Analysis of Factors Influencing Annual Unemployment Rate in the United States

Abstract: This paper investigates factors influencing annual US unemployment rate, with a specific interest in its relationship with civil unrest. Using data from various government and public databases, we fit a multiple linear regression model with an R^2_{adj} value of 0.7549. We found that aggregate stock market value, national debt, number of major incidents of civil unrest, and the interaction between three or more instances of civil unrest annually and percentage of news articles in US media covering protests were significant predictors of annual unemployment rate. We hypothesize that stock market value and national debt are strong predictors because low debt and high stock market value indicate a strong economy with high employment. Also, although more instances of civil unrest was individually associated with higher unemployment, possibly due to discontent stemming from the lack of jobs, increasing percentage of protest coverage was counterintuitively associated with a larger *decrease* in unemployment in years with three or more instances of civil unrest than in years with no instances of civil unrest. Further research could explore this finding or break down employment by demographics to examine how various protest motivations can impact groups differently.

Background and Introduction:

The Federal Reserve's primary mandate is to keep both inflation and unemployment rates low, where the unemployment rate is defined as the percentage of unemployed workers who are able and willing to work. The successful pursuit of this mandate results in consistent product prices and full employment, optimizing the economy. However, depending on the unemployment rate, the best approach could vary year to year and require different actions including changes to the interest rate, money supply, or buying/selling treasury bonds (Amadeo). Thus, we're interested in predicting annual unemployment rate through various predictors. We are particularly interested in how this decade's rise in civil unrest and protests have affected unemployment (Tharoor). By improving the forecasting of unemployment rates, the monetary policies implemented can be more impactful on the economy. This is especially important today, when recent protests and job losses due to the COVID-19 pandemic have created an economic crisis (Iacurci).

Data and Exploratory Analysis:

Data and Variables:

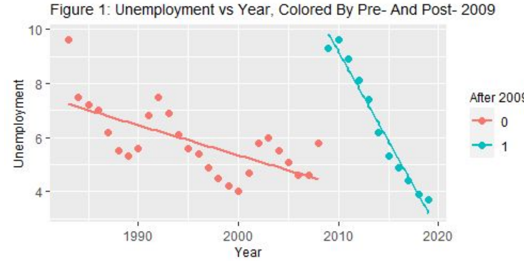
Our original data set included 49 observations, each corresponding to a year between 1971 and 2019, inclusive, and 11 variables: annual unemployment rate, stock market index, GDP, debt, inflation, energy prices of oil, gas, and electric, election year, protest intensity, and incidents of civil unrest. Our response variable is the annual US unemployment rate, calculated by taking the average of the monthly unemployment rates measured by the Bureau of Labor Statistics.[3] All variables with the exception of election year are quantitative.

The stock market index predictor variable represents the total monetary value of US equity markets, as defined by the December market cap of the Wilshire 5000 Total Market Index,[10] which combines the weighted performance of approximately 3500 publicly traded companies. The values in our data set are in billions of US dollars. The quarterly GDP in billions of chained 2012 dollars is from the Bureau of Economic Analysis.[6] For our purposes, we recorded just the annual Q4 data. Our aggregate national debt variable is also measured in billions of US dollars and is taken from the US Treasury.[9] The US inflation rate is measured in percentages and pulled from the Federal Reserve Bank of Minneapolis.[2] Energy prices are taken from the US Energy Information Administration.[7] We recorded annual electricity, oil, and gas prices, in cents per kilowatt hour, dollars per barrel, and dollars per thousand cubic feet, respectively. Election year is the only categorical variable collected as such and is binary, with 1 indicating that the year was an election year and 0 indicating that it was not.[8]

We were particularly interested in the effect of civil unrest on unemployment. The two predictor variables we found with respect to this focus were civil unrest and protest intensity. The first was collected by counting the number of recorded major incidents of civil unrest per year found through a Wikipedia article[1] listing incidents of major civil unrest in the US. The latter was pulled from the GDELT Project,[5] which measured the percentage of articles in the US news of each year that covered protests. Protest intensity was the only variable with missing values between 1971 and 2019, with no data before 1979. Because we chose to include protest intensity in our final model, we omitted data from 1971 through 1978, and our final analysis was based on 41 sampling units instead of 49.

Exploratory Data Analysis:

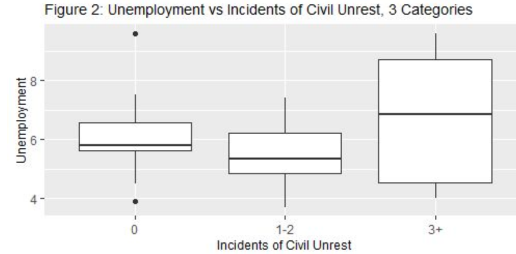
Figure 1: Unemployment vs Year, Colored By Pre- And Post- 2009



Our initial unemployment vs. year plot shows that the marginal relationship between year and unemployment, indicated by the slope, differs a lot across business cycles. Thus we added an indicator variable splitting the data into "before 2009" and "after 2009", as shown in Figure 1 to the right. Here, both time intervals have a negative relationship with unemployment, indicating that in this period of 41 years, more recent years are associated with lower levels of unemployment.

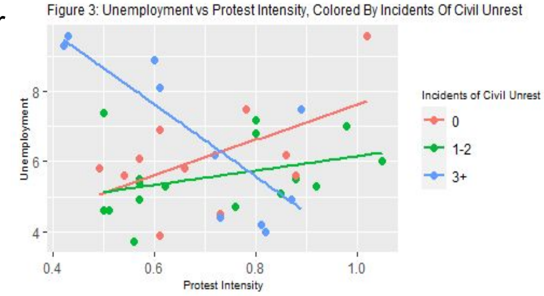
To study the relationship between civil unrest and unemployment rates, we plotted incidents by unemployment and found few years that saw 5 or more instances of civil unrest. As a result, we decided to make this predictor a categorical variable, with the categories being 0, 1-2, and 3 or more incidents of civil unrest, resulting in mostly uniform number of observations. This is plotted in the side-by-side boxplots in Figure 2 to the right. The spread of the distributions vary, and the centers of the distributions, marking the median unemployment rate associated with a each number of cases, first decrease then increase as the instances of civil unrest increase.

Figure 2: Unemployment vs Incidents of Civil Unrest, 3 Categories



Another variable that we used to study civil unrest was the intensity of protests. Initially, protest intensity didn't appear to have any relationship with the unemployment rate. However, in Figure 3 to the right, which groups by the three categories of civil unrest, we see much clearer relationships between protest intensity and unemployment. This indicates that an interaction variable between protest intensity and the number of instances of major civil unrest might be appropriate for explaining the variation in unemployment rates. With this interaction taken into account, it appears that when the number of cases of civil unrest in a year are between 0 and 2, higher protest intensity is associated with higher levels of unemployment, but when there are more than 3 cases of civil unrest in a year, higher protest intensity is associated with lower levels of unemployment.

Figure 3: Unemployment vs Protest Intensity, Colored By Incidents Of Civil Unrest



Model and Results:

Analytic Methods

A multiple linear regression was used to model annual US unemployment rate versus our various quantitative and categorical predictor variables. After examining relationships among the predictors, we decided against using each of the three energy prices as predictors, only keeping oil prices, as they were all heavily collinear with correlation coefficients above 0.8. The oil price variable was retained because it resulted in the highest R^2_{adj} in the initial model.

We fit our initial model using the remaining predictors as well as their interactions with incidents of civil unrest and the before/after 2009 indicator variable (Appendix A). This resulted in an initial R^2_{adj} value of 0.71. Diagnostics indicated violations of normality and constant variance, which we addressed by applying a Box-Cox transformation of -2 on the response variable, *Unemployment Rate*$^{-2}$ (Appendix B). Stepwise selection and best subset selection were then used in an attempt to obtain a simpler model, but this proved to be more challenging than expected. The "best" models calculated by R all had ten or more variables in them, making interpretability a challenge and often including interactions between variables without the variables themselves being in the model. To work around this, we decided to take the best model from best subset selection (Appendix C*)* and manually perform nested F-tests. Our primary aim was to retain variables that would be informative, significant, and account for the most variance, while also maintaining interpretability and keeping the goal of the project in mind.

Final Model and Results

Our final model predicts *Unemployment Rate*$^{-2}$ using *SMI, Debt, Protest Intensity, Instances of Civil Unrest,* and the interaction between *Protest Intensity* and *Instances of Civil Unrest*. At a significance level of $\alpha = 0.05$, every predictor is significant with the exception of *Protest Intensity* and the interaction between *Protest Intensity* and *1 or 2 instances of civil unrest*. See Appendix D for the summary table.

The model has an F-statistic of 18.6 with a p-value of 8.845e-10, indicating that the full model is effective as compared to the intercept-only model. The final R^2_{adj} value is 0.7549, reflecting the number of predictors in the model and the amount of variability in *Unemployment Rate*$^{-2}$ explained by the model. Our model diagnostics show no severe violations (see appendix F for details). However, there are minor issues that will be further discussed in the conclusion.

The table in Appendix E displays 95% confidence intervals for the slopes of each predictor. Some unsurprising conclusions can be inferred here, such as increasing SMI being associated with increasing *Unemployment Rate*$^{-2}$ (decreasing regular unemployment rate), and increasing debt being associated with decreasing *Unemployment Rate*$^{-2}$ (increasing regular unemployment rate), holding other variables constant. However, an unexpected but interesting conclusion that can be inferred here is that an increase in protest intensity given there are three or more instances of major civil unrest is associated with a greater increase in *Unemployment Rate*$^{-2}$, or decrease in regular unemployment, than years with no instances of civil unrest.

Conclusions and Discussion:

The goal of this project was to investigate the best predictors of annual US unemployment rate, focusing on the relationship between annual US unemployment rate and civil unrest. The aggregate value of the stock market, national debt, number of major incidents of civil unrest, protest intensity as measured by percentage of articles in the US media each year that covered protests, and the interaction between the latter two were important in predicting the US annual unemployment rate. Of these predictors, only protest intensity wasn't statistically significant given other variables.

Stock market value and national debt are strong predictors of unemployment rate because high stock market value and low national debt will likely create more available jobs. Civil unrest could be explained by high unemployment rates causing outrage amongst people, especially if the burden of high unemployment rates are not evenly distributed amongst socioeconomic classes. Moreover, increases in civil unrest and percentage of news covering protests are both individually associated with increases in unemployment. Yet the interaction between three or more instances of civil unrest and percentage of news coverage is associated with a larger decrease in unemployment than a year with no instances of civil unrest, a counterintuitive finding that might be interesting to research further.

There are several limitations to this study. In Figure 4 to the right, the aggregate stock market value and debt are highly multicollinear. Nonetheless, because each variable explains a unique aspect of the variation, removing either variable will significantly impact the R^2_{adj}. Furthermore, the best model by best subset selection was too complex. To simplify the model, we adjusted this "best" model and used a series of nested F-tests to methodically remove the variables that explained little unique variation. Consequently, the final model had an R^2_{adj} of 0.7549, which is lower than the best subsetted model's R^2_{adj} of 0.8789. There are also minor issues with model diagnostics. On the plots of residuals vs. debt and residuals vs. SMI, two clusters of points are present. Grouping by before/after 2009, the points become clearly separated (see Appendix F). Lastly, the response variable, rate, has restrictions on the values that can be taken. This is an unavoidable limitation.

Figure 4: Pairs Plot for Final Model

Further research could separate the data by economic cycle, allowing us to fix the previously stated grouping before/after 2009 limitation. Additional research could also investigate unemployment by various demographics. By breaking down unemployment, we can research which demographics are most impacted by various instances of civil unrest, and look into which protests are associated with the largest changes in annual unemployment rate.

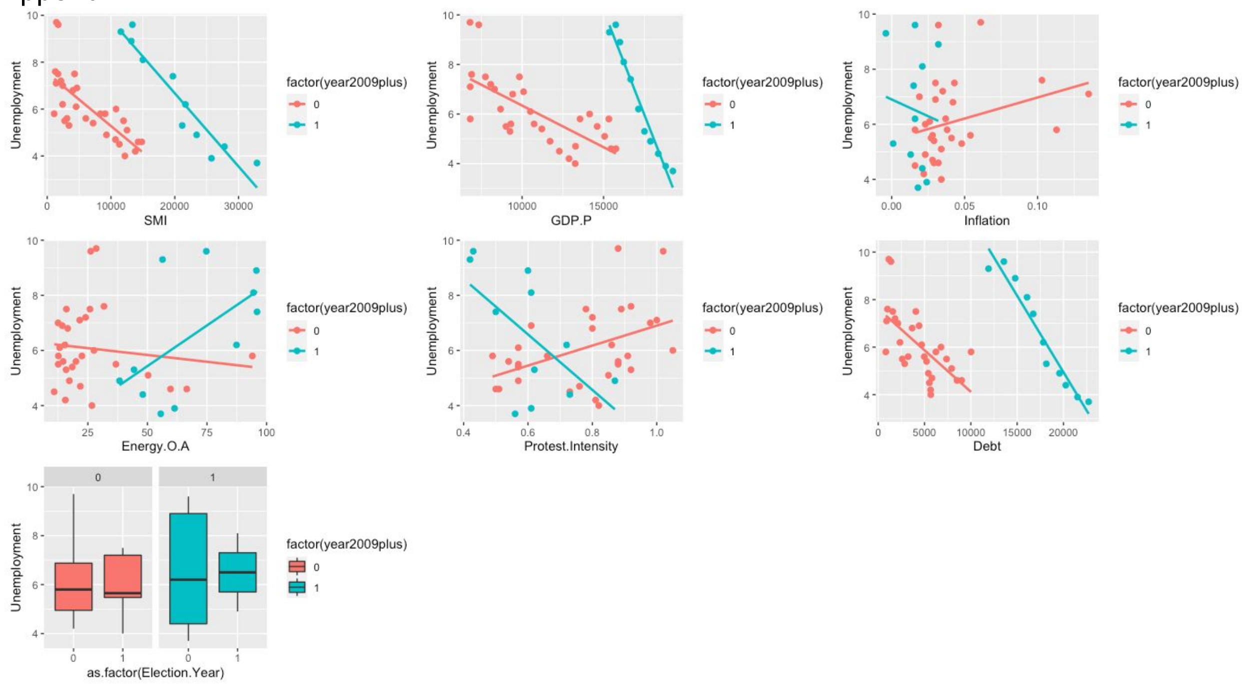
References

Background:
1. Amadeo, Kimberly. “Unemployment Rate, Effect, and Trends.” *the balance*, DotDash, 21 11 2020, https://www.thebalance.com/unemployment-rate-3305744.
2. Iacurci, Greg. “Here's why the unemployment rate is so important.” *CNBC*, NBCUniversal, 5 6 2020, https://www.cnbc.com/2020/06/05/heres-what-unemployment-rate-actually-means-and-why-its-important.html.
3. Tharoor, Ishaan. “A year of protests caps a decade of crisis and anger.” *The Washington Post*, 20 12 2019, https://www.washingtonpost.com/world/2019/12/20/year-protests-caps-decade-crisis-anger/.

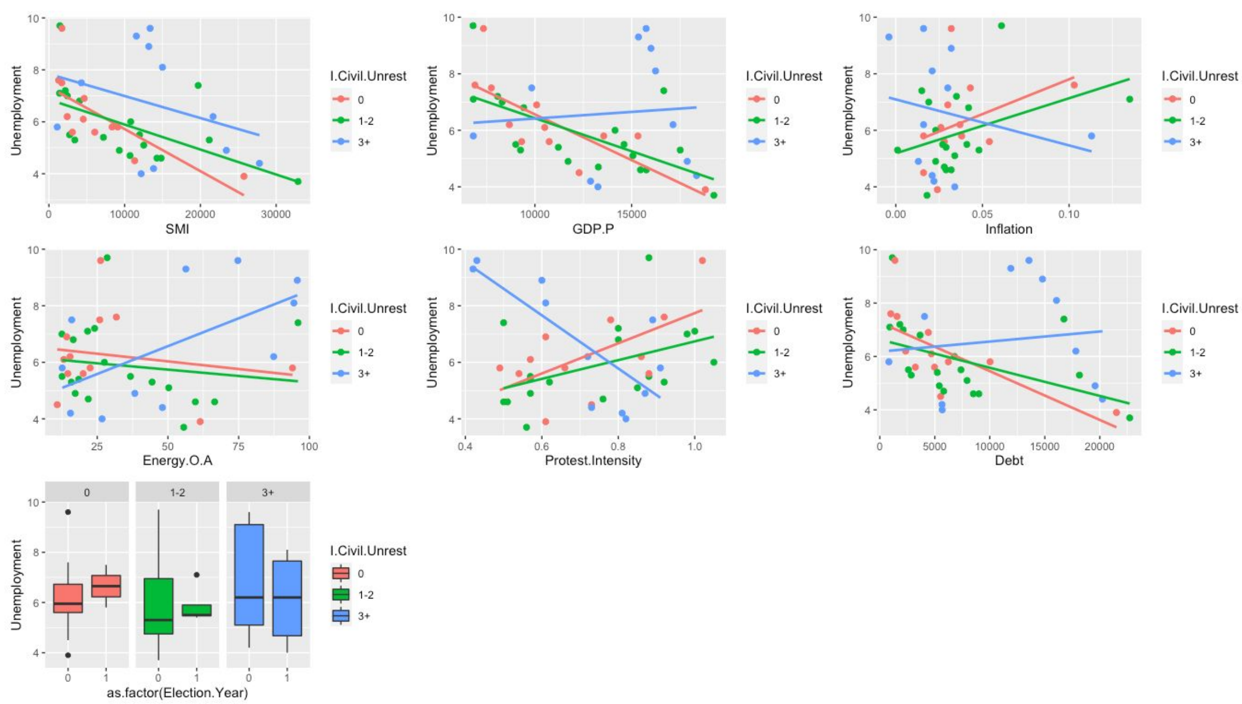
Data Sources:
1. “Cases of Civil Unrest in the United States.” *Wikipedia*, Media Wiki, 2020, https://en.wikipedia.org/wiki/List_of_incidents_of_civil_unrest_in_the_United_States.
2. “Consumer Price Index, 1913-.” *Federal Reserve Bank of Minneapolis*, Federal Reserve Bank, 2019, https://www.minneapolisfed.org/about-us/monetary-policy/inflation-calculator/consumer-price-index-1913-.
3. “Labor Force Statistics from the Current Population Survey.” *US Bureau of Labor Statistics*, United States Department of Labor, 2020, https://data.bls.gov/pdq/SurveyOutputServlet.
4. “Labor Force Statistics from the Current Population Survey.” *Bureau of Labor Statistics*, US Department of Labor, 2020. https://www.bls.gov/news.release/union2.toc.htm
5. “Mapping Global Protest Trends from 1979 - 2019.” *The Official GDELT Project Blog*, GDELT Project, 24 11 2019, https://blog.gdeltproject.org/mapping-global-protest-trends-1979-2019-through-one-billion-news-articles/.
6. “National Income and Product Accounts.” *United States Bureau of Economic Analysis*, United States Government, 2020, https://apps.bea.gov/iTable/iTable.cfm?reqid=19&step=2#reqid=19&step=2&isuri=1&1921=survey.
7. “Total Energy.” *US Energy Information Administration*, USEIA, 2020, https://www.eia.gov/totalenergy/data/browser/index.php?tbl=T09.08#/?f=A&start=1960&end=2019&charted=0-1-2-3-4.
8. “United States Presidential Elections.” *Wikipedia*, Media Wiki, 2020, https://en.wikipedia.org/wiki/United_States_presidential_election.
9. “US Historical Debt Outstanding - Annual.” *Treasury Direct*, United States Treasury, 2020, https://www.treasurydirect.gov/govt/reports/pd/histdebt/histdebt.htm.
10. “Wilshire 5000 Total Market Full Cap Index Historical Values.” *Investing.com*, Fusion Media Limited, 2020, https://www.investing.com/indices/wilshire-5000-total-market-historical-data.

Appendix

Appendix A



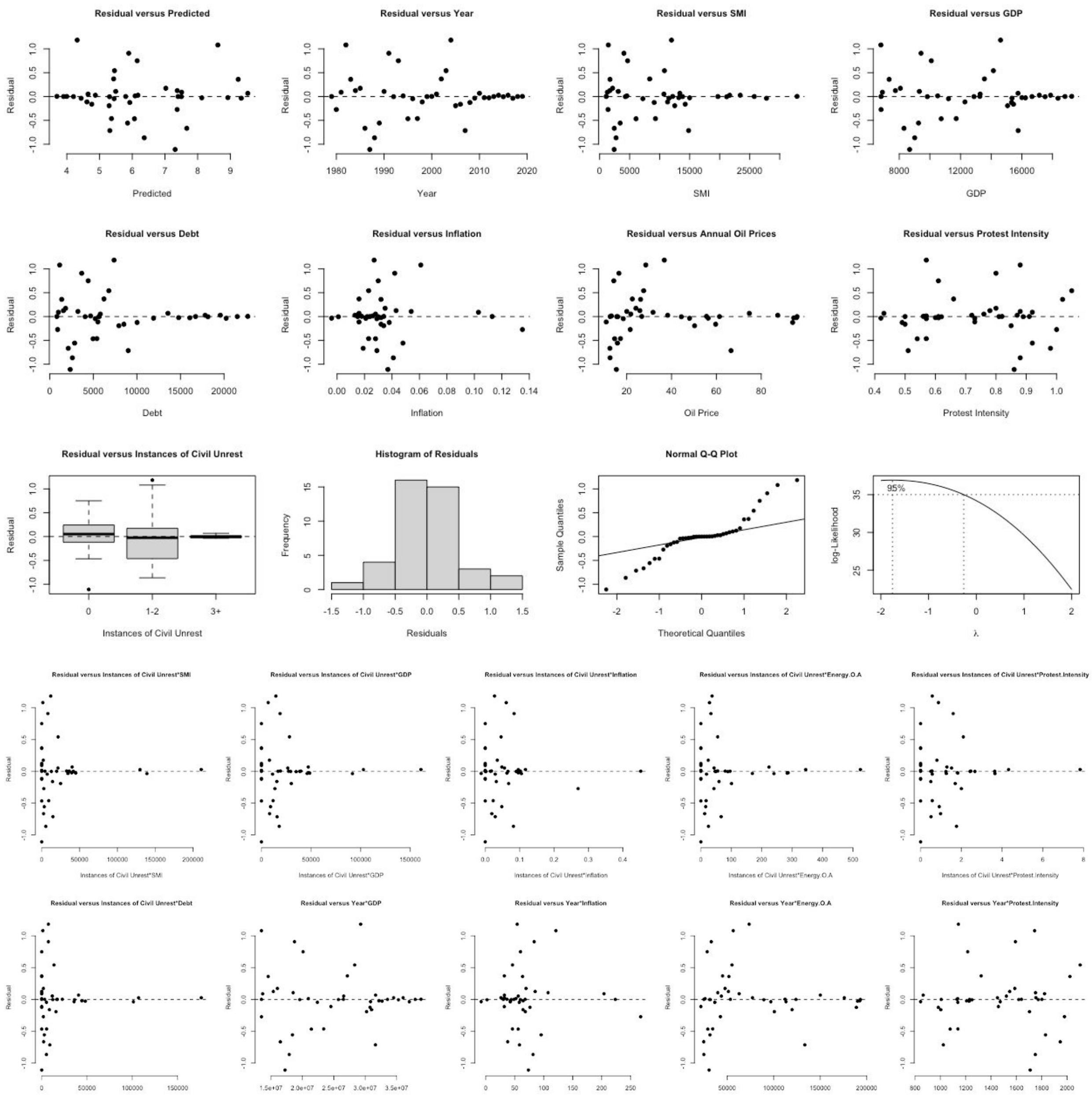
The non-parallel lines indicate possible interactions between before/after 2009 and GDP, inflation, oil prices, protest intensity, and election year. All were included in the initial model.



The non-parallel lines indicate possible interactions between instances of civil unrest and every other predictor. All were included in the initial model.

Appendix B

Initial model diagnostics



Model diagnostics show some concerning patterns regarding constant variance for some of the variables, as well as normality based on the QQ-plot. We applied a Box-Cox transformation of -2 to combat these concerns, and although there were still some issues regarding constant variance afterward, the separate clusters of points are clearly split by before and after 2009 (which is something we address later in the paper when discussing our final model).

Appendix C

	model	p	rsq	rss	adjr2	cp	bic	stderr
1	SMI	2	0.3488357	0.0065032327	0.3321392	101.8886568	-10.16158	0.012913154
2	SMI-E.O.A:2	3	0.7552344	0.0024444943	0.7423520	17.2067325	-46.56491	0.008020524
3	SMI-En.O.A-y	4	0.8099409	0.0018981359	0.7945307	7.5382311	-53.22295	0.007162470
4	SMI-Db-E.O.A:I.C.U1-D:I.C.U3	5	0.8383423	0.0016144890	0.8203803	3.4804213	-56.14538	0.006696784
5	SMI-In-En.O.A-E.O.A:I.C.U1-P.I:2	6	0.8546950	0.0014511731	0.8339372	1.9925066	-56.80430	0.006439106
6	SMI-In-En.O.A-E.O.A:I.C.U1-D:I.C.U3-GDP.P:2	7	0.8647989	0.0013502649	0.8409399	1.8374248	-56.04566	0.006301879
7	SMI-In-En.O.A-y-GDP.P:I.C.U3-E.O.A:I.C.U1-D:I.C.U3	8	0.8769707	0.0012287043	0.8508735	1.2412711	-56.20005	0.006101923
8	SMI-En.O.A-I.C.U3-y-GDP.P:I.C.U3-I:I.C.U3-D:I.C.U1-D:I.C.U3	9	0.8908988	0.0010896029	0.8636235	0.2705006	-57.41251	0.005835245
9	SMI-En.O.A-I.C.U3-y-GDP.P:I.C.U3-I:I.C.U1-I:I.C.U3-E.O.A:I.C.U1-D:I.C.U3	10	0.8974140	0.0010245354	0.8676309	0.8808635	-56.22347	0.005748872
10	SMI-En.O.A-I.C.U3-y-SMI:I.C.U3-GDP.P:I.C.U3-I:I.C.U3-D:I.C.U1-D:I.C.U3-GDP.P:2	11	0.9058608	0.0009401761	0.8744811	1.0792138	-56.03292	0.005598143
11	SMI-In-En.O.A-I.C.U3-y-SMI:I.C.U3-GDP.P:I.C.U3-I:I.C.U3-D:I.C.U1-D:I.C.U3-GDP.P:2	12	0.9094768	0.0009040625	0.8751405	2.3079400	-53.92526	0.005583419
12	SMI-En.O.A-I.C.U1-I.C.U3-y-SMI:I.C.U3-GDP.P:I.C.U3-I:I.C.U3-P.I:I.C.U1-D:I.C.U1-D:I.C.U3-GDP.P:2	13	0.9148288	0.0008506121	0.8783268	3.1664064	-52.71033	0.005511715
13	SMI-En.O.A-I.C.U1-I.C.U3-y-SMI:I.C.U3-GDP.P:I.C.U3-I:I.C.U3-P.I:I.C.U1-D:I.C.U1-D:I.C.U3-GDP.P:2-I:2	14	0.9182840	0.0008161044	0.8789393	4.4294308	-50.69472	0.005497826
14	SMI-In-En.O.A-I.C.U1-I.C.U3-y-SMI:I.C.U3-GDP.P:I.C.U3-I:I.C.U3-P.I:I.C.U1-D:I.C.U1-D:I.C.U3-GDP.P:...	15	0.9209532	0.0007894467	0.8783896	5.8601062	-48.34276	0.005510294
15	SMI-In-En.O.A-I.C.U1-I.C.U3-y-SMI:I.C.U1-SMI:I.C.U3-GDP.P:I.C.U3-I:I.C.U1-I:I.C.U3-P.I:I.C.U1-D:I.C....	16	0.9224447	0.0007745512	0.8759115	7.5419853	-45.41018	0.005566152
16	SMI-In-En.O.A-I.C.U1-I.C.U3-y-SMI:I.C.U1-SMI:I.C.U3-GDP.P:I.C.U3-I:I.C.U1-I:I.C.U3-P.I:I.C.U1-D:I.C....	17	0.9236182	0.0007628316	0.8726970	9.2916917	-42.32171	0.005637788
17	SMI-In-En.O.A-I.C.U1-I.C.U3-y-SMI:I.C.U3-GDP.P:I.C.U3-I:I.C.U3-E.O.A:I.C.U1-P.I:I.C.U1-D:I.C.U1-D:I....	18	0.9244770	0.0007542549	0.8686556	11.1085196	-39.07172	0.005726578
18	SMI-In-En.O.A-I.C.U1-I.C.U3-y-SMI:I.C.U1-SMI:I.C.U3-GDP.P:I.C.U3-I:I.C.U1-I:I.C.U3-P.I:I.C.U1-D:I.C....	19	0.9254060	0.0007449770	0.8643745	12.9103728	-35.86561	0.005819157
19	SMI-In-En.O.A-I.C.U3-SMI:I.C.U1-SMI:I.C.U3-GDP.P:I.C.U1-GDP.P:I.C.U3-I:I.C.U3-E.O.A:I.C.U3-P.I:I.C.U...	20	0.9272061	0.0007269991	0.8613449	14.5264209	-33.15359	0.005883792
20	SMI-In-En.O.A-I.C.U1-I.C.U3-y-SMI:I.C.U1-SMI:I.C.U3-GDP.P:I.C.U1-GDP.P:I.C.U3-E.O.A:I.C.U3-P.I:I.C....	21	0.9329831	0.0006693038	0.8659661	15.2942303	-32.83019	0.005784911
21	SMI-In-En.O.A-El.Y-I.C.U1-I.C.U3-y-SMI:I.C.U1-SMI:I.C.U3-GDP.P:I.C.U1-GDP.P:I.C.U3-E.O.A:I.C.U3-P....	22	0.9345714	0.0006534409	0.8622556	16.9554492	-30.10004	0.005864438
22	SMI-Db-In-En.O.A-El.Y-Pr.I-I.C.U3-y-SMI:I.C.U1-SMI:I.C.U3-GDP.P:I.C.U1-GDP.P:I.C.U3-E.O.A:I.C.U3-...	23	0.9384435	0.0006147696	0.8632079	18.1295505	-28.88766	0.005844131
23	SMI-Db-In-En.O.A-El.Y-Pr.I-I.C.U3-y-SMI:I.C.U1-SMI:I.C.U3-GDP.P:I.C.U1-GDP.P:I.C.U3-E.O.A:I.C.U1-...	24	0.9423956	0.0005753000	0.8644602	19.2866031	-27.89468	0.005817317
24	SMI-Db-In-En.O.A-El.Y-Pr.I-I.C.U3-y-SMI:I.C.U1-SMI:I.C.U3-GDP.P:I.C.U1-GDP.P:I.C.U3-I:I.C.U3-E.O.A...	25	0.9442629	0.0005566513	0.8606572	20.8883268	-25.53216	0.005898365
25	SMI-GDP.P-Db-In-En.O.A-El.Y-Pr.I-I.C.U3-y-SMI:I.C.U1-SMI:I.C.U3-GDP.P:I.C.U1-GDP.P:I.C.U3-E.O.A:I...	26	0.9463216	0.0005360908	0.8568576	22.4492185	-23.36165	0.005978243

Output of best subset selection function in R: The selected model was row 13 where R^2_{adj} = 0.8789 using the variables: stock market value, oil energy price, civil unrest indicator, before/after 2009, and the following interactions: stock market value and civil unrest, GDP and civil unrest, inflation and civil unrest, protest intensity and civil unrest, debt and civil unrest, GDP and before/after 2009, and inflation and before/after 2009. Some variables such as protest intensity were included in interaction terms but not individually.

Appendix D

COEFFICIENT SUMMARY	Estimate	Std. Error	t-value	p-value
Intercept	.04519	.01134	3.985	.000351 ***
SMI	4.552e-6	5.722e-7	7.995	3.56e-9 ***
Debt	-4.406e-6	7.325e-7	-5.998	9.71e-7 ***
I.Civil.Unrest1-2	-.02818	.01339	-2.105	.04297 *
I.Civil.Unrest3+	-.04604	.01519	-3.031	.004717 **
Protest.Intensity	-.02922	.01470	-1.987	.055255
I.Civil.Unrest1-2:Protest.Intensity	.03153	.01759	1.793	.082144
I.Civil.Unrest3+:Protest.Intensity	.05521	.02071	2.666	.011799 *

Significant Predictors at $\alpha = .05$: SMI, Debt, Incidents of Civil Unrest, interaction between 3 or more Instances of Civil Unrest and Protest Intensity

Appendix E

95% Confidence Intervals

Predictor	**2.5%**	**97.5%**
SMI	3.388e-06	5.716e-06
Debt	-5.901e-06	-2.911e-06
I.Civil.Unrest1-2	-5.542e-02	-9.463e-04
I.Civil.Unrest3+	-7.694e-02	-1.513e-02
Protest.Intensity	-5.913e-02	6.960e-04
I.Civil.Unrest1-2:Protest.Intensity	-4.247e-03	6.731e-02
I.Civil.Unrest3+:Protest.Intensity	1.307e-02	9.735e-02

Appendix F

Final model diagnostics