Quantifying the Relationship Between Economic Crises and Political Unrest

Andrew Furlong

April 29, 2022

Contents

Abstract

It is often assumed that banking crises (a phenomenon wherein a country's banking services do not have sufficient liquidity to cover all demand) have a relationship with political instability. However, political scientists have had difficulty quantifying this relationship, due to the many factors that play a role in these events. Using economic and political data provided by Dr. Daniel Hansen (CMU, IRP), our paper aims to build on preliminary research in this area, specifically focusing on more advanced statistical modeling techniques. This will allow us to predict revolutions at a high accuracy rate and identify variables which are strongly related to the onset of political instability.

We construct our models on our entire dataset rather than a subset of specific countries in order to focus on broader trends in the geopolicial climate and also learn more about which factors in particular are influential in predicting unrest. We constructed additional time series models such as the Hidden Markov Model to understand whether this played a role in the onset of revolutions.

Over the course of this project, we conclude that we can predict political instability using economics data, and we identify a number of variables in our dataset relating to economics, geography, and government structure that are strong predictors of political instability in countries around the globe.

1 Introduction

Throughout history, citizens have protested against various regimes. While there is a variety of causes for political instability, the most common reasons are often economic: The United States of America was founded after a tax revolt, the Soviet Union collapsed in part due to its unstable economy, and many South American revolutions started because civilians were upset over unjust economic conditions. Historically, many analyses of this relationship have focused on the impact of political instability on economic growth. While this research has yielded promising results, these analyses also overlook the potential to utilize the same data in order to understand the converse relationship: the impact of economic instability on political crises.

One of the main reasons that little focus is given to this task is its relative complexity. The statistical models used to explore the impact of political crises on economic growth are not sufficient for predicting the converse relationship. Additionally, there is more value in forecasting economic growth. For many researchers, understanding future economic performance is an important question for any potential investors in our global economy, and predicting political instability is a less relevant research topic.

There has, however, been some work in this area. Economics researchers at Harvard have observed a correlation between economic decline and political instability (*Political Instability and Economic Growth*), often using electoral volatility or public trust as measures of such instability, but failed to apply statistical modeling to this research. Additionally, there is evidence suggesting that there is a relationship between a nation's government system and its likelihood of undergoing a financial crisis (*Democracy and Financial Crisis*). In this thesis, we will build off of these conclusions and apply multiple statistical models to our data, with the focus of our research being on time series models such as Hidden Markov Models (HMMs).

Even though time-series models have been used in other disciplines, there is no documentation of researchers utilizing time series to model political instability, with good reason. Probability-based time series models are notoriously difficult to build, and are inflexible: each model would likely only generalize to a single country in our dataset (out of 271 total), meaning that it would be difficult to draw general conclusions from any one of them. Due to our application of statistical modeling when researching this topic, we expect this research to be relatively new in the field, and could be used as a starting point for future researchers.

This thesis aims to build on existing research in the relationship between economic crises and political instability by utilizing previously unused statistical methods. Specifically, we chose to focus on modeling country-year economic data and the presence of a revolution in that country.

In Section 2, we lay out the background of existing research in this field. We describe our data sources and preprocessing techniques in Section 3 and our methodology and extensions to the field in Section 4. We explore our models in Section 5, before exploring our results in Section 6. Finally, we lay out our conclusions and potential future extensions of this project in Section 7.

2 Background

Despite the limited research exploring the impact of economic crises on political instability, some research exists for us to use as a starting point. A number of economists have explored the impact of political instability on economic growth. In this section, we will highlight notable research that has been conducted in this field, as well as our potential outcome measures for political instability and why we decided to incorporate a selection of these variables into our final model.

2.1 Prior Research

There are several papers that have researched the relationship between economics and political instability, which have had success in detecting this relationship. In his paper *Democracy and Financial Crisis*, Philip Lipscy identifies multiple governmental factors that can influence the likelihood of debilitating financial crises. First, he notes that democratic nations with greater executive constraints are more likely to experience financial crises. This is likely due to the time required to respond to market fluctuations: nations such as the United States (where the executive branch does not have unilateral authority) must wait for congressional approval before combating any financial instability. However, nations such as China (where the legislative branch exists as a rubber stamp for the executive) are able to respond faster than democratic nations. This is best exemplified by the Financial Crisis in 2008, where the United States economy contracted significantly but China largely escaped any consequences.

Interestingly, Lipscy also notes that while democratic nations have a greater likelihood of financial crises, wealthy democratic nations (as seen in North America and Europe) do not have an increased risk of political instability stemming from financial crisis. From this, we can surmise that wealthy nations are often the most politically stable, and that relatively poorer nations (especially poorer democracies, which are often established after the fall of a dictator) are at the greatest risk of political instability following a financial crisis.

Similarly, Alesina et al. concluded in their paper *Political Instability and Economic Growth* that the relationship between political instability and economic conditions exists, and there is a definite correlation between increased political instability and economic growth in a country. While this is not directly related to our topic (as we wish to explore the converse of this relationship), they noted that political unrest influenced by income inequality (as measured by the Gini coefficient) is often the most significant factor in slowing growth. This supports the conclusions drawn from Lipscy, who hypothesized that poorer nations (which often have high levels of income inequality) are more likely to experience political instability.

Other conclusions drawn by Alesina et al. support the modeling techniques we will explain in Section 4. They note a nonlinear relationship between economic conditions and political instability, which supports our use of supervised learning techniques such as logistic regression and support vector machines, each of which is able to identify nonlinear relationships in the data. Furthermore, Alesina et al. note a compounding effect of economic growth on political instability. While this may seem trivial, this supports the use of time-series modeling techniques to predict economic growth, which has not been

tested in the field.

There are also other papers which attempt to use Markov Models to predict conflicts in various regions. Duncan et al. attempt to use a Markov model to predict conflict on the Asian continent in their paper *Markov Chain Models for Conflict Analysis*, where they attempted to use a Markov model to assess Sino-Indian relations between 1959 and 1964. While this paper was specifically focused on conflicts between China and India over a smaller time frame and failed to account for economic factors (due to its focus on international relations), this does lend credence to the ability for Markov Models to be used in predicting political crises. For this thesis, we will attempt to construct HMMs using our data.

2.2 Outcome Variable Selection

When we are measuring for political instability, there are many ways to do so. Data regarding political instability was collected in the International Country Risk Guide, an annual database built for financial investors looking to assess the risk of investing in various nations. The main variables for political instability in this dataset are assassinations, mass strikes, government crises, administrative purges, riots, anti-government demonstrations, and revolutions. Each of these variables measures a different aspect of political instability, and we could construct statistical models to predict any of these metrics.

When constructing our statistical models, we chose to focus on the number of revolutions in a country each year. This variable was chosen over the others for multiple reasons. First, revolutions are more significant than other measures of political instability, and have a greater potential to lead to widespread devastation. Therefore, accurately predicting future revolutions is of great importance for political scientists. Second, prior models of revolutions have been uninterpretable for political scientists. Prior attempts by CIA researchers at predicting revolutions have been limited to black-box algorithms, which yield accurate predictions but fail to explain how they found their results. Any attempts at producing interpretable models for revolutions and armed conflict would advance our understanding of these relationships.

3 Data

The data used for this thesis was provided in five sections, each of which focuses on different aspects of our research question. We were given economics, governing, and political instability data for countries between 1960 and 2017, which we will use for data visualization, as well as for both supervised and unsupervised learning.

3.1 Data Sources

Our first dataset contains information pertinent to the economic conditions in every country between 1960 and 2017. This includes GDP, growth, and debt values per year. This dataset was supplemented with measurements of the corruption present in every country and whether the nation was undergoing any sort of financial crisis. Additionally, we were provided access to the polity dataset, a resource which aggregates the authority of a regime, assigning scores from -10 (representing a hereditary monarchy) to 10 (a full

democracy). The polity dataset shares the individual category scores for each country, which allows us to consider the structure and openness of a regime when building our statistical models. We were also provided access to the International Country Risk Guide, which contains information on which countries are at risk of undergoing political crises in a given year. This was used in our supervised learning as our target variables for political instability.

3.2 Data Merging and Cleaning

Prior to analysis, we merge our data into one table. First, we standardize country names in each dataset. This is particularly difficult, due to the changing nature of world geopolitics and inconsistent nature in data collection. Next, we join each of our five datasets so that each entry in our final dataset contains annual information on the economic and political information for each country in our dataset. In order to do this, we construct an additional variable in each dataset containing a specific numeric code for each country-year combination. With this variable, we were able to merge the five datasets using `dplyr` functions in `R`.

Within our new data table, we conduct additional data cleaning. We identify over 100 variables which we considered unnecessary, due to their failing one of the following three tests: are too many data points missing, is this data redundant, or is it uninformative?

The data cleaning process reduces the dimensionality of our dataset by 62%, which leaves us with one problem: the amount of missing data present. Approximately 32% of our dataset is missing data, which can significantly impact the predictive ability of any model we construct. To solve this problem, we turn to multiple imputation, a technique which imputes an incomplete column (the target column) by generating 'plausible' synthetic values given other columns in the data. While we were initially concerned about the effect multicollinearity would have on multiple imputation, numerous researchers have proven that when utilizing a CART (Classification and Regression Tree) framework, multicollinearity is minimized. Thus, we proceed with Multiple Imputation to reduce the amount of missing data.

3.3 Final Dataset

Our final dataset has approximately 13000 rows and 97 columns, which capture different information about each country on an annual level. We have 57 years worth of data in this final dataset, which contain measurements on 197 countries. It is worth nothing that before we constructed any statistical models, we filtered the first ten years of the data, due to the fact that only 163 countries were measured in these first ten years. We now have a centralized dataframe containing have measurements on the economy, system of government, levels of corruption, and political instability of all countries, which will be invaluable when we proceed with our unsupervised and supervised learning techniques in Section 4.

4 Methodology

We will now utilize both supervised and unsupervised learning techniques on our dataset. These differences should allow us to understand different relationships in the data, and identify important variables for predicting revolutions. Unsupervised learning (which is conducted in Section 4.1) seeks to understand relationships between our covariates without knowing anything about our dependent variable (the number of revolutions). This can help us identify any patterns or similarities between countries, which can help us when we conduct supervised learning. Supervised learning seeks to predict our dependent variable using our covariates, which allows us to understand which variables influence the presence of a revolution in a country. By implementing multiple statistical learning techniques, we hope to maximize the amount of information we obtain from this dataset.

4.1 Unsupervised Learning

Before constructing any statistical models (which are detailed in Section 4.2), we first use unsupervised learning techniques such as clustering to identify trends in our dataset. We utilize partition-based clustering for this research project (provided in the `Cluster` package in `R`), due to its ability to cluster using factor variables in order to identify similarities or dissimilarities which can be used to further our understanding of political instability. Before we implement our clustering algorithm, we removed the following variables: country name, region, and revolutions (an indicator variable identifying countries which had at least one revolution in a given year).

Once these variables are removed (and we were confident any confounding factors were removed), we proceed with our clustering algorithm. Additionally, we cluster on individual years of data for two reasons. First, using a single year of data for each round of clustering allows us to understand trends in a single year, which is more relevant than clustering over all forty years on our dataset. Additionally, using a single year of data for clustering allows us to construct graphs identifying various clusters, which can further our understanding of political instability and any pertinent factors.

Below is a plot of our clusters for 1994 (a year with the greatest number of revolutions worldwide), and a plot of the clusters for 2017 (the most recent year in our dataset).

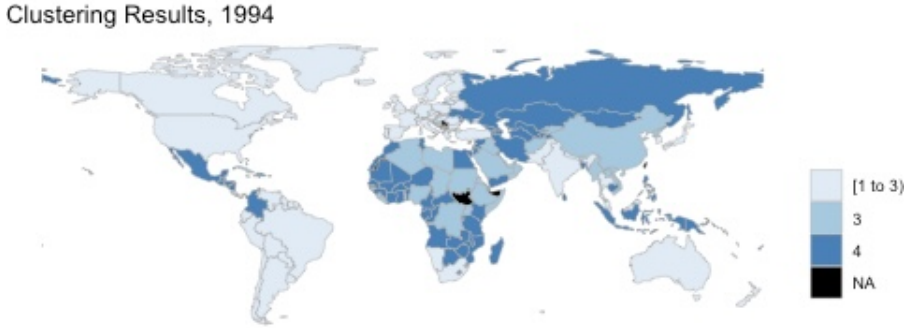

Figure 1: Clustering Results for 1994

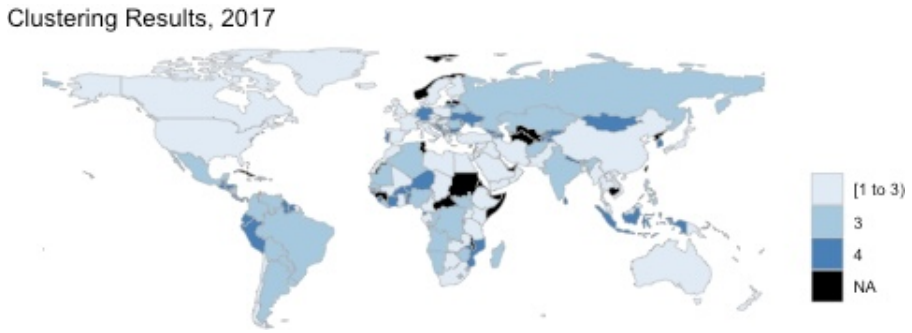

Figure 2: Clustering Results for 2017

The optimum number of clusters for each year in our dataset is 3-4, as shown above. This optimum was selected based on which number of clusters minimized the dissimilarities within each cluster, and is more robust than the standard clustering algorithm (measuring squared euclidean distance) due to its ability to incorporate categorical variables into the clustering algorithm. In both figures, we can identify specific geographic trends in our clustering algorithm, which is notable due to the fact that these variables were removed from the input dataset.

We can also begin to segment our countries into various groups based on the results from our clustering algorithm. Clusters 1 and 2 are often countries with strong economies, and often tend to have more political stability. This includes countries such as the United States, most EU countries, China, and Saudi Arabia. Cluster 3 countries tend to have developing economies, and are also more likely to be struggling with political instability (especially in our 1994 graph). This includes nations such as Brazil, China (in 1994), and Northern African countries. Finally, countries in the fourth cluster tend to have the weakest economies and are the most at risk of experiencing a revolution. This includes Central American countries, many Middle Eastern nations, and southeast Asian countries. We utilize the results of our clustering when we implement supervised learning models.

4.2 Supervised Learning

After completing our unsupervised learning, we turned to supervised learning techniques, which can allow us to understand specific variables that have strong predictive ability. Our first such model was the logistic regression model, which assesses the probability of a revolution in a country during a specific year. We chose this model for multiple reasons. First, the non-linear relationship between economic conditions and political crises described in Section 2.1 suggests that a logistic regression model (which is non-linear) will perform well. Second, a logistic regression model is very interpretable: we can use the coefficient values from our model to understand the significance of each variable in predicting revolutions. For our logistic regression and random forest models, we used a probability threshold of 0.5 to determine whether there was a revolution or not in a country.

We start by constructing a multivariate logistic regression model on our data: significant coefficient values are displayed in Appendix A.3 (note: not all coefficient values can be displayed due to the number of variables in our data). We observe that the following

variables increase the likelihood of a revolution in a given year: the Gini coefficient (a measure of income inequality), the beginning of a banking crisis, and the presence of a credit bubble. The accuracy for this model was 97%. While this appears to be significant, our sensitivity value (representing the percent of properly classified revolutions) of 87% suggests that our model is struggling due to unbalanced classes in our dataset (less than 15% of the entries in our dataset are countries experiencing a revolution in any given year). For this reason, we construct a weighted logistic regression model, which accounts for the unbalanced classes in our dataset. The weighting procedure for this model is detailed in Appendix A.1. Significant coefficient values for this model are displayed in Appendix A.3. This model has an accuracy of 95%, but a sensitivity of 98%, suggesting that incorporating class weights allows us to correctly identify revolutions more often, which is more important in the context of this project.

After constructing both logistic regression models, we proceed constructing a random forest model. While this model is less interpretable than the logistic regression models, the random forest does provide us with a measure of the relative importance of each variable in making our predictions. The variable importance plot for our random forest model is presented in Appendix A.3. We observe that the variables measuring GDP growth, a nation's openness, and the log of debts to credits are the most important for making predictions in this model, which confirms the results from our logistic regression model. The random forest model has an accuracy of 97%, coupled with a sensitivity of 87%, which speaks to the predictive ability of this model.

Before moving to HMMs, we first want to assess our ability to make predictions about future instability from a previous year's data. The process for 'stepping-back' our predictions is explained in Appendix A.2. We used a weighted logistic regression model to complete this task, measuring the accuracy and sensitivity for each year in our set-back model, which is displayed below.

Years Held Back	Accuracy	Sensitivity
0	96%	96%
1	79%	92%
2	75%	92%
3	74%	91%
4	74%	91%
5	74%	91%

Table 1: Accuracy and Sensitivity Values for Held-Back Logistic Regression

While the accuracy decreased as we increased the years in our set-back model, the sensitivity held level. This means that we are able to predict future political instability from a current year's economic measurements (a high sensitivity value reflecting the ability to identify states undergoing revolutions accurately), which supports exploring the use of time-series modeling.

4.3 Hidden Markov Models

When working with time-series data, we must first understand that measurements of political instability should not be considered static. In any country, political instability does not just depend on the current economic conditions, but also previous economic and

political conditions. Revolutions do not happen overnight - they are caused by longer-term trends within a country. Because of this, we suggest utilizing the HMM, which relies on the assumption that the current information (economic and/or political conditions in the current year) depends on prior data (conditions in previous years). We implement our HMM in `RStudio` using the `momentuHMM` package. For the purpose of this project, we construct individual models for a select few countries in our dataset (identified based on the Unsupervised Learning in Section 4.1) which we believe to be representative of the data as a whole. This model takes in our economic and government measurements as covariates, and predicts the presence of a revolution as an outcome variable. We constructed multiple HMMs for multiple countries in our dataset: Afghanistan, Cameroon, Russia, and Mexico. These countries were chosen based on the different levels of political instability present, so we could understand our model performance when faced with a variety of socioeconomic conditions. A graphical explanation of the HMM framework is provided in Appendix A.4

5 Analysis

5.1 Establishing our Baseline Model

Due to the overwhelming size of our dataset, we constructed our baseline model using a small number of our covariates: GDP growth, government openness, debt to GDP ratio, executive constraints, corruption, elections, Gini coefficient, beginning of a banking crisis, ethnic tensions, and socioeconomic conditions. Without reducing the number of covariates in our dataset, our models would be too computationally complex. Once our models have been built, we are able to evaluate their predictive ability and identify important variables.

5.2 Model Predictive Ability and Covariates

After constructing our models, we work to predict the occurrence of a revolution in a given country each year. Because we constructed our models using the `momentuHMM` package in `R`, we are able to do this using the built-in functionality provided by the package. We then compare the predicted results with the true data, and receive the following accuracy values:

Country	Overall Accuracy	Percent of Revolutions Identified
Afghanistan	78%	55%
Cameroon	72%	50%
Russia	61%	33%
Mexico	68%	25%

Table 2: Accuracy of our Null Model

When we compare this to the accuracy values for our logistic regression and random forest models, we observe that our HMM performs worse than both of these models. Additionally, we observe that our HMM performs worse than our set-back model when predicting revolutions. Therefore, we can conclude that the HMM would need additional tuning if we wish to be competitive.

Additionally, we are provided with a 95% confidence interval for the coefficients in our model. When we assess our covariates, we observe that the variables measuring GDP growth, the Gini coefficient, and elections are the most relevant for making predictions.

5.3 Increasing Complexity I: Moving Beyond the Null Model

Now that we have our HMM, we are able to increase the complexity in order to improve its performance. Our first step towards increasing the complexity of our model will involve an additional round of optimization for our parameter values, which should increase our accuracy. Using the `momentuHMM` package, this is a simple process, where we initialize our model the same way as before. The only difference is that for our initial coefficient values, we input our final coefficient values from the previous section. The performance of our results is displayed below:

Country	Overall Accuracy	Percent of Revolutions Identified
Afghanistan	81%	75%
Cameroon	75%	100%
Russia	67%	66%
Mexico	72%	55%

Table 3: Accuracy of our Improved HMM

This model performs better than our null model, but more work is necessary if we are to yield any relevant conclusions from it. While we could re-optimize our model until we reach a satisfactory accuracy level, this process is computationally complex and will be left for anyone who wishes to explore this data in the future.

5.4 Increasing Complexity II: Additional Covariates

Now that we have a more complex model, we are in a position to increase the number of covariates that we are using to predict the presence of a revolution. We provide additional data measuring geographic region, internal conflicts, socioeconomic conditions, and polarization for our new model. The performance of this model is displayed below:

Country	Overall Accuracy	Percent of Revolutions Identified
Afghanistan	87%	82%
Cameroon	81%	100%
Russia	71%	75%
Mexico	79%	66%

Table 4: Accuracy of our HMM with Additional Covariates

This model performs much better than our previous two HMMs, but struggles to reach the same performance levels as any of our logistic regression or random forest models. This suggests that it may be difficult to predict revolutions using time-series data (at our current level of understanding and computational power).

When we assess the 95% confidence interval for the coefficients in our model, we observe that the same variables we used in our null model are relevant when making predictions. This is similar from the variable importance we obtained from our logistic regression and random forest models, which speaks to the importance of these variables in predicting revolutions.

6 Results

6.1 Validation and Model Strength

After constructing our predictive models, we assess their performance on held-out data in our dataset. For our logistic regression and random forest models, this was 20% of our data which was randomly taken from our entire dataset. For our HMMs, this was the last 20% (ten years) of data for each model. The results are displayed below.

Model	Validation Accuracy
Logistic Regression	92%
Random Forest	94%
HMM - Afghanistan	82%
HMM - Cameroon	78%
HMM - Russia	66%
HMM - Mexico	75%

Table 5: Validation Accuracy for Our Models

While our validation accuracy is lower than the accuracy for our training data, this is to be expected. Overall, our models were able to perform well with entirely new data, which means they can be used in future settings.

6.2 Model Interpretability

While overall accuracy is important when developing statistical models, it is necessary for us to identify important variables for predicting revolutions in a specific country. Fortunately, this is a simple process for our logistic regression and random forest models models, which can provide us with coefficient values (logistic regression) or variable importance plots (random forest) for our convenience. The top ten variables for each model are displayed below.

Covariate	Log-Odds Coefficient Value
Government Religion: Christian	38.8
Government Religion: Islamic	6.05
Credit Bubble	1.83
Region: Arab States	1.39
Start of Banking Crisis	1.37
Gini Coefficient	1.33
Region: Middle East	1.19
Growth	1.24
Polarization	1.05
Socioeconomic Conditions	1.04

Table 6: Coefficient Values for Weighted Logistic Regression

Covariate	Variable Importance Score
Log of Debt to GDP Ratio	16.98
Gini Coefficient	14.32
Internal Conflict	16.60
Region	10.81
Socioeconomic Conditions	10.13
Polarization	7.56
Growth	7.36
Elections	4.05
Start of Banking Crisis	3.08
Credit Bubble	2.02

Table 7: Variable Importance Scores for Random Forest

As we observe above, there is some overlap as to which covariates are deemed important in both of our models. We observe that many economic metrics rank in the top ten for both models, as do variables measuring socioeconomic conditions, geography, and government structure. This will be discussed in depth in Section 7.

7 Conclusion

This thesis was undertaken with the goal of furthering our understanding of revolutions across the globe, as well as the factors which can influence the likelihood of one. Over the course of this project, we have conducted both supervised and unsupervised learning, in hopes of identifying which variables (if any) can be used to predict a revolution. We have identified multiple factors which appear to increase the likelihood of a revolution in a country, some of which are highlighted below.

7.1 Geographic Region

Geographic region plays an important role in predicting political instability - it was one of the most important variables in both our stationary models, and with good reason. Throughout the past half century, certain regions (especially South America, Africa, and the Middle East) have been more politically unstable than the rest of the world, and continue to face turmoil to this day. Anyone who wishes to explore the relationship between economic conditions and political instability further should account for these region-level effects.

7.2 Election Year

In many developing democracies, an election year is a danger point. Various parties who are unsatisfied with election results may chose to disregard the results of the democratic process and attempt to seize power for themselves. This is evident if we look at countries such as Myanmar (which experienced a military coup in 2020 following their losses in the national legislature), Honduras (the President was kidnapped by the military in 2009), and even the United States, when supporters of President Donald Trump attempted to overturn the results of a democratic election.

7.3 Gini Coefficient

The Gini coefficient, a measure of global wealth inequality, is a strong measure of predicting political instability. Countries with higher levels of wealth inequality are more likely to suffer from political instability, as citizens will become frustrated with the wealth disparities they are facing and may rise up against their rulers. Researchers at the University of Chicago found that a 1-standard-deviation increase in the Gini coefficient significantly increases support for any potential revolution, a sign of how people may have a preference for revolt when income inequality is high. This was exemplified in my thesis, where high income inequality was a strong predictor of a revolution.

7.4 GDP Growth

GDP growth, a measure of how much a countries economy grows each year, is a strong measurement of political instability. Countries with slow growth (or even negative growth) will struggle to provide resources to their citizens, who may wish for a change in their government, especially if it is more autocratic. Many dictatorships in our data (e.g, Russia and China) provide their citizens with robust GDP growth, which may reduce the likelihood of these citizens revolting.

7.5 Religion

As we observed in Table 6 (Section 6.2), religion in government (which was stored as a factor variable in our data) is a strong indicator of revolutions. Often, non-secular governments will force religion beliefs on a (sometimes heterogeneous) population, which may resent this. This hypothesis is supported by research done which ties religious forces to unrest in the Middle East (revolutions in this region were often aided by Western Powers, which increased their severity and frequency), as well as civil unrest in Europe (e.g., the IRA was an organization which was founded based in Protestant oppression of Catholics in Northern Ireland). While religion and government are two forces which are unlikely to ever be separated, it is clear that this is a strong predictor of unrest.

7.6 Takeaways

We can conclude that there are many factors that can be used to predict revolutions, but the relationship is quite nuanced. While we are able to predict a revolution with high accuracy, we have only scratched the surface of understanding this relationship. We have identified certain factors, both geographic and economic, which are related to the onset of a revolution, which could be used by any political scientists or statisticians who wish to continue this work.

8 Future Work

While this thesis has advanced our understanding of the causes of revolutions, there is still much work that could be done to build on what we have started here. First, one could implement more advanced statistical models, which are outside of the scope of this thesis. This would allow us to make more accurate predictions, and would provide us with a more nuanced understanding of the relative importance of our variables. Second, working on

an interdisciplinary team with a combination of statisticians and political scientists would allow for a more in-depth understanding of the data and relevant variables, which would accelerate the pace at which discoveries are made.

9 Acknowledgments

The author would like to thank his thesis advisor, Professor Freeman, for his continuous support throughout this process. The author would also like to thank his family (particularly his parents) who worked as a sounding board for ideas and offered words of encouragement whenever necessary, and his girlfriend, Emily Zhang, who helped motivate him to meet deadlines and maintain his enthusiasm throughout this process. Finally, the author would like to extend his profound gratitude to Daniel Hansen of the Institute for Politics and Strategy for providing him with data and resources to conduct this research.

A Appendix I

A.1 Model Weighting Procedure

In order to correct any problems stemming from unbalanced classes in our dataset, we apply the following formulas to calculate weights for our outcome measures:

$$\text{Positive Weight: } \frac{1}{\frac{Number\ of\ data\ points\ with\ revolution}{Number\ of\ data\ points}}$$

$$\text{Negative Weight: } \frac{1}{\frac{Number\ of\ data\ points\ with\ no\ revolution}{Number\ of\ data\ points}}$$

Positive weight is assigned to all data where there is an active revolution, and negative weight is assigned to all data where there is no active revolution. This procedure allows us to remedy any problems stemming from unbalanced classes, as our model will now consider revolutions to be more important than non-revolutions.

A.2 Set-Back Model Explanation

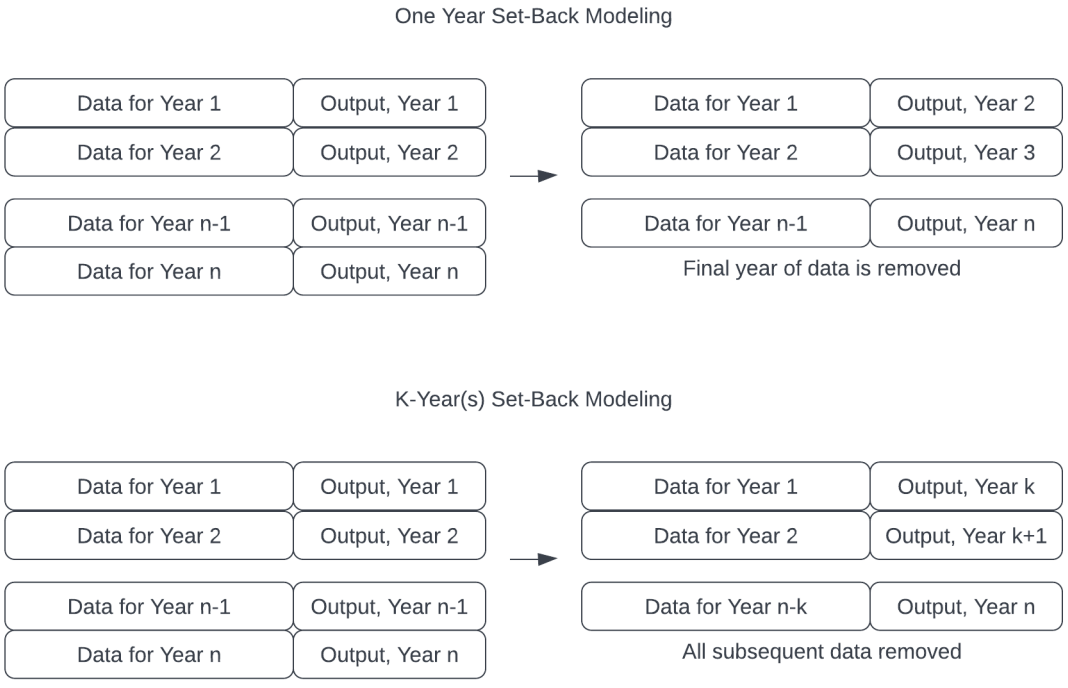

Figure 3: Diagram of Set-Back Modeling Procedure

A.3 Model Summaries

Covariate	Log-Odds Coefficient Value
Government Religion: Christian	60.52
Gini Coefficient	23.2
Government Religion: Islamic	13.1
Region: Arab States	8.82
Start of Banking Crisis	2.1
Credit Bubble	1.49
Region: Middle East	1.32
Growth	1.21
Polarization	1.13
Socioeconomic Conditions	1.07

Table 8: Coefficient Values for Logistic Regression

Covariate	Log-Odds Coefficient Value
Government Religion: Christian	38.8
Government Religion: Islamic	6.05
Credit Bubble	1.83
Region: Arab States	1.39
Start of Banking Crisis	1.37
Gini Coefficient	1.33
Region: Middle East	1.19
Growth	1.24
Polarization	1.05
Socioeconomic Conditions	1.04

Table 9: Coefficient Values for Weighted Logistic Regression

Covariate	Variable Importance Score
Log of Debt to GDP	16.98
Gini Coefficient	14.32
Internal Conflict	16.60
Region	10.81
Socioeconomic Conditions	10.13
Polarization	7.56
Growth	7.36
Elections	4.05
Start of Banking Crisis	3.08
Credit Bubble	2.02

Table 10: Variable Importance Scores for Random Forest

A.4 Hidden Markov Model Explanation

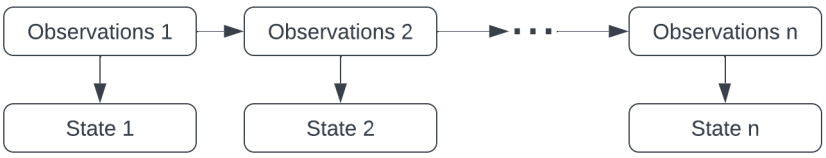

Figure 4: Diagram of Hidden Markov Model

Hidden Markov Models are probabilistic models which allow us to compute the joint probability of a set of hidden states (economic and political measurements) and observed states (whether there is a revolution or not). Matrices measuring the state transmission probabilities (probability of going from a revolution to either a revolution or no revolution, and vice versa) and emission probabilities (probability distributions for each of our covariates) are created during the model optimization process.

One our model is optimized, we can go through our data year by year to determine a probability value for our dependent variable's value at any given year, and can develop a predicted sequence of dependent variables, using the following math:

$$\text{Transition Probability: } p(x_t|x_{t-1})$$
$$\text{Observation probabilities: } p(y_t|x_t)$$

We can calculate the joint probability of X (our observations) and Y (our states) using the following formula:

$$P(X,Y) = p(x_1)\prod_{t=1}^{n-1} p(x_{t+1}|x_t)\prod_{t`=1}^{n} p(x_{t`}|x_t`)$$

Finally, we are left with probability values for $P(Y)$, which we can use to predict Y values (based on whether $P(Y)$ is greater than or less than 0.5).

References

[1] Alesina, A., Ozler, S., Roubini, N., &; Swagel, P. (1992). Political instability and economic growth. *Journal of Economic Growth*, 1(2), 189–211. https://doi.org/10.3386/w4173

[2] Bodea, C., Elbadawi, I., &; Houle, C. (2016). Do civil wars, coups and riots have the same structural determinants? *International Interactions*, 43(3), 537–561. https://doi.org/10.1080/03050629.2016.1188093

[3] Cross-national time-series data archive. (2015). *Choice Reviews Online*, 52(07). https://doi.org/10.5860/choice.187502

[4] Duncan, G. T., &; Siverson, R. M. (1975). Markov chain models for Conflict Analysis: Results from Sino-Indian relations, 1959-1964.*International Studies Quarterly*, 19(3), 344. https://doi.org/10.2307/2600315

[5] *Hidden markov models for regime detection using R.* QuantStart. (n.d.). Retrieved April 27, 2022, from https://www.quantstart.com/articles/hidden-markov-models-for-regime-detection-using-r/

[6] *ICRG methodology - PRS Group.* (n.d.). Retrieved April 27, 2022, from https://www.prsgroup.com/wp-content/uploads/2014/08/icrgmethodology.pdf

[7] Lipscy, P. Y. (2017). Democracy and financial crisis. *SSRN Electronic Journal*, 72(4), 937–968. https://doi.org/10.2139/ssrn.1900126

[8] MacCulloch, R. (2005). Income inequality and the taste for revolution. *The Journal of Law and Economics*, 48(1), 93–123. https://doi.org/10.1086/426881

[9] Qian, N., Nunn, N., &; Wen, J. (2019, May 10). *Why economic crises trigger political turnover in some countries but not others.*Kellogg Insight. Retrieved April 27, 2022, from https://insight.kellogg.northwestern.edu/article/why-economic-crises-trigger-political-turnover-in-some-countries-but-not-others

[10] Qiao, F., Li, P., Zhang, X., Ding, Z., Cheng, J., &; Wang, H. (2017). Predicting social unrest events with Hidden Markov models using GDELT. *Discrete Dynamics in Nature and Society*, 2017, 1–13. https://doi.org/10.1155/2017/8180272

[11] Scartascini, C., Cruz, C., &; Keefer, P. (2018). The database of Political Institutions 2017 (DPI2017). https://doi.org/10.18235/0001027

[12] Stack Exchange. (1961, June 1). *Time-series machine learning methods and R packages.* Cross Validated. Retrieved April 27, 2022, from https://stats.stackexchange.com/questions/71350/time-series-machine-learning-methods-and-r-packages