Quantitative Predictors of Armed Conflict

Andrew Hoagland

June 5, 2021

1 Introduction and Motivation

By and large the world is getting safer. A large literature has shown that since the end of WWII, and especially since the end of the Cold War and fall of the USSR in 1991, the prevalence of violent conflict and violent deaths have decreased dramatically compared to historical averages. Pinker (2011) showed that violence had precipitously declined in the past century and audaciously claimed that ”today we may be living in the most peaceable era in our species' existence.” Although this is likely true, one look at the nightly news tells us that violent conflict still plagues different parts of the world, even if it may be smaller in scale and less deadly than in the past. In April 2021, the President of Chad, Idriss Deby was killed during a clash between government forces and Muslim rebels in his country's north. As of May 17, at least 246 people died in the latest flare-up of the ongoing conflict between Israel and Palestine. In recent anti-government protests in Colombia, over 40 people have tied, all but one of them being civilians. Clearly, armed and violent conflict persists and will persist into the future, despite the world becoming safer on average.

Researchers have long studied that factors that are correlated with and can be used to help forecast future armed conflict. Three sets of correlates are mostly strongly correlated with future armed conflict within a country: economic (i.e. poverty rate, infant mortality rate, GDP), demographic (i.e. population, level o education, the presence of several ethnic identities living under a single government), and political (i.e. level of corruption, type of government). The literature emphasizes the roll of ethnic tensions (Kalyvas (2008), Fearon and Laitin (2003)), economic greed or grievances (Collier and Hoeffler (1998), Collier and Hoeffler (2002), Collier and Hoeffler (2004)), and geographical factors and natural resource endowments (Ross (2004), Fearon (2005)). Blattman and Miguel (2010) provides a survey of the variables correlated with conflict, while World Bank (2011) outlines the economic and policy implications of identifying such variables.

2 Data and Empirical Approach

Conflict predictions studies thus far have primarily focused on various forms of regression. Hewitt (2008, 2010), Rost et al. (2009), and Goldstone et al. (2010) provide various multinomial logistic regression approaches for predicting the onset of future conflict. Hegre et al. (2010) use a multinomial logit model to account for non-linearities and simultaneously predicted the onset and termination of conflicts to forecast global and regional incidences of armed conflict. Regression approaches have proven fruitful in the past and have the benefit of providing coefficients with p-values so that researchers can measure the marginal effect and statistical significance of each variable. Because this past literature is so rich, I decided against a regression approach, as I believe it unlikely that I could produce novel results and contribute to it. Furthermore, the outbreak of civil unrest is quite random, and the data surrounding it is noisy, and thus regression approaches can suffer from poor out-of-sample accuracy, which can be particularly harmful in creating forecasts. Finally, most regression models thus far have focused solely on predicting conflict within the next year, whereas I hoped to see if reasonably accurate forecasts were possible over an expanded time horizon.

As previous studies have already identified which variables are most strongly correlated with internal conflict, I hoped to build on these models by using the identified variables to create a predictive model. Rather than identifying variables of interest, I wanted to focus solely on out-of-sample performance in hopes of creating as accurate a model as possible. In order to do so, I used decision tree models implemented in R. Decision forests were chosen for two primary reasons. First, by splitting at each node according to a binary cutoff for a single variable, decision trees are useful for identifying and incorporating non-linearities in the data, which I suspected likely existed. Second, in aggregating the predictions of several different trees, with each trained on a random sub-sample of the training observations and random sub-sample of the available explanatory variables, decision forests tend to have relatively high out-of-sample accuracy and avoid over-fitting to historical, which I wanted to avoid in order to make the model as applicable as possible going forward.

In constructing my model, I build largely in the work of Hegre et al. (2013), mentioned above. Specifically, I used their Predicting War, 2010-2015 dataset to train the model. The data set includes observations from 1970 to 2009 from 172 different countries and includes projections for the included variables based on UN data from 2010 to 2050, although did not use them in my analysis. For each year-country observation, the dataset includes over 350 historic, geographic, economic, and demographic predictor variables. The variable of interest being predicted is a categorical variable indicating whether there was a small, large, or no conflict. That is, the model predicts only the probability of internal violent conflict, which also happens to currently be the most deadly form. The data, as well as the distinction between the various forms of conflict, come from the Peace Research Institute Oslo, which publishes data on combat deaths down to the city level, including the location of battles, the combatants involved, the geographic location, if it was part of a larger conflict, and, most importantly to us, the number of deaths involved. Based on the PRIO's classification system, a small conflict is defined as one involving between 25 and 999 battle-related deaths, whereas a large conflict is one involving 1000 or more battle-related deaths. I also draw largely from Hegre et al.'s model and variable selection, performing no variable selection myself, but instead using those in their robust model. The exact variables used in the final model will be discussed below in more detail.

Although I draw on the work of Hegre et al., there are several important distinctions between their model and mine. As mentioned above, I use a classification decision forest model whereas they use a multinomial logit model. Additionally, the reported probabilities they report (i.e. what the model is predicting) is the probability of conflict within a country in the next year. While my primary model does the same thing, I also included two models that predict the instance of a conflict in a given country within the next three and five years, respectively. Finally, Hegre et al. used a binary classification framework, predicting small and large conflicts separately. That is, the regression predicts a dummy variable taking value 1 if there is a small conflict and 0 otherwise and separately predicts a dummy variable taking value 1 if there is a large conflict and 0 otherwise. My model instead uses a multiclass classification framework, predicting both types of conflict together, classifying each observation as either belonging to the small, large, or no conflict category. By far the most common category is no conflict. 83% of all observations correspond to there being no conflict in a given country, meaning that the model would be 83% accurate if it always predicted no conflict, and thus this is the baseline probability against which the models were compared.

3 Model

My model uses a decision forest framework, in which several classification trees are created, with each separately voting on which category the observation belongs and the votes being aggregated to output a probability and label. In a decision tree, each node considers a single variable and separates observation based on a binary cutoff. At each node the model's subjective probability of each category is updated before a final label is predicted based on the most probable classification at the terminal nodes.

In implementing the decision forest, I considered two types of classification tree aggregations, random forests (which I proceeded with for most of the quarter) and gradient boosted trees. Both form aggregations of decision trees, but random forests do so by creating n shallow (meaning they consider relatively few variables) trees all-at-once, while a gradient boosted tree model creates a single weak learner (meaning it predicts at or just above the baseline) and then iteratively adds other trees, focusing on observations and classes that perform poorly on the last iteration of the model. In both, each tree is trained on a random subset of the available observations in the testing set and a random subset of the explanatory variables available. In theory, gradient boosted trees should do better on tricky observations.

Each type of model was trained on all observations from 1970 to 2000 and tested on all observations from 2001 to 2009. The training-test cutoff of 2000 was chosen because the hyper-parameter tuning showed that test set accuracy plateaued after this year, it offered a roughly 3:1 training to test set size ratio, and historically I believed it to be a significant break, coming after the end of the civil and ethnic conflicts that arose in former Soviet states and right on the cusp of terrorism and rogue-stated conflict in the 21st century. In selecting between the two models, each first had its hyper-parameters tuned via a random search. The random forest contained 1000 trees (ntree = 1000), trying to split on 6 different variables at each node (mtry = 6) and maximizing accuracy (of correct predictions/ of observations in the test set). The xgboost (extreme gradient boosting) model used a learning rate of 0.19, had maximum depth of 7 nodes, and minimized according to a multi-class log-loss function, running up to 10,000 iterations but stopping early if the error not reduced for 100 consecutive rounds.

Table 1: Model Accuracy

Metric	Gradient.Boosted	Random.Forest
Accuracy	90.323	90.483
TP Rate for Small	49.686	55.975
TP Rate for Large	39.623	32.075

A table comparing the performance of the models is shown above. The models perform very similarly in terms of total accuracy. While the random forest performs 6 percentage points better in terms of the true positive rate for small conflicts, the xgboost model performs 7 percentage points better in terms of true positive rate for large conflicts. The consequences of a false negative (i.e. predicting no conflict but a large conflict occurring) are likely worse than the consequences of a false positive (i.e. predicting a large conflict but no conflict occurring) and the consequences of a large conflict are obviously much larger than those of a small conflict. Thus, I chose the xgboost model because of its superior accuracy on large conflicts.

The base xgboost model included 37 explanatory variables that predicted the incidence of conflict. It was built over 137 iterations, meaning there are 137 different classification trees in the ensemble. The 37 features include: an lagging indicator for the presence of small/large conflicts two years prior; ln(Duration of Peace); ln(Infant Mortality Rate); ln(Population); the percentage of the population that is a youth; the level of education in a country; neighboring country's IMR, education level, percentage of youth; indicator variables for whether the country produces oil, has an ethnic majority, has a neighbor involved in a conflict, is in West Asia or North Africa, is in Western Africa, is in Southern Africa; and various interaction terms.

In addition to the base model, four variations were created:

- Small conflict binary classifier: Instead of the multiclass classification of the base model, this model performed binary classification for whether there was a small conflict in a given country in a given year or not.

- Large conflict binary classifier: Instead of the multiclass classification of the base model, this model performed binary classification for whether there was a large conflict in a given country in a given year or not.

- Conflict within three years: This model was similar to the base model in that it performed multiclass classification, but it predicted used leading variables to predict whether there would be a certain type of conflict within in a country in any of the next three years

- Conflict within five years: The same as the three year model above but using a time horizon of five years.

4 Results and Analysis

Overall, the models performed significantly better than the baseline probabilities found by always guessing the most common class. This was true for each of the five models described in the previous section, but was particularly surprising for the models predicting conflict on a longer time horizon. The valuation metrics for the base and binary classification models was obtained by inputting the test set spanning 2001 to 2009 into each model to obtain predictions and comparing them to the true values. For the three-year model, the test set included data from 2001 to 2006, and for the five-year model the test-set included data from 2001 to 2004. The trimming of the test set for the final two models was done due to the nature of the leading indicator, which defaulted to no conflict and resulted in inflated accuracy for the longer time horizons.

Table 2: Model Performance

Model	Accuracy	Baseline
Base	90.323	83.0381
Binary (Small)	92.034	88.536
Binary (Large)	96.577	94.502
Three Year	87.846	78.175
Five Year	87.240	74.883

A summary of the performance of the model is shown in Table 2. The base, binary (small), binary (large), three year, and five year models beat the baseline by 7, 3.5, 2, 9.5, and 12.5 percentage points, respectively. As can be expected when simplifying a multi-class classification problem to a binary classification problem, the accuracy for each of the binary models is higher than that of the baseline, with the small version beating it by 1.7 percentage points and the large version beating it by 6.3 percentage points. Although the binary classifiers yielded a higher accuracy than the base model, this seems largely to be a function of problem simplification, not overall better performance. Indeed, the binary classifiers actually had a lower true positive rate for their respective types of conflict than the multi-class base model.

Because uncertainty and noise inexorably increase as we increase the time horizon, the three year and give year models unsurprisingly perform worse than the base model. What was surprising, however, was the magnitude of this decrease in performance. The three year accuracy was still 87.846% while the five year accuracy was 87.240%, a much smaller drop than I had anticipated. The drop in accuracy from the three year to the five year model is likely insignificant. This suggests that as the time horizon is initially expanded beyond a year, accuracy takes a noticeable hit, but as it increases beyond that the decrease in performance is much more gradual. The two forward-looking models perform especially well compared to the baseline probability of always predicting new conflict, as the likelihood of having no conflict over a three or five year time horizon decreases as compared to looking only at the next year. Overall, the performance of the three-year and five-year models was the most surprising finding from the entire project, and the accuracy of models with longer time-horizons is cause for cautious optimism about the feasibility and efficacy of this entire endeavor.

Table 3: Countries with Highest Risk of Conflict, 2009

Country	P.Conflict	Actual
Sudan	0.954	Small
Somalia	0.939	Large
Myanmar	0.934	Small
Democratic Republic of the Congo	0.931	None
Ethiopia	0.920	Small
Uganda	0.916	Small
Niger	0.849	None
Sri Lanka	0.844	Large
Afghanistan	0.816	Large
India	0.815	Large

Table 3 above exemplifies the predictions and provides a good sanity check for the model. The probability of conflict within a country is found by summing the probability that there will be a small conflict plus the probability that there will be a large conflict. One of the primary benefits of this model is that it not only produces a prediction for each country, but also returns the underlying prediction probabilities for each type of conflict from which the labels are determined. Due to this, we can analyze not only the overall prediction but also the chance of conflict. At the time, Sudan had been plagued by genocide in its Darfur region for several years and was on the brink of secession, with the Republic of South Sudan being formed in early 2011. Perhaps the most surprising inclusion on the list is India, which seems to be an outlier, at least in terms of economic and demographic variables, although its massive size makes the presence of at least one conflict killing at least 25 people likely. This also A full world map color-coded by the predicted conflict as well as a heat map for the overall probability of any type of conflict for 2009 can be found in the appendix.

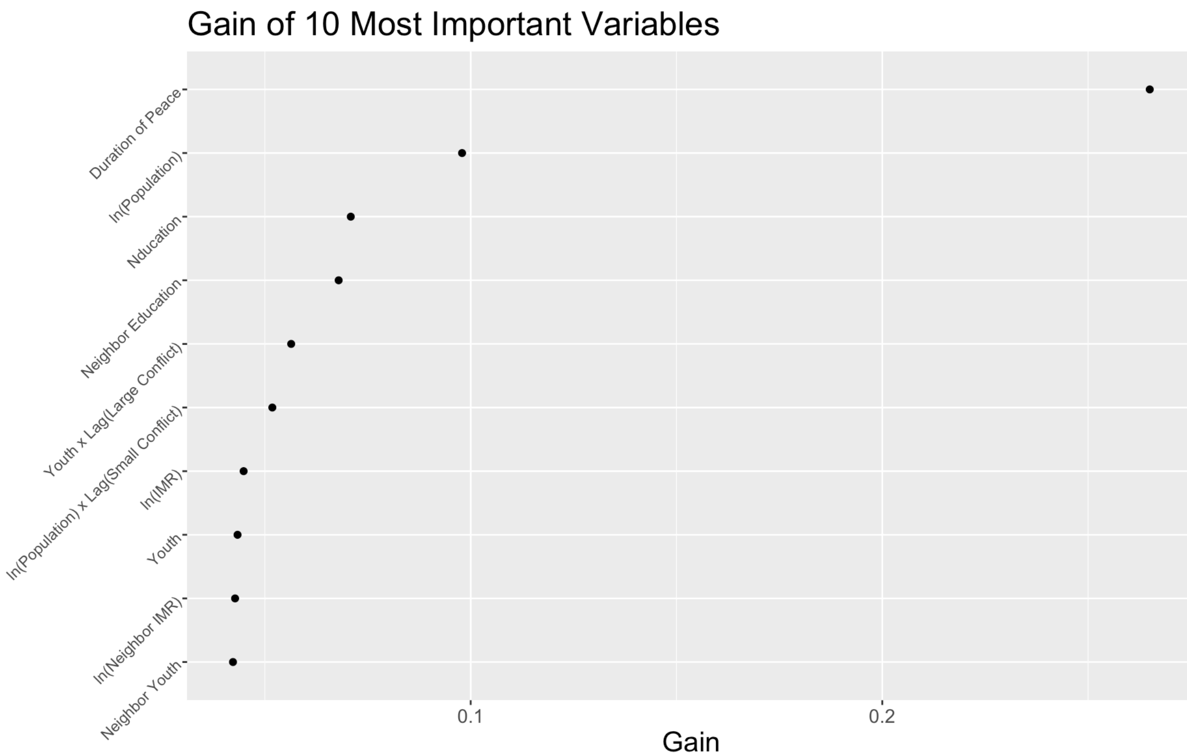

Figure 1: Plot of Variable Importance

Before discussing variable importance, is it important to note a few qualifications. First, unlike with regression, the importance of a variable does not relate to a physical value, i.e. a x% increase in the population size results in a y% increase in the chance of conflict. Rather, the measure of variable importance used in xgboost models is gain, which is essentially a weighted measure of the variable's frequency and influence across all component trees, with the sum of the gain for all included variables equalling 1. Second, because xgboost models are produced stochastically, the reported gain for each variable will differ slightly from model to model unless the random seed is set beforehand. The same is also true for the predictions, with the subjective confidence in each category changing slightly from model to model. Third, the gain is always positive and thus it does not tell us whether an increase in a variable is associated with an increase or decrease in the probability of conflict. The direction of its influence must be inferred from intuition or an associated regression study (this assumes that the effect of a variable is monotonic, which is not necessarily true with a classification tree).

By far the most important variable was the duration of peace in a country measured in years; the longer a country has gone with a conflict , the lower the chance of a conflict in the future. The huge influence of duration of peace is likely for two reasons. First, if it equals 0, it means that a conflict occurred in the past year and it is thus likely to persist. The absence of the lagging variables indicating the presence of a small or large conflict two years prior is likely in the ten most important variables is likely because their effect is picked up by duration of Peace. Second, a large duration of peace allows a country to rebuild its political and economic infrastructure, making future conflict less likely. The second most important variable was the size of the population. Simply, the number of reasons for conflict likely increase as the size of the population increases. Demographic variables such as the level of education and youth bulge in a country are also quite important, as are indicators of poverty like infant mortality rate. Surprisingly, the level of education and youth bulge of neighboring countries are important, while the presence of a conflict in a neighboring country is not. The interaction terms have no clear interpretation. The absence of regional indicator variables indicate that there is not significant regionality in the data, or, perhaps more likely, that the explanatory vairables are strongly correlated to region and the effect is picked up by them. Ultimately, we should not put too much stock into the reported variable importance, the purpose of this model is to predict as accurately as possible, not to determine the marginal effect of each co-variate.

Figure 2 shows that thhe accuracy of predictions does not seem to decrease significantly as the year of the observation becomes further from the time frame of the training data. This suggests that the model is likely insensitive to the particularities of the time it is trained on, although more analysis is necessary for concluding this.

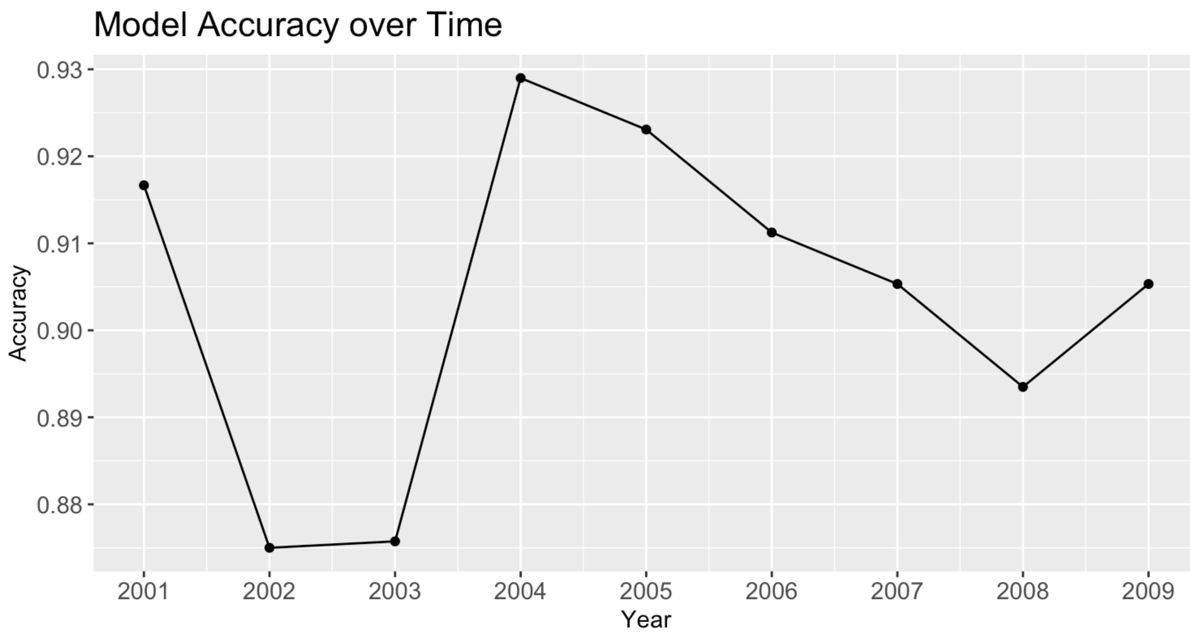

Figure 2: Plot of Model Accuracy Over Time

5 Conclusion

Before delving into the conclusions, implications, and future directions of research for this project, it is important to note the model's limitations. The primary limitation of these models is that they are only trained on data from 1970 to 2000 and tested on data from 2001 to 2009. Although the accuracy of the model does not seem to be particularly sensitive to how close an observation is temporally to the training set, it must be tested on more recent data before drawing a conclusion on its current efficacy. While I had intended to construct the Hegre et al. dataset for 2010 to 2020, joining so many different data sets (upwards of 10) was too complicated. Another limitation of the model is that is likely to miss systemic events that have spillover effects across countries. For example, few of the countries involved in the Arab Spring were predicted to have a particularly high chance of conflict in 2009, although the series of revolutions that began in 2011 would claim over 60,000 lives. Ultimately, the predictions from the model should be treated as forecasts rather than set-in-stone predictions, for armed conflict arises from the interaction of thousands if not millions of people, and is thus inherently random and difficult to predict.

The strong performance of all models, but especially the base, three year, and five year models are a solid proof-of-concept. They indicate that it is possible to predict the likelihood of conflicts ex ante up to several years and be significantly more accurate than the baseline probability of always causing the most common outcome. It is apparent that historic variables are the most important in producing accurate predictions, with a country's chances of experiencing violent conflict decreasing significantly as the duration of peace increases, although their predictive power decreases significantly as the time horizon expands. Additionally, demographic and economic variables are also influential, which is consistent with both intuition and the theory of why conflicts arise. While these results are promising, previous studies have already proven the concept of conflict prediction. I believe the most novel and important results from my model to be the accuracy of predictions as the time horizon increases.

The trend seen over the past 70 years of a decrease in the frequency of armed conflict and battle-related deaths is likely to continue into the future. The importance of the duration of peace indicates that conflicts are likely to relapse and, at the very least, current conflict makes future conflict more likely. Because the duration of peace is so important, nations and international governing bodies such as the UN and World Bank should provide aid to countries that recently experienced conflict. Whatever the form, whether it be political advisors, monetary aid, or the presence of peacekeeping troops, is likely to help rebuild the country's infrastructure and avoid future conflict. Even a brief period of peace and dramatically decrease the likelihood of future violence.

There are several ways in which this model could be improved, built upon, or extended. First and foremost, it should be trained and tested on data that is as recent as possible if it is to be used to forecast future conflicts as of 2021. The model incorporates historical, economic, and demographic variables, but does not include political variables, so the inclusion of factors like the type of government and level of corruption within a country is likely to improve performance. Given reliable data on the matter, the model could also be used to measure the effectiveness of outside intervention (i.e. UN peacekeeping troops, invasion by a neighbor) on decreasing the chance of conflict in the future. Finally, the model could be extended to predict interstate conflicts, although there are several complicating factors when more than one state is involved, and thus it would likely need to be significantly overhauled before it is ready to do so.

6 Appendix

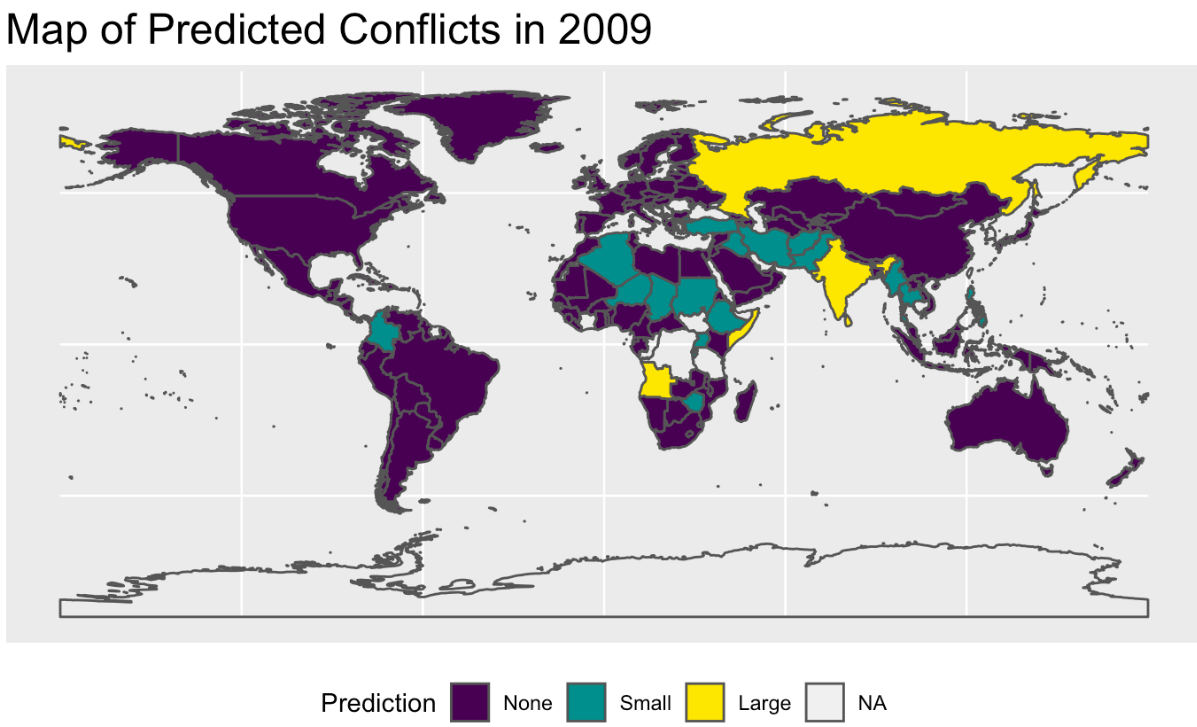

Figure 3: Map of Predicted Types of Conflict. Note that transparent regions were either not included in the model or did not join correctly.

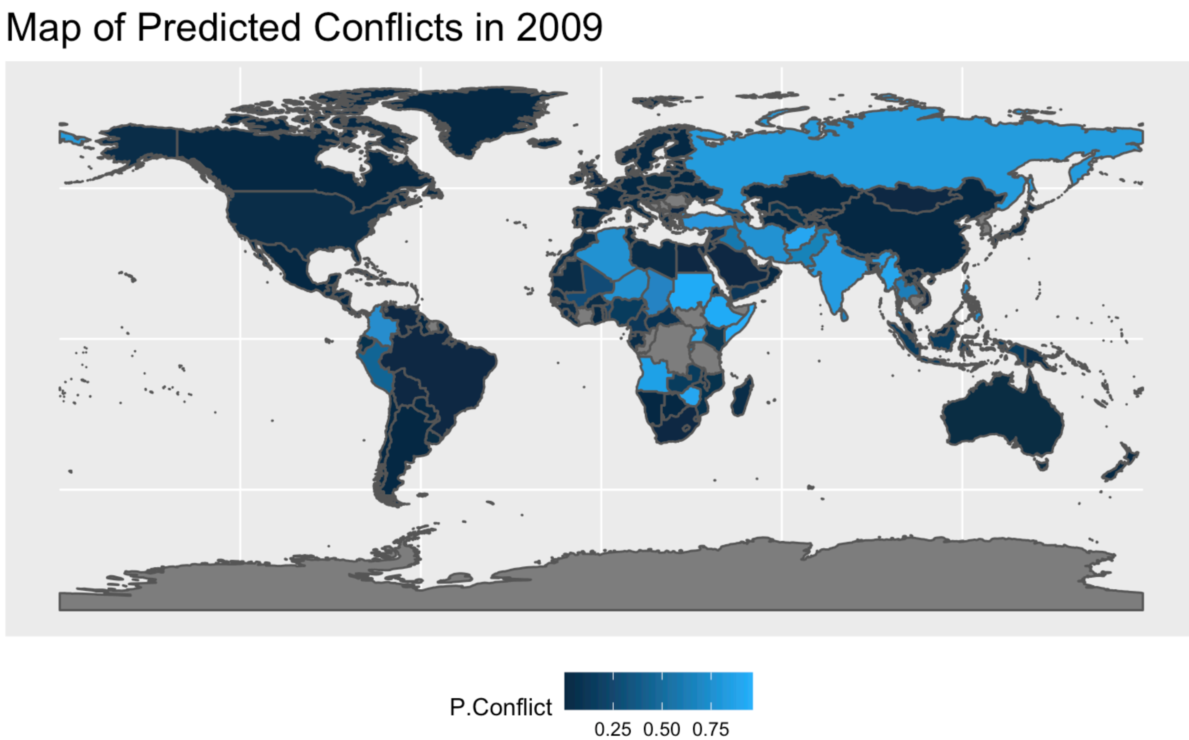

Figure 4: Map of Predicted Probability of Conflict. Note that gray regions were either not included in the model or did not join correctly.

7 References

Blattman, Christopher, and Edward Miguel. 2010. ”Civil War.” Journal of Economic Literature 48:1, 3-57.

Collier, P, and A. Hoeffler. 2002. ”On the Incidence of Civil War in Africa.” Journal of Conflict Resolution 46, 13–28.

Collier, P., and A. Hoeffler. 2004. ”Greed and grievance in Civil War.” Oxford Economic Papers 56, 563–595.

Collier, Paul, and Anke Hoeffler. 1998. ”On economic causes of Civil War.” Oxford Economic Papers 50(4), 563-573.

Fearon, J. D., and D. D. Laitin. 2003. ”Ethinicity, Insurgency, and Civil War.” Am. Polit. Sci. Rev. 97,75–90.

Fearon, J. D. 2005. ”Primary Commodities Exports and Civil War.” Journal of Conflict Resolution.

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

Hegre, Havard, Joakim Karlsen, Havard Mokleiv Nygard, Havard Strand, and Henrik Urdal. (2013) Predicting Armed Conflict, 2010-2050. International Studies Quarterly 57: 250–270.

Hewitt, J. Joseph. (2008) The Peace and Conflict Instability Ledger: Ranking States on Future Risks. In Peace and Conflict 2008, edited by J. Joseph Hewitt, Jonathan Wilkenfeld and Ted Robert Gurr. Boulder, CO: Paradigm.

Hewitt, J. Joseph. (2010) The Peace and Conflict Instability Ledger: Ranking States on Future Risks. In Peace and Conflict 2010. Boulder, CO: Paradigm Publishers.

Kalyvas, S. N. 2008. ”Ethnic Defection in Civil War.” Comparative Political Studies 41, 1043–1068.

Pinker, Steven. The Better Angels of Our Nature: A History of Violence and Humanity. Penguin Books, 2012.

Ross, M. 2004. ”What Do We Know about Natural Resources and Civil War? .” Journal of Peace

Research 41, 337–356.

Rost, Nicolas, Gerald Schneider and Johannes Kleibl. (2009) A Global Risk Assessment Model for Civil Wars. Social Science Research 38(4): 921–933.