Predicting Intrastate Conflict Using Machine Learning

Liz Masten and Cian Stryker

Abstract

Compared to causal modeling, using Machine Learning techniques to predict violent civil conflict is nascent within the broader conflict literature. In this paper, we run three Machine Learning models – OLS Regression, Ridge Regression, and Random Forest – on a dataset of our own making to predict instances of violent intrastate conflict. We find our Random Forest model to be the most accurate, but at the expense of high false negative occurrences, to which we assign greater weight considering their ramifications. Both our OLS and Ridge models outperform Random Forest in false negatives, and we assess our Ridge model to be an optimal compromise of high accuracy and low false negative scores. We conclude this paper with caveats to our models and remarks about the limitations and analytical trappings of Machine Learning applications to conflict prediction.[i]

Introduction and Literature Review

According to the Uppsala Conflict Data Program, the number of violent intrastate conflicts has increased over the past 50 years. This increase has brought death and instability, driven mass displacement, caused famines, and due to the protracted nature of many of these conflicts, destroyed the hopes of entire generations. Aside from the human suffering these conflicts cause within their borders, they send shocks into the international system. Because of the stakes involved, it follows that many scholars have tried to assign causality to these conflicts. However, the traditional metrics of causality and statistical significance may not be the best method for

[i] The project can be found in on GitHub: https://github.com/CianStryker/Predicting_Conflict_Project

identifying predictors of conflict and some are eschewing these techniques all together in favor of machine learning models that provide out-of-sample predictions.[1,2] In this paper, we join this trend towards predictive modeling and use three Machine Learning models – OLS regression, Ridge Regression, and Random Forest – to best predict violent civil conflict.

How does one winnow down all measurable attributes surrounding conflict to a reasonable number of indicators? We looked to the rich conflict literature as a starting point. Economists have long pointed to indicators such as GDP and income inequality, with GDP being one of the most reliable indicators of conflict. While most literature suggests that GDP and conflict are negatively correlated, others argue that conflict increases with wealth.[3,4,5] Income inequality can spur economic grievances, causing the poor to rebel, and profitable natural resources may be exploited by rebel groups thus increasing conflict.[6,7,8]

Others reject the resource curse argument in favor of institutional variables, such as weak state institutions.[9] They argue that factors which limit state control and cohesiveness include large population, regime type, regime change, and political instability. Others suggest that weak states can act as breeding grounds for terrorists, increasing domestic and transnational terrorist activity.[10,11] While there is precedent to reject the significance of identity and grievance-based causes of conflict, others find empirical support for the value of indicators such as ethnicity and other demographic qualities.[12,13] Demographic composition may have an indirect effect on conflict, making it an important indicator of conflict that performs poorly in traditional studies.[14]

The literature on civil conflict lacks strong empirical conclusions and a unifying framework across the field.[15] This paper seeks to contribute to this discipline by first identifying potential predictors of civil conflict while avoiding the trappings of statistical significance that make the discipline seem disjointed. Secondly, we run three Machine Learning models on these predictors

- OLS regression, Ridge Regression, and Random Forest - to attempt to find the optimal model to predict conflict based on our data. Lastly, we offer a robust critique of our models and the potential trappings of using Machine Learning methods for conflict studies more broadly.

Data Gathering

We divided our predictors into four categories: demographic, economic, institutional, and instability, sourcing them from nine separate datasets. The first predictors we chose to include came from the World Bank and included GDP (Current USD), GDP per capita, population size, population growth rate, infant mortality, male population percentage, and percent of income from natural resources.[16] Those predictors address general economic development, gender diversity, and the potential for natural resource dependence to affect conflict. We then included UN Human Development Index (HDI) values to offer a more robust analysis of development.[17] We wanted to also address how ethnic diversity may impact conflict as well, so we included a measure of ethnic fractionalization using "The Historical Index of Ethnic Fractionalization (HIEF)."[18]

For our institutional indicators, we included two Stockholm International Peace Research Institute (SIPRI) predictors including state military spending and military imports to account for the impact military spending may have on conflict.[19] Some literature suggests the presence of international peacekeepers has a negative correlation to conflict onset, so we included a binary predictor for peacekeeping presence using the Military and non-Military Interventions Dataset.[20][21] To account for regime type, we used both Freedom House's freedom index and several indicators from The Center for Systemic Peace's Polity V dataset, including measurements of political competitiveness, regulation, and durability.[22][23]

To investigate instability, we used the National Consortium for the Study of Terrorism and Responses to Terrorism's Global Terrorism Database (GTD).[24] For our outcome variable, conflict, we aggregated the number of intrastate conflict events per country per year using the Uppsala Conflict Data Program (UCDP/PRIO)'s Armed Conflict Dataset, version 20.1.[25] Our codebook can be found in Appendix A.

One of the key issues we had to address was that for each predictor there was a very high percentage of missing data. In fact, our dataset did not contain any rows without a missing observation. This is unfortunately a common issue for scholars who study civil conflict or international issues more generally. There is simply not exhaustive granular information for each country in the world for a consistent number of years, and collection and reporting mechanisms frequently do not exist in conflict zones. This lack of data consistency means that scholars of civil conflict often choose to impute their data in order to avoid large percentages of missing entries. Due to the ubiquity of imputation in this field, combined with the necessity to remove missing values for machine learning models to run, we chose to impute as well. To be as accurate as possible, we used R's MissForest package.[26] This method uses a Random Forest approach to imputation, first imputing the missing values using the mean, then separating the imputed and original values into a test and training dataset, respectively. The package then trains a Random Forest model on the imputed dataset to finalize the imputations. The process iterates to optimize the imputations. We let the model run until the package's automatic stopping criteria was reached, giving us 8 iterations. This was a relatively computationally intensive process which took 1.5 hours.

Models

Using our dataset, we decided to use three different algorithms for predicting conflict onset represented by our outcome variable *n_conflict*, or the number of conflicts that occurred within a year. We decided to use a multivariate linear regression model, a ridge regression model, and

Random Forest. For diversity in our modeling choices, we chose OLS and Ridge regressions because they are parametric and offer higher bias with hopefully lower variance, and Random Forest, which is non-parametric and therefore offers a lower bias but possibly higher variance.

For each model we divided our data randomly into a training data set and a test data set before running the model on the training data. After this, we used that model to predict *n_conflict* values for our test data. We then took the Mean Square Error (MSE) from each prediction model and then compared the number of accurate predictions, false negatives, and false positives between the real test data outcome values and our models' predictions to see which model performed the best. Using this approach, we were able to compare the accuracy and usefulness of each model for predicting the onset of civil conflict.

Our first model was simply a multivariate OLS linear regression that measured our outcome variable, *n_conflict*, as a result of every other predictor in our dataset. OLS regression is generally easy to interpret and has the highest bias of any machine learning technique, which means it may underfit the underlying data, but it will also continue to operate reasonably well even when new data is introduced. We concluded that the data were non-linear and as such, a linear regression would likely fit our training data poorly. However, we thought this model had a good chance of respectfully fitting the test data and would offer an interesting comparison to our two other models.

Second, we ran a ridge regression to adjust for some non-linearity. We once again split our data into a training and test data set, but then used cross validation to determine the optimal lambda coefficient for the ridge regression. Using that optimal lambda, we ran our ridge regression on the training data and used it to predict values for our outcome variable in our test data.

Third, we ran a Random Forest model. Scholars using Machine Learning to study civil conflict tend to prefer Random Forest models.[27][28][29] A Random Forest model is an ensemble learning method which can process regression and classification tasks. This model also reduces the likelihood that highly correlated variables will cause multicollinearity issues due to the randomness of the splitting criteria for individual decision trees (using a random subset of predictors rather than the entire ensemble of predictors). It is also a highly flexible modeling choice, which we thought would fit our highly non-linear and not uniformly distributed data well. Because a Random Forest model uses bagging to train an ensemble of decision trees, this model was the most computationally taxing of the three, taking approximately 25 minutes to run.

Results

Our OLS model performed surprisingly well with a relatively low MSE. This model had a generally high overall accuracy with low false positive and false negative rates. Considering our research objective of predicting conflict, we consider overall accuracy and false negative rate to the most important outcomes.

Predicting Civil Conflict

Linear Model Prediction Results

	Value	Percent
False Positive	*64*	*2.75%*
Accurate	*2176*	*93.59%*
False Negative	*85*	*3.66%*
MSE	*0.1071*	

Our Ridge Regression outperformed our OLS model in terms of MSE and slightly better in both false positives and overall accuracy. It did slightly worse than OLS, however, with false negatives.

Predicting Civil Conflict
Ridge Regression Prediction Results

	Value	Percent
False Positive	*53*	*2.28%*
Accurate	*2181*	*93.81%*
False Negative	*91*	*3.91%*
MSE	*0.0974*	

We had high hopes for our Random Forest model considering the ubiquity of its use in the discipline. It did have the lowest MSE of the three models, but it had the highest false negative score, at almost 10%. This is more than double the false negative score of the OLS and Ridge models, making this the most accurate predictor of conflict but the least useful overall for our purposes.

Predicting Civil Conflict
Random Forest Prediction Results

	Value	Percent
False Positive	*3*	*0.13%*
Accurate	*2094*	*90.06%*
False Negative	*228*	*9.81%*
	0.0577	

Analysis and Critique

When we look at the basic predictive accuracy in terms of the percentage of accurate predictions and the number of false negatives or false positives, our linear models outperform random forest considerably. Although somewhat rare, it is entirely possible for the models with the lower MSE to actually be outperformed by others with higher MSE but overall much better predictive accuracy.[30] For our purposes overall predictive accuracy is far more important considering the importance and potential consequences of civil conflict. With this assumption, therefore, we can rule out random forest as our model of choice. It is harder, however, to decide between linear regression and ridge regression because their values are so close. We believe that ridge regression is the superior model overall due to the slightly higher base predictive accuracy, even considering its slightly higher false positive rate.

Interestingly, our results suggest that the higher bias models of OLS and Ridge regressions are considerably more useful for predicting the onset of civil war than a lower bias model such as random forest, even though Random Forest has the lowest MSE. This may be due to overfitting in our Random Forest model, which is not terribly surprising considering that its non-parametric structure fits the signal and noise of our training data at the expense of the test data.

We suggest that another explanation for this may simply be in the underlying data itself. As mentioned in the *Data Gathering section*, missing data plague the study of civil conflict, which is not surprising considering the challenges of reliable data collection in conflict zones. This makes running Machine Learning models impossible, so we and other scholars address missing data via imputation. While we used a more complex method of imputation as opposed to a simple mean/median formula, problems remain. Imputation can distort the overall distribution of data and cause researchers to underestimate variance, thereby negatively impact the analytical accuracy of our models.[31]

Another major issue has to do with historical accuracy and coding continuity. States change over time with the creation of new countries, like Sudan and South Sudan, or they can fracture and then reconstitute, like North and South Yemen. Countries can dissolve like the USSR, or undergo name changes that signify radically different regime types, like Zaire. The coding difficulties associated with these types of countries is made all the more unfortunate considering that they offer the richest source of conflict data, so it is imperative to find a coding scheme that is both historically accurate and compatible with existing data.

Scholars and international organizations do not address this issue in a consistent manner, which causes issues for researchers like us who combine datasets. Some databases do not include data for countries that no longer exist, but do include data for individual countries that exist now even during years when they did not exist before. For example, the World Bank database does not contain Yugoslavia but does contain the present-day countries that formed after the breakup of Yugoslavia. Because administrative divisions in Yugoslavia reported economic data separately, the World Bank was able to geographically attribute data to present-day countries in that region and back-date their formation to that of Yugoslavia. A similarly notable omission in the World Bank dataset is the USSR. This creates interesting historical reimaginations.

Other databases choose to include previous states such as Yugoslavia and only provide data for countries starting from the year in which they were officially established. To reconcile these differences, we chose to code these "problem" countries in such a way that combined values of component countries to now-defunct states when possible, and when new state boundaries would not allow us to do this, we absorbed historical states into their present-day geographic equivalents. This is an intractable problem in conflict studies, and we do not claim to have made the best methodological choice. However, we feel that we made the best choice considering the scope of the project and the data available. Because any model is only as good as the underlying data, any conclusions we can draw from our model specifically and conflict modeling in general should be grounded in a healthy skepticism of their overall validity.

Conclusion

The empirical study of conflict has long sought to identify the causes of civil conflict and predict its next occurrence. The field is much more established around causal modeling than prediction, with the application of Machine Learning techniques a more recent development. This paper attempted to modestly contribute to this emerging field by predicting violent intrastate civil conflict using Machine Learning models. We ran three models – OLS Regression, Ridge Regression, and Random Forest – on a novel dataset created for this project. We found that our Ridge Regression model produced an optimal balance of low MSE and reasonable false negative predictions. However, there are fundamental problems with our data which are reflective of the limitations in the field more broadly. Ultimately, these issues are beyond the scope of this paper but are important considerations as the efforts to predict conflict move forward.

Works Cited

[1] Ward, Michael D., Brian D. Greenhill, and Kristin M. Bakke. "The perils of policy by p-value: Predicting civil conflicts." *Journal of peace research* 47, no. 4 (2010): 363-375.

[2] Ward, Michael D., and Andreas Beger. "Lessons from near real-time forecasting of irregular leadership changes." *Journal of Peace Research* 54, no. 2 (2017): 141-156.

[3] Basuchoudhary, Atin, Scott Shipp Hall, James T. Bang, and William F. Shughart II. "Predicting State Failure: Different Pathways into the Abyss."

[4] Garfinkel, Michelle R., and Stergios Skaperdas. "Economics of conflict: An overview." *Handbook of defense economics* 2 (2007): 649-709.

[5] Dal Bó, Ernesto, and Pedro Dal Bó. "Workers, warriors, and criminals: social conflict in general equilibrium." *Journal of the European Economic Association* 9, no. 4 (2011): 646-677.

[6] Collier, Paul, and Anke Hoeffler. "Greed and grievance in civil war." *Oxford economic papers* 56, no. 4 (2004): 563-595.

[7] Gurr, Ted Robert. "Sources of rebellion in Western societies: Some quantitative evidence." *The Annals of the American Academy of Political and Social Science* 391, no. 1 (1970): 128-144.

[8] Horowitz, Donald. "Ethnic Groups in Conflict, University of California Press Ed." (1985).

[9] Fearon, James D., and David D. Laitin. "Ethnicity, insurgency, and civil war." *American political science review* (2003): 75-90.

[10] Piazza, James A. "Draining the swamp: Democracy promotion, state failure, and terrorism in 19 Middle Eastern countries." *Studies in Conflict & Terrorism* 30, no. 6 (2007): 521-539.

[11] Basuchoudhary, Atin, and William F. Shughart. "On ethnic conflict and the origins of transnational terrorism." *Defence and Peace Economics* 21, no. 1 (2010): 65-87.

[12] Cederman, Lars-Erik, Nils B. Weidmann, and Kristian Skrede Gleditsch. "Horizontal inequalities and ethnonationalist civil war: A global comparison." *American Political Science Review* (2011): 478-495.

[13] Kuhn, Patrick M., and Nils B. Weidmann. "Unequal we fight: between-and within-group inequality and ethnic civil war." *Political science research and methods.* 3, no. 03 (2015): 543-568.

[14] Blimes, Randall J. "The indirect effect of ethnic heterogeneity on the likelihood of civil war onset." *Journal of Conflict Resolution* 50, no. 4 (2006): 536-547.

[15] Basuchoudhary, Atin, James T. Bang, Tinni Sen, and John David. *Predicting Hotspots: Using Machine Learning to Understand Civil Conflict*. Rowman & Littlefield, 2018.

[16] "World Bank Open Data | Data," accessed November 12, 2020, https://data.worldbank.org/

[17] "Human Development Index (HDI) | Human Development Reports," United Nations Development Programme, accessed November 12, 2020, http://hdr.undp.org/en/content/human-development-index-hdi

[18] Lenka Drazanova, "Historical Index of Ethnic Fractionalization Dataset (HIEF)" (Harvard Dataverse, 2019), https://doi.org/10.7910/DVN/4JQRCL.

[19] "SIPRI Databases | SIPRI," Stockholm International Peace Research Institute, accessed November 12, 2020, https://www.sipri.org/databases

[20] "MILINDA - Military and Non-Military Interventions Dataset 1947-2016," Georg-August-Universitat Gottingen, accessed November 12, 2020, http://lehrstuhlib.uni-goettingen.de/milinda.html

[21] Lisa Hultman, Jacob D. Kathman, and Megan Shannon, "United Nations Peacekeeping Dynamics and the Duration of Post-Civil Conflict Peace," *Conflict Management and Peace Science* 33, no. 3 (July 1, 2016): 231–49, https://doi.org/10.1177/0738894215570425.

[22] "Polity IV Dataset," Center for Systemic Peace, accessed November 12, 2020, https://www.systemicpeace.org/index.html

[23] "A Leaderless Struggle for Democracy," Freedom House, accessed November 12, 2020, https://freedomhouse.org/report/freedom-world/2020/leaderless-struggle-democracy;

[24] "START Global Terrorism Database," DHS | University of Maryland, accessed November 12, 2020, https://www.start.umd.edu/research-projects/global-terrorism-database-gtd

[25] "UCPD/PRIO Armed Conflict Dataset, Version 20.1," Department of Peace and Conflict Research, Uppsala University," accessed November 12, 2020, https://ucdp.uu.se/

[26] Stekhoven, Daniel J., and Maintainer Daniel J. Stekhoven. "Package 'missForest'." (2012).

[27] Muchlinski, David, David Siroky, Jingrui He, and Matthew Kocher. "Comparing random forest with logistic regression for predicting class-imbalanced civil war onset data." *Political Analysis* (2016): 87-103.

[28] Colaresi, Michael, and Zuhaib Mahmood. "Do the robot: Lessons from machine learning to improve conflict forecasting." *Journal of Peace Research* 54, no. 2 (2017): 193-214.

[29] Basuchoudhary, Atin, James T. Bang, Tinni Sen, and John David. *Predicting Hotspots: Using Machine Learning to Understand Civil Conflict*. Rowman & Littlefield, 2018.

[30] James Mccaffrey, "Mean Squared Error versus Predictive Accuracy," *James D. McCaffrey* (blog), August 20, 2011, https://jamesmccaffrey.wordpress.com/2011/08/20/mean-squared-error-versus-predictive-accuracy/.

[31] Paul Lodder, "To Impute or Not Impute, That's the Question," *Advising on Research Methods: Selected Topics 2013*, 2014, https://research.tilburguniversity.edu/en/publications/to-impute-or-not-impute-thats-the-question.

APPENDIX A:
Explanation of Variables

GENERAL INDICATORS:

Country
(*country*)
Character

Taken from World Bank GDP data. Includes only sovereign, internationally recognized countries with the exception of the Palestinian Territories; 203 countries are included.

Year
(year)
Numeric

Taken from World Bank GDP data. Years run from 1960-2019.

Continent
(*continent)*
Factor

Taken from World Bank but standardized with the R package *countrycode*.

DEMOGRAPHIC INDICATORS:

Population
(population)
Numeric

Taken from World Bank GDP data.

Population Growth Rate
(population_growth_rate)
Numeric

Taken from World Bank GDP data.

Infant Mortality
(infant_mortality)
Numeric

Taken from World Bank GDP data.

Gender Breakdown
(percent_male_population)
Numeric

Taken from World Bank GDP data. Values represent the percent of the population that is male.

Human Development
(hdi)
Numeric

Taken from the UN Human Development Index.

ECONOMIC INDICATORS- GENERAL:

GDP
(gdp_current_us)
Numeric

Taken from World Bank GDP data. Total GDP value per country in USD.

GDP Per Capita
(gdp_per_capita)
Numeric

Taken from World Bank GDP data. GDP value per country per capita in USD.

Unemployment
(unemployment)
Numeric

Taken from World Bank GDP data.

ECONOMIC INDICATORS- RESOURCE CURSE:

Oil Rent as Percent of GDP
(oil_rent_percent_gdp)
Numeric

Taken from World Bank GDP data.

Natural Gas as Percent of GDP
(natural_gas_percent_gdp)
Numeric

Taken from World Bank GDP data.

Mineral Rent as Percent of GDP
(mineral_rent_percent_gdp)
Numeric

Taken from World Bank GDP data.

Natural Resource Rent as Percent of GDP
(natual_resources_rent_percent_gdp)
Numeric

A combination of the previous rent predictors taken from World Bank GDP data

ECONOMIC INDICATORS: MILITARY SPENDING AND ARMS SALES

Military Spending
(military_spending)
Numeric

Total amount spent on military expenses (in USD), taken from the Stockholm International Peace Research Institute.

Military Imports
(military_imports_TIV)
Numeric

Total value of military hardware (in USD) imported into a country, taken from the Stockholm International Peace Research Institute.

REGIME INDICATORS:

Freedom House Freedom Index
(freedom_index)
Numeric

Taken from Freedom House's Freedom Index. Values are between 0 and 100 with lower values indicating higher levels of freedom. Years run from 1993-2016.

Polity Change
(change)
Integer

Taken from the Center for Systemic Peace's Polity V dataset. Values represent the total change in polity value from the previous year. Values run from 1 to 35. Higher numbers indicate greater Polity changes, which indicate greater instability.

Executive Transfers
(xrreg)
Factor

Taken from the Center for Systemic Peace's Polity V dataset. Values represent the regulation of executive transfers.

1 = Unregulated. changes occur via forceful seizes of power (ex: coups)
2 = Designated. Changes occur via political elite.
3 = Regulated. Changes occur via competitive elections.
-88 = Transitional Period. Denotes period of time between polities, after one polity has dissolved but before the new polity has been completely established.
-77 = Interregnum Period. Denotes complete collapse of central authority.
-66 = Interruption Period. Denotes periods of foreign occupation or attempts to create distinct federations within the country separate from the sovereign governing body.

Executive Selection
(xrcomp)
Factor

Taken from the Center for Systemic Peace's Polity V dataset. Values represent the competitiveness and mechanisms of executive selection.

1 = Selection. Changes occur via heredity succession.
2 = Transitional. Changes occur via a combination of heredity succession and elections.
3 = Election. Changes occur via competitive elections.
-88 = Transitional Period. Denotes period of time between polities, after one polity has dissolved but before the new polity has been completely established.
-77 = Interregnum Period. Denotes complete collapse of central authority.
-66 = Interruption Period. Denotes periods of foreign occupation or attempts to create distinct federations within the country separate from the sovereign governing body.

Executive Recruitment
(xropen)
Factor

Taken from the Center for Systemic Peace's Polity V dataset. Values represent the openness of chief executive recruitment.

1 = Closed. Designated by hereditary succession only. Values are coded after two generations.
2 = Dual- Designated. Designated by hereditary succession plus selection of a chief minister by other means NOT including elections.
3 = Dual- Elected. Designated by hereditary succession plus selection of a chief minister by competitive elections.
4 = Open. Designated by competitive elections or elite designation.

-88 = Transitional Period. Denotes period of time between polities, after one polity has dissolved but before the new polity has been completely established.
-77 = Interregnum Period. Denotes complete collapse of central authority.
-66 = Interruption Period. Denotes periods of foreign occupation or attempts to create distinct federations within the country separate from the sovereign governing body.

Political Competition
(parreg)
Factor

Taken from the Center for Systemic Peace's Polity V dataset. Values represent the degree of institutionalization, or regulation, of political competition

1 = Unregulated. No enduring national political organizations, no regime controls on political activity.
2 = Multiple Identity. Stable political groupings exist that may share common interests.
3 = Sectarian. Political demands and benefits are centered around ethnic or religious identifications.
4 = Restricted. Regular political institutional participation is operations exclusively and with respect to factional groups.
5 = Regulated. Little use of coercion in political institutions, healthy level of political institutional competition.
-88 = Transitional Period. Denotes period of time between polities, after one polity has dissolved but before the new polity has been completely established.
-77 = Interregnum Period. Denotes complete collapse of central authority.
-66 = Interruption Period. Denotes periods of foreign occupation or attempts to create distinct federations within the country separate from the sovereign governing body.

Political Competition Restriction
(parcomp)
Factor

Taken from the Center for Systemic Peace's Polity V dataset. Values represent the extent of government restriction on political competition and the ability for citizenry to pursue alternative policy choices.
0 = Not Applicable. Occurs when Polities are previously coded as *Unregulated*.
1 = Repressed
2 = Suppressed
3 = Factional
4 = Transitional
5 = Competitive
-88 = Transitional Period. Denotes period of time between polities, after one polity has dissolved but before the new polity has been completely established.
-77 = Interregnum Period. Denotes complete collapse of central authority.
-66 = Interruption Period. Denotes periods of foreign occupation or attempts to create distinct federations within the country separate from the sovereign governing body.

Executive Constraints
(xconst)
Factor

Taken from the Center for Systemic Peace's Polity V dataset. Values represent the mechanisms of institutional constraints on executive authority and decision making.

1 = Unlimited authority
2 = Intermediate between 1 and 3
3 = Slight to moderate limitations
4 = Intermediate between 3 and 5
5 = Substantial limitations
6 = Intermediate between 5 and 7
7 = Executive parity or subordination to regulatory groups (ex: legislature)
-88 = Transitional Period. Denotes period of time between polities, after one polity has dissolved but before the new polity has been completely established.
-77 = Interregnum Period. Denotes complete collapse of central authority.
-66 = Interruption Period. Denotes periods of foreign occupation or attempts to create distinct federations within the country separate from the sovereign governing body.

Polity
(polity2)
Factor

Taken from the Center for Systemic Peace's Polity V dataset. Values represent adjusted time-series Polity values, coded for Polity's fifth iteration of their Polity dataset.

-10: -6 = Autocracies
-5: 5 = Anocracies
6: 10 = Democracies
-88 = Transitional Period. Denotes period of time between polities, after one polity has dissolved but before the new polity has been completely established.
-77 = Interregnum Period. Denotes complete collapse of central authority.
-66 = Interruption Period. Denotes periods of foreign occupation or attempts to create distinct federations within the country separate from the sovereign governing body.

Regime Durability
(durable)
Integer

Taken from the Center for Systemic Peace's Polity V dataset. Values represent the number of years since the most recent regime change (defined by a three point change in Polity2 score over a period of three years or less). Can also denote the end of a transition period defined by the lack of stable institutions.

INSTABILITY INDICATORS:

Number of Unique Conflicts
(n_conflict)
Numeric

Taken from the UCDP/PRIO Armed Conflict Dataset Codebook, version 20.1. Variable created by taking *type_of_conflict*, excluding interstate conflicts, then taking the sum of instances of conflict per country/ year.

Regime Fragmentation
(authority_challenge)
Factor

Taken from the Center for Systemic Peace's Polity V dataset. Created from the Polity V *change* variable, this is a combination of authority interruption, authority collapse, authority transition, and regime change (polity in transition), state disintegration, state transformation, state creation, and state demise. This was to (1) keep the "change" scores numeric from -20:20 for better interpretability; (2) make imputation possible; and (3) make an instability metric that was more robust than state failure (*sf*), which was missing 98% of its observations.

0 = Regime did not experience fragmentation
1 = Regime experienced fragmentation

Peacekeeping Presence
(peace_keeping)
Factor

Taken from the MILINDA - Military and Non-Military Interventions Dataset.

0 = No international peacekeeping presence
1 = International peacekeeping presence

Terrorist Attacks: State Target
(state_target)
Integer

Taken from the Global Terrorism Database. Created from the *targtype1* variable, this is the sum of state targeted terrorist attacks per country/ year. We considered state targets to be: government(general), police, military, government(diplomats), and violent political parties.

Terrorist Attacks: Nonstate Target
(nonstate_target)
Integer

Taken from the Global Terrorism Database. Variable created from the sum of nonstate targeted terrorist attacks per country/ year. We considered nonstate targets to be: businesses, abortion-related, airports/ aircraft, educational institutions, food or water supply, journalists/ media, maritime, NGO, private citizens/ property, religious figures/ institutions, telecommunications, terrorists/ non-state militias, tourists, transportation, and utilities.

Terrorist Attack Death Toll
(terrorism_kill)
Integer

Taken from the Global Terrorism Database. Number of people killed in terrorist attacks.

Terrorist Attack Wounded
(terrorism_wound)
Integer

Taken from the Global Terrorism Database. Number of people wounded in terrorist attacks.