Forecasting Civil Wars: A Pre-Analysis Plan*

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

1 Introduction

We use event data from the World-Wide Integrated Crisis Early Warning System (ICEWS) to forecast the onset of civil war over 1-month and 6-month windows. While attempts to forecast civil wars are not new, most models rely on what we might call a country's "fundamentals"—structural characteristics such as per capita GDP, natural resources, regime type, or ethnolinguistic fractionalization. Some of these variables have been shown to be statistically significant correlates of civil war, but most are poor predictors (Ward, Greenhill and Bakke 2010).

Moreover, most attempts to forecast civil war are only ever evaluated retrospectively—i.e. after the researcher has already observed the "true" distribution of civil wars against which the forecasts are being compared. Further, the specific model parameters used to generate those forecasts are rarely (if ever) defined *ex ante* in something akin to a pre-analysis plan. As a result, it is often difficult to evaluate the prospective performance of even the most prominent models (e.g. those produced by the Political Instability Task Force), and the possibility of "fishing" for results—ensuring accuracy by over-fitting through multiple unreported rounds of training and testing on the same dataset—looms large.

Our intention here is not to accuse other researchers of wrongdoing. Rather, it is to observe that the same concerns about fishing that have catalyzed the use of pre-analysis plans in experimental and quasi-experimental research are very much applicable to forecasting as well. While some predictions are publicized in the media (e.g. Nate Silver's famous forecasts of U.S. presidential election results), and others are posted on researchers' personal or institutional websites (e.g. the website of the Early Warning Project[1]), we are unaware of any previous effort to pre-register predictions and the specific model parameters used to generate them in a formal, publicly-available scholarly venue. Nor are we aware of any

[1] See `http://www.earlywarningproject.com/risk_assessments`, accessed February 2, 2016.

previous effort to pre-register predictions in a way that cannot be subsequently revised without documentation and approval by a third party.

2 Data

2.1 Predictors

We use ICEWS event data to code predictors for civil war onset at the country-month level, with actors classified into one of three categories: government,[2] rebels[3] or opposition.[4] We limit our analysis to events that occur between two actors within the same country, and to the period beginning January 1, 2001[5] and ending December 31, 2015. For the period through 2013 we use the public version of ICEWS posted on March 18th, 2015. For 2014 and 2015 we use an embargoed version of the dataset made available to us through the PITF.[6]

We assess the relative performance of five models:

1. The *escalation* model, composed of country-month event counts for (1) low-level violence, (2) non-violent repression, (3) demands and (4) accommodations. Low-level violence includes all events in which (1) government is the source and opposition is the target; (2) opposition is the source and government is the target; (3) government

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

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

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

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

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

is the source and rebels are the target; or (4) rebels are the source and government is the target. Non-violent repression and accommodations include only events in which government is the source and either opposition or rebels is the target, and demands includes only events in which either opposition or rebels is the source and government is the target. See the appendix for the CAMEO codes comprising each of these event counts.

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

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

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

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

For the *Quad*, *Goldstein* and *CAMEO* models we include all events in which (1) government is the source and opposition is the target; (2) opposition is the source and government is the target; (3) government is the source and rebels are the target; or (4) rebels are the source and the government is the target.

The *Quad*, *Goldstein* and *CAMEO* models incorporate all event types in the ICEWS dataset (albeit in different ways), and thus constitute more atheoretical benchmarks against which to assess the *escalation* model, which, unlike the other three, is motivated by theoretical priors about escalation dynamics. (See the accompanying working paper for further details.) We include the *average* model as well because ensembles of predictions gleaned from multiple models often outperform those of a single model alone (Montgomery, Hollenbach and Ward 2012).

2.2 Dependent variable

The dependent variable throughout is civil war onset, coded 1 for each country-month in which a civil war began, 0 for each country-month of peace, and missing for each country-month of ongoing conflict (excluding the month of onset). These data are gleaned from the Sambanis (2004) dataset on civil war onsets, updated through the end of 2015.

2.3 Estimation

We generate forecasts using random forests, an ensemble method for regression and classification that is now standard in the machine learning literature (Breiman 2001; Hastie, Tibshirani and Friedman 2001). We run random forests with five terminal nodes and 100 observations per tree, sampled without replacement, and with 100,000 trees per forest. We train each model on a subset of the data beginning January 1, 2001 and ending December 31, 2007, forecasting over intervals of either one or six months. We test the models on the remaining subset of the data. We implement our analysis using the randomForest package in R.

2.4 Model performance

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

Table 1: Out-of-sample AUCs

1-month forecasts

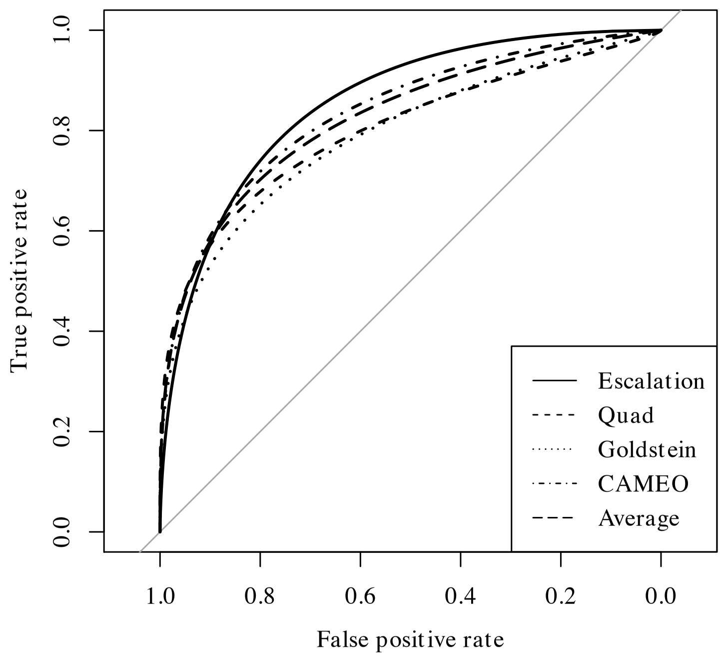
Escalation	Quad	Goldstein	CAMEO	Avg.
0.85	0.80	0.79	0.84	0.82

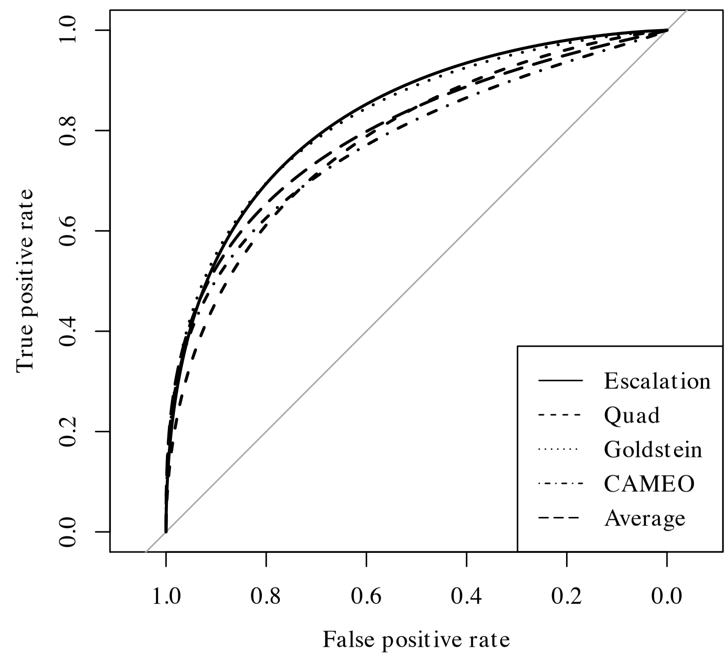
6-month forecasts

Escalation	Quad	Goldstein	CAMEO	Avg.
0.83	0.78	0.83	0.77	0.79

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

Figure 1: Out-of-sample ROC curves





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

Table 2: 2016 prediction rankings for 6-month *escalation* model

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

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

Figure 1 presents the ROC curves themselves, smoothed for ease of interpretation. We will update Table 1 and Figure 1 every six months, beginning in July 2016.

2.5 Forecasts for first half of 2016

Finally, Table 2 provides "true" prospective forecasts for the countries at highest risk of civil war in the first half of 2016, using predictors measured through December 31, 2015. For compactness and ease of exposition we report results for the top 30 highest-risk countries only, based on cross-country rankings generated over 6-month windows using the *escalation* model alone. Below the top 30 the predicted probabilities are all so similar, and so small, that we interpret them as indistinguishable. We will update this table and assess the retrospective performance of our model every six months. To do this we will (i) compare our predictions to the "true" distribution of civil wars around the world at the end of each 6-month period and calculate the resulting AUC[7], and (ii) more qualitatively evaluate the extent to which our predictions match patterns of instability in the highest-risk countries during each 6-month period.

References

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

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

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

[7]We will do this in two ways: (i) by coding all new or ongoing civil wars as 1s, regardless of their month of onset, and (ii) by coding as 1s only those civil wars whose month of onset occurred in the corresponding 6-month window. Because onsets are rare, we anticipate that this latter coding will be very sparse.

Leetaru, Kalev and Philip A. Schrodt. 2014. "GDELT: Global Data on Events, Location and Tone, 1979-2012.".

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

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

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

Appendix

CAMEO codes for predictors

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

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

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

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

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

1232:	Reject request to change policy
1233:	Reject request for rights
1234:	Reject request for change in institutions, regime
124:	Refuse to yield, not specified below
1241:	Refuse to ease administrative sanctions
1321:	Threaten with restrictions on political freedoms
1322:	Threaten to ban political parties or politicians
137:	Threaten with repression
172:	Impose administrative sanctions, not specified below
1721:	Impose restrictions on political freedoms

- ***Demands***, comprising:

104:	Demand political reform, not specified below
1042:	Demand policy change
1043:	Demand rights
1044:	Demand change in institutions, regime

- ***Accommodations***, comprising:

035:	Express intent to yield, not specified below
0342:	Express intent to change policy
0343:	Express intent to provide rights
0344:	Express intent to change institutions, regime
034:	Express intent to institute political reform, not specified below
080:	Yield, not specified below

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

Actors classified as "government"

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

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

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

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

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

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

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

- Municipal Party