncontiguous state, oil exporter, new state, lagged instability measure, lagged Polity IV score, ethnic fractionalization, religious fractionalization. The lags are one year.

[3]Probabilities in year N are given by $P(\text{onset in year } N) = 1 - (1 - P(\text{onset in 1999}))^{N-1999}$. If a war occurs in year M, between 1999 and year N, the exponent is set to $(N - M)$. We had to exclude a number of countries since they did not have a predicted probability for 1999 due to missing data. This is the approach which results in the probability map found in Fearon & Laitin.

[4]The onset and prior war variables were taken from UCDP's intrastate war dataset.

[5]Using the out-of-sample approach with a threshold of 0.05.

[6] Some of these are in the public domain and include recent efforts. A recent report by the Army Environmental Policy Institute (2011) lists no fewer that twelve ongoing projects that touch on some aspect of forecasting in the environmental realm (somewhat loosely conceptualized).

[7]Laitin and Fearon 2009

[8]More information about our modeling approach can be found at `mdwardlab.com` which describes the Crisis Prediction (CRISP) modeling activities and provides links to a variety of forecasting activities.

[5]There is, among the policy community, a great optimism that these kinds of models will provide a unified way of observing as well as predicting social behavior; see Maybury 2010.

[6]See the recent articles published by the National Academy on this topic Fischhoff and Chauvin 2011.

[9]See `http://eventdata.psu.edu/`.
[10]We use split-population, duration models.

References

Azar, Edward. 1980. "The Conflict and Peace Data Bank (COPDAB) Project." *Journal of Conflict Resolution* 24:143–152.

Bennett, D. Scott and Allan C..C. Stam. 2009. "Revisiting Predictions of War Duration." *Conflict Management and Peace Science* 26(3):256.

Bond, Doug, Joe Bond, Churl Oh, J. Craig Jenkins and Charles L. Taylor. 2003. "Integrated Data for Events Analysis (IDEA): An Event Typology for Automated Events Data Development." *Journal of Peace Research* 40(6):733–745.

Brandt, Patrick T..T., Michael. Colaresi and John R..R. Freeman. 2008. "The Dynamics of Reciprocity, Accountability, and Credibility." *Journal of Conflict Resolution* 52(3):343.

Bremer, Stuart A. and Thomas R. Cusack. 1995. *The Process of War: Advancing the Scientific Study of War.* Luxembourg and Philadelphia, PA: Gordon and Breach.

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

Bueno de Mesquita, Bruce. 2002. *Predicting Politics.* Columbus, Ohio: Ohio State University Press.

Bueno de Mesquita, Bruce. 2010. "Judging Judgment." *Critical Review* 22(4):355–388.

Campbell, James E. 1992. "Forecasting the Presidential Vote in the States." *American Journal of Political Science* 36(2):386–407.

Cederman, Lars, Halvard Buhaug and Jan Ketil Rød. 2009. "Ethno-nationalist dyads and civil war: A GIS-based analysis." *Journal of Conflict Resolution* 53(4):496–525.

Choucri, Nazli and Thomas W. Robinson, eds. 1978. *Forecasting in International Relations: Theory, methods, problems, prospects.* W.H. Freeman.

Davies, John L. and Ted Robert Gurr. 1998. *Preventive measures: building risk assessment and crisis early warning systems.* Rowman & Littlefield Publishers.

de Marchi, Scott, Christopher Gelpi and Jeffrey D. Grynaviski. 2004. "Untangling Neural Nets." *American Political Science Review* 98(02):371–378.

Doran, Charles F. 1999. "Why Forecasts Fail: The Limits and Potential of Forecasting in International Relations and Economics." *International Studies Review* 1(2):11–41.

Enders, Walter and Todd Sandler. 2005. "After 9/11: Is it All Different Now?" *The Journal of Conflict Resolution* 49(2):259–277.

Fearon, James D. and David D. Laitin. 2003*a*. "Ethnicity, Insurgency, and Civil War." *The American Political Science Review* 97(1):75–90.

Fearon, James D. and David D. Laitin. 2003*b*. "Ethnicity, Insurgency and Civil War." *American Political Science Review* 97(1):75–90.

Feder, Stanley. 1995. Factions and Policon: New Ways to Analyze Politics. In *Inside CIA's Private World: Declassified Articles from The Agency's Internal Journal, 1955-1992*, ed. H. Bradford Westerfield. New Haven: Yale University Press pp. 274–292.

Fischhoff, Baruch and Cherie Chauvin, eds. 2011. *Intelligence Analysis: Behavioral and Social Scientific Foundations.* Washington, D.C.: National Academies Press.

Freeman, John R. and Brian L. Job. 1979. "Scientific Forecasts in International Relations: Problems of Definition and Epistemology." *International Studies Quarterly* 23(1):113–143.

Gelman, Andrew and Jennifer Hill. 2007. *Data Analysis Using Regression and Multilevel/Hierarchical Models.* Analytical Methods for Social Research Cambridge, UK and NYC: Cambridge.

Gerner, Deborah J., Philip A. Schrodt and Ömür Yilmaz. 2009. "Conflict and Mediation Event Observations (CAMEO) Codebook." *Available at http:// eventdata.psu.edu/data.dir/cameo.html* .

Gleditsch, Kristian Skrede and Michael D. Ward. 2010. "Contentious Issues and Forecasting Interstate Disputes." Presented to the 2010 Annual Meeting of the International Studies Association.

Gleditsch, Nils Petter, Peter Wallensteen, Mikael Eriksson, Margareta Sollenberg and Håvard Strand. 2002. "Armed Conflict 1946-2001: A New Dataset." *Journal of Peace Research* 39(5):615–637.

Gneiting, Tilmann and Adrian E. Raftery. 2007. "Strictly Proper Scoring Rules, Prediction, and Estimation." *Journal of the American Statistical Association* 102(477):359–378.

Greenhill, Brian D., Michael D. Ward and Audrey Sacks. 2011. "The Separation Plot: A New Visual Method for Evaluating the Fit of Binary Models." *American Journal of Political Science* 55(4):991–1003.

Gurr, Ted Robert and Barbara Harff. 1996. "Early warning of communal conflict and humanitarian crisis." *Tokyo: United Nations University Press, Monograph Series on Governance and Conflict Resolution* .

Gurr, Ted Robert and Mark I. Lichbach. 1986. "Forecasting Internal Conflict: A Competitive Evaluation of Empirical Theories." *Comparative Political Studies* 19(1):3–38.

Hegre, Håvard, Håvard Nygård, Håvard Strand, Henrik Urdal and Joakim Karlsen. 2013 in press. "Predicting Armed Conflict, 2011-2050." *International Studies Quarterly* tbd(tbd):tbd.

Hildebrand, David K., James D. Laing and Howard Rosenthal. 1976. "Prediction Analysis in Political Research." *The American Political Science Review* 70(2):509–535.

Institute, Army Environmental Policy. 2011. "Environmental Change and Fragile States: Early Warning Needs, Opportunities, & Intervention." Fort Belvoir, VA: AEPI Report.

King, Gary and Langche Zeng. 2001. "Improving Forecasts of State Failure." *World Politics* 53(4):623–658.

Krause, George A. 1997. "Voters, Information Heterogeneity, and the Dynamics of Aggregate Economic Expectations." *American Journal of Political Science* 41(4):1170–1200.

Laitin, David D. and James D. Fearon. 2009. *Oxford Handbook of Political Science.* New York, N.Y.: Oxford University Press chapter Integrating Qualitative and Quantitative Methods, pp. 1166–1186.

Leblang, David and Shanker Satyanath. 2006. "Institutions, Expectations, and Currency Crises." *International Organization* 60(01):245–262.

Leng, Russell J. 1993. *Interstate crisis behavior, 1816-1980: realism versus reciprocity.* Cambridge and New York: Cambridge University Press.

Lynch, Peter. 2006. *The Emergence of Numerical Weather Prediction: Richardson's Dream.* Cambridge: Cambridge University Press.

Maybury, Mark. 2010. Social Radar for Smart Power. Technical Papers 10-0745 Mitre Corporation Bedford, MA: .

McClelland, Charles A. and Gary Hoggard. 1969. Conflict Patterns in the Interactions among Nations. In *International Politics and Foreign Policy*, ed. James N. Rosenau. New York: The Free Press pp. 711–724.

McGowan, Pat, Harvey Starr, Gretchen Hower, Richard L. Merritt and Dina A. Zinnes. 1988. *International Data as a National Resource.* Vol. 14 of *International Interactions: Special Issue* New York: Taylor and Francis.

Milstein, Jeffery S. 1974. *Dynamics of the Vietnam War.* Columbus, Ohio: Ohio State University Press.

Milstein, Jeffrey S. and William C. Mitchell. 1968. "Dynamics of the Vietnam Conflict." *Peace Research Society (International) Papers* 10:163–NNN.

Murdie, Amanda. 2011. Coercion, Capacity, and Coordination: A Risk Assessment Model of the Determinants of Political Violence. Technical report Council on Foreign Relations Washington D.C.: .

O'Brien, Sean P. 2002. "Anticipating the Good, the Bad, and the Ugly: An Early Warning Approach to Conflict and Instability Analysis." *The Journal of Conflict Resolution* 46(6):791–811.

O'Brien, Sean P. 2010. "Crisis Early Warning and Decision Support: Contemporary Approaches and Thoughts on Future Research." *International Studies Review* 12(1):87–104.

Pevehouse, Jon C. and Joshua S. Goldstein. 1999. "Serbian compliance or defiance in Kosovo? Statistical analysis and real-time predictions." *The Journal of Conflict Resolution* 43(4):538–546.

Pinheiro, J.C. and D.M. Bates. 2009. *Mixed-Effects Models in S and S-Plus.* Statistics and Computing New York: Springer.

Raleigh, Clionadh, Andrew Linke, Håvard Hegre and Joakim Karlsen. 2010. "Introducing ACLED: An Armed Conflict Location and Event Dataset: Special Data Feature." *Journal of Peace Research* 47(5):651–660.

Richardson, Lewis F. 1960. *Statistics of Deadly Quarrels.* Chicago and Pittsburgh, PA: Quadrangle/Boxwood.

Richardson, Lewis Frye. 1922. *Weather Prediction by Numerical Process.* Cambridge: Cambridge University Press.

Sarkees, Meredith Raid, Frank Whelon Wayman and David J. Singer. 2003. "Inter-State, Intra-State, and Extra-State Wars: A Comprehensive Look at Their Distribution Over Time, 1816-1997." *International Studies Quarterly* 47(1):47–70.

Schrodt, Philip A. and Deborah J. Gerner. 2000. "Using Cluster Analysis to Derive Early Warning Indicators for Political Change in the Middle East, 1979-1996." *American Political Science Review* 94(4):803–818.

Singer, J. David and Michael D. Wallace, eds. 1979. *To Augur Well: Early Warning Indicators in World Politics.* Beverly Hills, CA: SAGE.

Sorokin, Pitirim A. 1957. *Social and Cultural Dynamics.* London: Owen.

Stevens, Jacqueline. 2012. "Political Scientists Are Lousy Forecasters." *New York Times* June 23.

Tetlock, Philip. 2005. *Expert Political Judgment: How Good is it? How Can we Know?* Princeton, NJ: Princeton University Press.

Tetlock, Phillip E. 2010. "Second Thoughts about *Expert Political Judgment*: Reply to the Symposium." *Critical Review* 22(4):467–488.

Vincent, Jack E. 1980. "Scientific Prediction versus Crystal Ball Gazing: Can the Unknown be Known?" *International Studies Quarterly* 24(3):450–454.

Ward, Michael D., Randolph M. Siverson and Xun Cao. 2007. "Disputes, Democracies, and Dependencies: A Reexamination of the Kantian Peace." *American Journal of Political Science* 51(3):583–601.

Wright, Quincy. 1942. *A Study of War.* Reprinted 1957 ed. Chicago: University of Chicago Press.

Michael D. Ward: Department of Political Science
Current address: Duke University, Durham, NC, 27708, USA
E-mail address: michael.d.ward@duke.edu

Nils W. Metternich: Department of Political Science
Current address: Duke University, Durham, NC, 27708, USA
E-mail address: nm86@duke.edu

Cassy Dorff: Department of Political Science
Current address: Duke University, Durham, NC, 27708, USA
E-mail address: cassy.dorff@duke.edu

Max Gallop: Department of Political Science
Current address: Duke University, Durham, NC, 27708, USA
E-mail address: max.gallop@duke.edu

Florian Hollenbach: Department of Political Science
Current address: Duke University, Durham, NC, 27708, USA
E-mail address: florian.hollenbach@duke.edu

Anna Schultz Department of Political Science
Current address: Duke University, Durham, NC, 27708, USA
E-mail address: anna.schultz@duke.edu

Simon Weschle: Department of Political Science
Current address: Duke University, Durham, NC, 27708, USA
E-mail address: simon.weschle@duke.edu