ECEN 360 - Final Report

Investigating the correlation between the stock market and federal elections

Jameson Adams (DAEN student, jamdadam@tamu.edu, SPY researcher)
Fletcher Newman (DAEN Student, fletcht13@tamu.edu, Dow Jones/Apple Researcher)
Preston Kouyoumdjian (DAEN, prestonkou12@tamu.edu NASDAQ Researcher)

Executive Summary

The problem statement that our project is attempting to address is to find if there is a correlation between the state of the stock market on certain days leading up to federal election day, and then the outcome of the election, specifically in terms of how many votes the incumbent party receives compared to the challenging party. Our second problem statement was to see if the correlation was strong between the two we could accurately predict the outcome of the 2024 presidential election.

Our overarching objectives were to perform EDA on various major stocks and federal election data, find correlation by utilizing linear regression models between individual stocks and the upcoming elections at different periods, and finally create multi-linear regression models of all of the stocks pooled together separated by the different periods to attempt to see if there is a correlation in the market overall. If there was, we would then create a predictive linear regression model.

Our methodology to complete our objectives is to primarily use python and pandas through google collab. Pandas supplemented by matplotlib was used to clean our data, and separate it into our independent variables, so we could then perform EDA. The growth variables we used were a 1 year time period, 6 months time period, 3 month time period and finally a one-month time period. For these time periods we analyzed six stocks. These were: Apple, Google, Facebook/Meta, Dow Jones, SPY, and NASDAQ. The dependent variable we were attempting to correlate/predict was incumbent ratio. Incumbent ratio was found by subtracting incumbent votes by non-incumbent votes dividing by non-incumbent votes and multiplying by 100 to see if the float was greater than or less than zero. Finally to produce our linear regression models we used statsmodels library which fit our restricted data options. Statsmodels library was also used for our prediction model.

The key findings that we found are that there is no clear correlation between the state of the stock market at any time period and the outcome of that year's election. The time period with the highest correlation was 3 months with an R^2 of .345. This led us to create the prediction model from the 3 month data, but trained off of the six month data because we are currently 6 months out from the election. The output of which was that the incumbent ratio would be 4.392. This means that we predict Joe Biden would win the presidency because he is the incumbent. This is not the most accurate model though due to the low R^2 value, and not optimal training conditions.

Introduction:

To restate our objectives and problem statement, we will use stock price data from major stock exchanges to determine whether or not there is a correlation between the growth of certain stocks and the performance of the leading party in office during a particular election cycle (ie. whether or not the incumbent party maintains their status after an election). After performing our initial EDA and regression analysis, we will then create a model to try and predict (within a certain degree of accuracy) whether or not the incumbent party wins the presidential election using the stock movements within this year.

For the sake of this investigation, we will only focus on the elections that happen at the federal level. More specifically, elections for congress, which happen every 2 years, and presidential elections, which occur every 4 years. Trying to gauge which party the general population will vote for relies on several factors. One of which is the general performance of the economy during their party's electoral term. By observing large stock indices such as the NASDAQ 100 or DOW J, we can get a rough estimate of how good the economy is doing day by day. These stock indices cover a wide range of companies across multiple industries whose performance heavily relies on government policy and regulation.

In the United States, 61% of people have investments within the stock market, and if they see the price of these indices drop dramatically many of them may blame the government for not doing a better job looking out for their interests. As a result, many of them may consider voting for the opposing party. On the other hand, if these particular indexes are doing well then people may think that the president and Congress are passing strong economic legislation that benefits the American people. As a result, the incumbent government may be more favored in the upcoming election cycle.

It is worth noting that the economy is one of many factors that motivate people to vote for a particular candidate. The goal of this report is to try and use these stock indices to track their growth across several periods and use the data to develop a numerical quantification on whether or not the growth of these stocks plays a significant impact on whether or not the incumbent party maintains its influence in office. The results from our experiment will provide more nuance in how we analyze political trends.

Methodology:

To start, our group determined the major stocks that would be analyzed and fed into our prediction model. Below are the six stocks that our group decided on.

Google Apple Dow Jones NASDAQ

SPY Facebook

These stocks were decided on based on different factors such as size, density, population, and relevance. Our group's knowledge of the stock market came in handy when making this decision.

After our group decided which stocks to analyze, the next step was to search for data sets for these stocks. This proved to be more difficult than expected, as we were searching for data sets that went back as far as the year 2000 (or whenever the company started trading publicly), and most of the data sets found either did not go back this far or were not open sourced and required payment. After lots of scouring, appropriate data sets were found that fit the correct specifications[2,3,4,5,6,7]. Below is an example of what the Dow Jones data set looked like in its raw form[2].

	Date	Price	Open	High	Low	Vol.	Change %
304	04/01/2000	10,733.92	10,863.28	11,425.45	10,201.53	4.05B	-1.72%
305	03/01/2000	10,921.93	10,128.11	11,234.65	9,731.81	4.75B	7.84%
306	02/01/2000	10,128.31	10,937.74	11,118.93	9,836.06	3.66B	-7.42%
307	01/03/2000	10,940.53	11,501.85	11,750.28	10,701.64	202.32M	0.00%
308	01/01/2000	10,940.54	11,501.85	11,750.28	10,701.64	3.83B	-4.84%

The next step in our methodology was to clean these data sets. Null values were removed, although there were not many, and odd formatting issues were fixed. For example in the Dow Jones data frame shown above, the volume column was of type ‘object’ because the data had been entered with a ‘B’ or an ‘M’ appended to the end to represent ‘billion’ or ‘million’ respectively. This was fixed by removing all ‘B’ and ‘M’ characters and multiplying the leftover integer by its appropriate power of 10 (4.05B becomes 4,050,000,000 and 202.32M becomes 202,320,000). A similar process was used to remove the '%' character from the 'Change %' column values. All of this work was done using the Pandas library in Python.

Next, federal election data was found to analyze the success of incumbent parties in federal elections[1]. Data sets for each election from 2020-1982 were found (federal elections take place every two years), and each years' data set contained the amount of votes for the democratic and republican parties for each state/territory as shown below (only 5 out of 56 rows are shown).

	State	Democratic Candidates	Republican Candidates	Other Candidates
0	AL	2378911.0	4249258.0	63217.0
1	AK	459702.0	573189.0	34391.0
2	AS	1959.0	9790.0	0.0
3	AZ	5017928.0	4937863.0	55101.0
4	AR	754417.0	2382784.0	454525.0

The 'Other Candidates' column was ignored for this model, as the likelihood that a third party candidate would win the presidential election is very low. The total votes by each state for each party were summed for every year, and the resulting sums were concatenated into one data frame (all using the Pandas library in Python). A section of the resulting data frame is shown below.

Year	Democratic	Republican	Incumbent	Non-incumbent	incumbent_Win	Incumbent_Ratio
2012	176167260.0	158605650.0	176167260.0	158605650.0	1	5.245828
2014	56891574.0	64681903.0	56891574.0	64681903.0	0	-6.407918
2016	179822615.0	167731170.0	179822615.0	167731170.0	1	3.479014
2018	113103585.0	85199865.0	85199865.0	113103585.0	0	-14.071223
2020	205527359.0	195419135.0	195419135.0	205527359.0	0	-2.521091

'Incumbent' and 'Non-Incumbent' columns were added that represent the number of votes the incumbent and non-incumbent parties received that year. These columns were used to calculate the 'Incumbent Ratio', which will be the dependent variable that our model will predict. The incumbent ratio is a statistic that our group came up with that represents the percent of votes the incumbent party received relative to the non-incumbent party and was calculated with the following equation.

$$Incumbent\ Ratio = \frac{Incumbent\ Votes - Non\ Incumbent\ Votes}{Total\ Votes} \times 100$$

As one can see from the equation, a positive Incumbent ratio results in the incumbent party winning the election, while a negative incumbent ratio results in the non-incumbent party winning the election. Our group came up with this statistic because it communicates which party won and the margin of their victory (the distance of the ratio from 0) in one number. This is a good statistic for a machine learning model to predict when predicting the outcome of elections because it will tell us which party is likely to win and how likely it is that that party would end up winning (represented by the relative margin of victory, which is the absolute value of the incumbent ratio).

In order to better specify stock growth relative to upcoming elections, four independent variables were created. All of these variables consist of a stocks percent growth in a period of time leading up to the election. Our group decided that one year, six months, three months, and one month growth from the election would be appropriate time periods to evaluate stock growth relative to the election date. The reasoning behind these time periods was based on how long it takes for stock prices to noticeably change, and at what time period before the election voters begin paying attention to election campaigns. An example of the resulting joined data frame for the Dow Jones growth periods is shown below (the two tables are in one joined data set, but are separate screenshots because of their length).

Year	Democratic	Republican	Incumbent	Non-incumbent	incumbent_Win	Incumbent_Ratio
2012	176167260.0	158605650.0	176167260.0	158605650.0	1	5.245828
2014	56891574.0	64681903.0	56891574.0	64681903.0	0	-6.407918
2016	179822615.0	167731170.0	179822615.0	167731170.0	1	3.479014
2018	113103585.0	85199865.0	85199865.0	113103585.0	0	-14.071223
2020	205527359.0	195419135.0	195419135.0	205527359.0	0	-2.521091

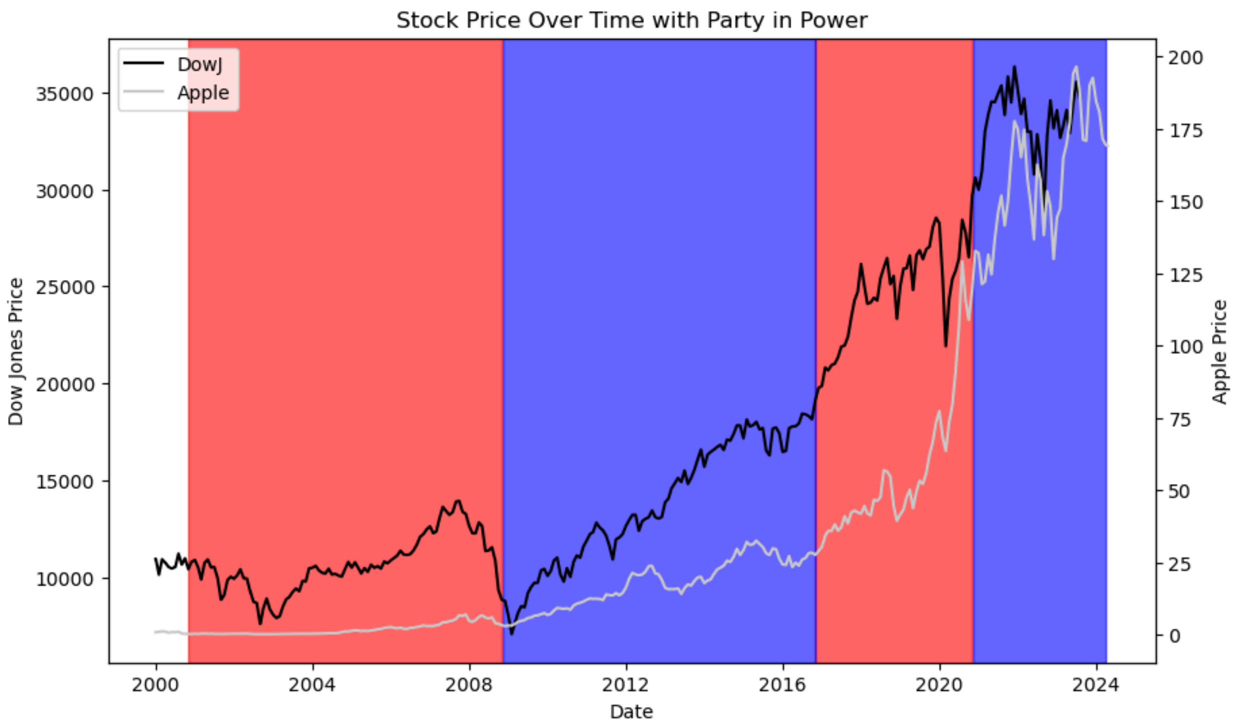
1_Year_DowJ_Growth	6_Month_DowJ_Growth	3_Month_DowJ_Growth	1_Month_DowJ_Growth
1141.45	-0.008947	0.006703	-0.026012
1844.77	0.046559	0.047567	0.019989
478.88	0.020327	-0.015975	-0.009135
1738.52	0.037929	-0.011922	-0.053454
-544.63	0.081349	0.002765	-0.048303

Now that the data frames had been cleaned, each of our group members created predictive models for each of the individual stocks assigned to them. This was done using OLS regression from the statsmodels.api library in python. r^2 values were observed for each model to see how accurate the prediction model would be if just the individual stock was used.

Next, all of the stocks each group member had analyzed were put into one data frame, and prediction models were made using the same OLS regression method as before. Four models were created, each using one year, six months, three months, and one month stock growth data respectively. After observing which model resulted in the highest election prediction accuracy, a prediction was made for the next upcoming election victor.

Results:

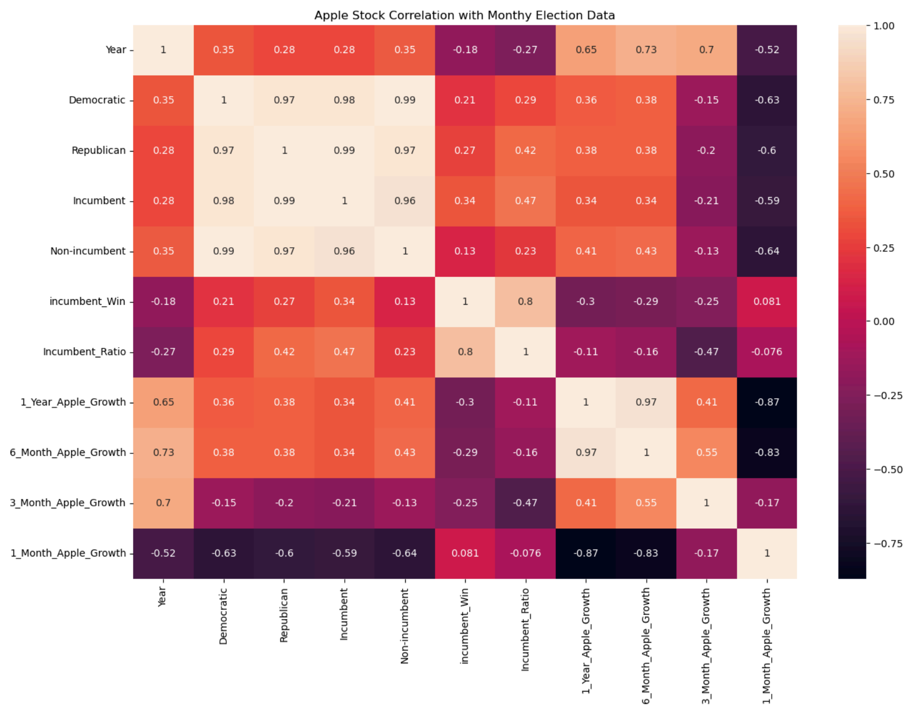
To start, shown below is a graph of the stock growth of the Dow Jones and Apple with the incumbent party color overlaid. This figure was used to make quick observations of economic trends for certain parties in power and for periods near elections.

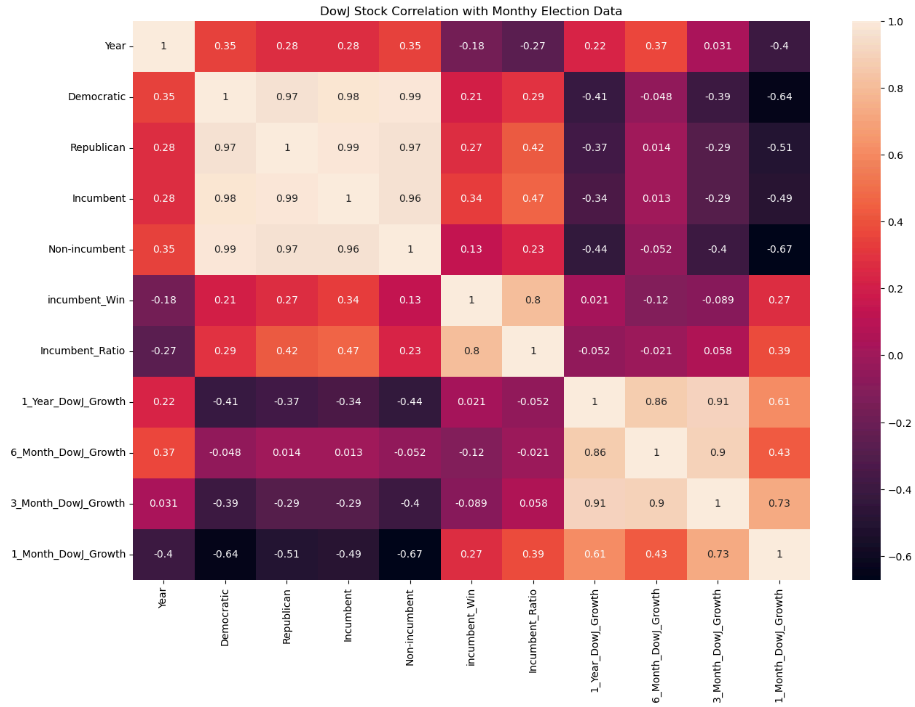


Republican : Red
Democratic : Blue

A few trends can be observed in the graph, one being the economic recession in 2008 resulting in a change of power from the republican to the democratic party. Other small dips in the market can be seen right before November elections in most other election years that the incumbent party lost.

Correlation models for the Dow Jones and Apple with the election data were also created. This is represented in the heat maps below. The correlation models were made using the one year, six month, three month, and one month stock growth data. The main correlation values of interest will be the overlapping squares between the incumbent ratio and various stock growth periods.





From the heat maps, it can be seen that there is little correlation between the incumbent ratio and stock growth during any period. The highest correlation seems to be with three month and one month growth for Apple and Dow Jones respectively (-0.47 and 0.39). This information tells us that short term stock market growth closer to the election will have more of an effect on the election results than long term growth.

Below are the results from the prediction models for each of the individual stocks. These prediction models took in the percent growth from one year, six months, three months, and one month before that years' election as independent variables to predict the incumbent ratio for that year.

```
                            OLS Regression Results
==============================================================================
Dep. Variable:        Incumbent_Ratio   R-squared:                       0.238
Model:                            OLS   Adj. R-squared:                 -0.270
Method:                 Least Squares   F-statistic:                    0.4691
Date:                Sat, 27 Apr 2024   Prob (F-statistic):              0.758
Time:                        18:35:05   Log-Likelihood:                -34.029
No. Observations:                  11   AIC:                             78.06
Df Residuals:                       6   BIC:                             80.05
Df Model:                           4
Covariance Type:            nonrobust
===============================================================================
                  coef    std err          t      P>|t|      [0.025      0.975]
-------------------------------------------------------------------------------
const          -4.1897      3.940     -1.063      0.329     -13.830       5.451
1_year_SPY      0.3320      0.299      1.111      0.309      -0.399       1.063
6_month_SPY    -0.0659      0.374     -0.176      0.866      -0.982       0.850
3_month_SPY    -0.6767      0.963     -0.702      0.509      -3.034       1.681
1_month_SPY     0.0052      1.301      0.004      0.997      -3.178       3.189
==============================================================================
Omnibus:                        2.317   Durbin-Watson:                   2.862
Prob(Omnibus):                  0.314   Jarque-Bera (JB):                1.317
Skew:                          -0.582   Prob(JB):                        0.517
Kurtosis:                       1.767   Cond. No.                         39.4
==============================================================================
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:        Incumbent_Ratio   R-squared:                       0.565
Model:                            OLS   Adj. R-squared:                 -0.016
Method:                 Least Squares   F-statistic:                    0.9731
Date:                Sat, 27 Apr 2024   Prob (F-statistic):              0.530
Time:                        18:50:35   Log-Likelihood:                -22.591
No. Observations:                   8   AIC:                             55.18
Df Residuals:                       3   BIC:                             55.58
Df Model:                           4
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const         -6.1978      3.097     -2.001      0.139     -16.054       3.658
3_monthG       0.5574      0.363      1.536      0.222      -0.598       1.712
1_monthG      -0.9794      0.783     -1.250      0.300      -3.473       1.514
1_yearG        0.2091      0.265      0.790      0.487      -0.634       1.052
6_monthG      -0.4323      0.432     -1.001      0.391      -1.807       0.942
==============================================================================
Omnibus:                        2.211   Durbin-Watson:                   2.650
Prob(Omnibus):                  0.331   Jarque-Bera (JB):                0.535
Skew:                          -0.634   Prob(JB):                        0.765
Kurtosis:                       3.017   Cond. No.                         44.1
==============================================================================
```

Nasdaq

```
                            OLS Regression Results
==============================================================================
Dep. Variable:        Incumbent_Ratio   R-squared:                       0.185
Model:                            OLS   Adj. R-squared:                 -0.359
Method:                 Least Squares   F-statistic:                    0.3399
Date:                Sat, 27 Apr 2024   Prob (F-statistic):              0.842
Time:                        18:55:03   Log-Likelihood:                -34.402
No. Observations:                  11   AIC:                             78.80
Df Residuals:                       6   BIC:                             80.79
Df Model:                           4
Covariance Type:            nonrobust
=================================================================================
                    coef    std err          t      P>|t|      [0.025      0.975]
---------------------------------------------------------------------------------
const            -4.5919      3.062     -1.499      0.184     -12.086       2.902
1_year_NASDAQ     0.2035      0.209      0.976      0.367      -0.307       0.714
6_month_NASDAQ   -0.0659      0.200     -0.330      0.752      -0.554       0.422
3_month_NASDAQ   -0.4536      0.708     -0.641      0.545      -2.185       1.278
1_month_NASDAQ    0.3259      0.612      0.532      0.614      -1.172       1.823
==============================================================================
Omnibus:                        0.057   Durbin-Watson:                   3.150
Prob(Omnibus):                  0.972   Jarque-Bera (JB):                0.245
Skew:                          -0.125   Prob(JB):                        0.885
Kurtosis:                       2.313   Cond. No.                         38.9
==============================================================================
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:        Incumbent_Ratio   R-squared:                       0.570
Model:                            OLS   Adj. R-squared:                  0.385
Method:                 Least Squares   F-statistic:                     3.090
Date:                Sun, 28 Apr 2024   Prob (F-statistic):             0.0991
Time:                        18:44:39   Log-Likelihood:                -30.887
No. Observations:                  11   AIC:                             69.77
Df Residuals:                       7   BIC:                             71.37
Df Model:                           3
Covariance Type:            nonrobust
=====================================================================================
                        coef    std err          t      P>|t|      [0.025      0.975]
-------------------------------------------------------------------------------------
const                -3.4702      1.778     -1.952      0.092      -7.673       0.733
1_year_Facebook       0.1805      0.141      1.283      0.240      -0.152       0.513
6_month_Facebook     -0.0993      0.110     -0.901      0.397      -0.360       0.161
3_month_Facebook      0.1415      0.185      0.764      0.470      -0.297       0.580
1_month_Facebook      1.0738      0.457      2.350      0.051      -0.007       2.154
==============================================================================
Omnibus:                        0.272   Durbin-Watson:                   1.022
Prob(Omnibus):                  0.873   Jarque-Bera (JB):                0.232
Skew:                           0.257   Prob(JB):                        0.890
Kurtosis:                       2.509   Cond. No.                     1.72e+16
==============================================================================
```

DOW JONES

```
                            OLS Regression Results
==============================================================================
Dep. Variable:        Incumbent_Ratio   R-squared:                       0.376
Model:                            OLS   Adj. R-squared:                 -0.040
Method:                 Least Squares   F-statistic:                    0.9037
Date:                Sun, 28 Apr 2024   Prob (F-statistic):              0.517
Time:                        22:47:51   Log-Likelihood:                -32.932
No. Observations:                  11   AIC:                             75.86
Df Residuals:                       6   BIC:                             77.85
Df Model:                           4
Covariance Type:            nonrobust
======================================================================================
                         coef    std err          t      P>|t|      [0.025      0.975]
--------------------------------------------------------------------------------------
const                 -0.7909      2.618     -0.302      0.773      -7.197       5.616
1_Year_DowJ_Growth    -0.0019      0.003     -0.702      0.509      -0.009       0.005
6_Month_DowJ_Growth   48.4628     50.143      0.966      0.371     -74.234     171.160
3_Month_DowJ_Growth  -94.5324    115.580     -0.818      0.445    -377.345     188.281
1_Month_DowJ_Growth  109.0353     63.989      1.704      0.139     -47.541     265.611
==============================================================================
Omnibus:                        6.980   Durbin-Watson:                   2.455
Prob(Omnibus):                  0.031   Jarque-Bera (JB):                2.804
Skew:                           1.070   Prob(JB):                        0.246
Kurtosis:                       4.240   Cond. No.                     1.20e+05
==============================================================================
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:        Incumbent_Ratio   R-squared:                       0.268
Model:                            OLS   Adj. R-squared:                 -0.220
Method:                 Least Squares   F-statistic:                    0.5499
Date:                Sun, 28 Apr 2024   Prob (F-statistic):              0.707
Time:                        22:50:39   Log-Likelihood:                -33.808
No. Observations:                  11   AIC:                             77.62
Df Residuals:                       6   BIC:                             79.60
Df Model:                           4
Covariance Type:            nonrobust
=======================================================================================
                          coef    std err          t      P>|t|      [0.025      0.975]
---------------------------------------------------------------------------------------
const                  -1.4179      2.569     -0.552      0.601      -7.703       4.867
1_Year_Apple_Growth    -0.2678      0.769     -0.348      0.740      -2.150       1.614
6_Month_Apple_Growth    0.2516      1.067      0.236      0.821      -2.360       2.863
3_Month_Apple_Growth   -1.3670      1.451     -0.942      0.382      -4.917       2.183
1_Month_Apple_Growth   -0.8878      2.181     -0.407      0.698      -6.226       4.450
==============================================================================
Omnibus:                        1.105   Durbin-Watson:                   1.525
Prob(Omnibus):                  0.575   Jarque-Bera (JB):                0.858
Skew:                          -0.573   Prob(JB):                        0.651
Kurtosis:                       2.252   Cond. No.                         23.8
==============================================================================
```

Looking at the r^2 values for each model, there is not much accuracy or correlation for any individual stock. The highest correlations observed are the Facebook and Google models with r^2 values of 0.570 and 0.565 respectively.

Now all of the individual stock data were pooled together to make prediction models based on the four different growth periods. The results of these models are shown below.

__One Year__

```
                            OLS Regression Results
==============================================================================
Dep. Variable:        Incumbent_Ratio   R-squared:                       0.075
Model:                            OLS   Adj. R-squared:                 -0.541
Method:                 Least Squares   F-statistic:                    0.1222
Date:                Sat, 27 Apr 2024   Prob (F-statistic):              0.969
Time:                        18:55:39   Log-Likelihood:                -35.095
No. Observations:                  11   AIC:                             80.19
Df Residuals:                       6   BIC:                             82.18
Df Model:                           4
Covariance Type:            nonrobust
```

Six Months

```
                            OLS Regression Results
==============================================================================
Dep. Variable:         Incumbent_Ratio   R-squared:                       0.081
Model:                             OLS   Adj. R-squared:                 -0.531
Method:                  Least Squares   F-statistic:                    0.1326
Date:                 Sat, 27 Apr 2024   Prob (F-statistic):              0.965
Time:                         18:55:48   Log-Likelihood:                -35.060
No. Observations:                   11   AIC:                             80.12
Df Residuals:                        6   BIC:                             82.11
Df Model:                            4
Covariance Type:             nonrobust
```

Three Months

```
                            OLS Regression Results
==============================================================================
Dep. Variable:         Incumbent_Ratio   R-squared:                       0.345
Model:                             OLS   Adj. R-squared:                 -0.092
Method:                  Least Squares   F-statistic:                    0.7899
Date:                 Sat, 27 Apr 2024   Prob (F-statistic):              0.572
Time:                         18:55:53   Log-Likelihood:                -33.199
No. Observations:                   11   AIC:                             76.40
Df Residuals:                        6   BIC:                             78.39
Df Model:                            4
Covariance Type:             nonrobust
```

One Month

```
                            OLS Regression Results
==============================================================================
Dep. Variable:         Incumbent_Ratio   R-squared:                       0.227
Model:                             OLS   Adj. R-squared:                 -0.288
Method:                  Least Squares   F-statistic:                    0.4408
Date:                 Sat, 27 Apr 2024   Prob (F-statistic):              0.776
Time:                         18:58:41   Log-Likelihood:                -34.108
No. Observations:                   11   AIC:                             78.22
Df Residuals:                        6   BIC:                             80.21
Df Model:                            4
Covariance Type:             nonrobust
```

Again looking at the r^2 values, these models do not seem to display much accuracy. Three month and one month growth periods seem to be most correlated with election outcomes while six month and one year growth periods seem to have little to no correlation at all with election outcomes. This shows that when voter opinion is formed or swayed based on the economy, short term change seems to drive their vote more than long term change.

Using the six month growth model, an incumbent ratio for the upcoming 2024 election was predicted. Although the three month model is more accurate, we are currently six months out from the 2024 election at the time this report is being written so the most accurate data available is the six month growth data. Using this data, our model predicted an incumbent ratio value of ***4.39282***. This means that the incumbent party (the Democratic party represented by Joe Biden) will have a vote tally 4.39282% larger than the non-incumbent party (the Republican party represented by Donald Trump). In short, our prediction is that Joe Biden will continue his presidency for a second term after the 2024 election.

Discussion:

Looking at our results it is evident that there is no clear correlation between stock market data for any of the chosen stocks, and the outcome of the presidential election. Even the most highly correlated stocks never reached what many would consider the minimum threshold of an r^2 value. All of the etfs had r^2 values ranging from .185 to .376. The non etf stocks ranged from .268 to .570. It is interesting to see however that all of the non-etf stocks had a higher r^2 value than the etf-stocks. This means that although the overall market is not highly correlated perhaps there are certain individual stocks that would serve to be a better fit for a linear regression model. It is surprising to see that the non-etfs seemed to have a higher correlation than the etfs. We initially assumed that the etfs would have a higher correlation because they track the market as a whole which is one way that people measure how the economy is doing. The state of the economy is almost always a noteworthy issue for anytype of federal election, so we assumed it would have a higher correlation. Clearly this was not the case. In the future different stock options could be explored without heavily changing our model.

Due to the low r^2 value, the value of our predicted incumbent ratio might not be entirely accurate. This problem is further compounded by the fact that our highest pooled correlation value of r^2 = .324 came when the election was three months away, but it is currently 6 months out from the election, so we were not able to use our highest correlated training data. There is nothing we could do to fix this besides wait and check the results of the model later this year to see if the results would change, and if so by how much.

Another limitation that we experienced was a lack of data. Due to us wanting to include more recent companies like Google, Meta, and Apple in our data sets to get an accurate representation of what the market is actually like nowadays, we faced the limitation of only being able to use election data since the year 2000. This only gave us 10 data points to work with. Future research could be done just focusing on the stocks that have been a publicly traded company for much longer, and to use election data going back to their ipos.

Comparing our results to the objectives that we laid out in the beginning of this project, we were able to successfully perform EDA and get multiple diagrams/visualizations of the market etfs and individual high-profile stocks. In addition, we did find the correlation between these stocks and our self-created incumbent ratio by utilizing linear regression models. Based on our linear regression models we were also able to make a linear regression predictive model that although had multiple limitations, was able to predict this year's presidential election winner. Thus we completed each of our objectives.

Conclusion:

In conclusion, we have determined that there is no statistically significant relationship between stock price growth and incumbent party election outcome. In both the individual and the pooled linear regression models, the majority of the generated linear regression models had low r^2 values associated with them, indicating that the model did not fit very well.

Although there were some instances (such as the 3-month pooled stock growth) where the correlation coefficient could be used to draw some conclusions about the underlying relationship. The vast majority of the underlying dataset did not lead to any notable relationships between these 2 variables. Although we were able to extrapolate our data and calculate a predicted incumbent ratio of 4.39282, the high level of uncertainty within our regression models indicates that this value may not be very accurate.

As mentioned previously, a huge limitation within our experiment was due to a lack of data available for analysis. A majority of stocks that are very highly traded did not go public until after 2004 leaving us with a very short election window to be able to run analysis on.

There are so many factors beyond the economy that dictate the outcome of federal elections, having this low of a correlation value is not significant enough to prove that this was the most significant factor deciding whether or not the incumbent party stays in office. However, we believe that if we can run this model again in the future (whenever we have more election data matching our stocks) we will be able to derive a more accurate prediction for future elections giving us a more nuanced perspective when trying to analyze political favoritism during key points before the election.

Further research into more accurate models could lead to a greater insight into other political areas such as state or local elections by tracking economic indicators within the respective area.

Overall, we believe that even though our models did not provide much direct insight into the upcoming election. Given the proper time and data, our methodology should be able to provide us with greater insight into future election cycles.

Future Work:

There is still much future work that could be done on this project. The first future add on would be running this model when we are actually 3 months away from an election. This would allow us to have the most optimal testing conditions for our model. It would be interesting to compare the results of the prediction model with what we got then compared to what the "correct" timing for our model would be. In addition, other future work that could be done is assessing other stocks to see if they would be a better fit for the linear regression models and the prediction model. Finally, without changing the core of our model, we could try to assess different time periods then the ones we chose. This would hopefully give us a better fit then any of the ones we initially chose. Interesting time periods that come to mind would be nine months to see if the trend in our models with further out time periods being less correlated held true, and stock data from the weeks leading up to the election. This would be interesting because it could show us if there is a shocking drop or rise in the economy if it affected the election at all as candidates make their final pitches to the American People.

These models could also be applied to state elections and focus on companies that are chartered or mainly located in one state. This would be very interesting because each state has different policies, so we would be able to compare states to one another. This model could also potentially have a higher correlation because some states are more direct when dealing with specific companies unlike the federal government. A potential problem in such a model would be that companies would not attempt to lobby, or stay in the state and instead they would relocate to other states. This would make it hard to gather enough data for one state to make an effective model. Another potential problem is that some major companies are not centralized which would make it difficult to see the impact the company is having on one individual state's election. Even with these problems in mind, we still believe that this could be an interesting avenue to explore.

One final direction we could take our model would be to apply it to other countries. This would be interesting because many other countries have parliamentary governments. This means they have much more than just two primary parties. This would allow us to see if the state of the economy as expressed as the stock market has a greater or lesser impact on the reelection of the incumbent party when compared to our governmental system with just the two parties. This would bring a whole host of complications and would involve a complete overhaul of our models, and even potentially different datasets that had access to foreign companies.

References:

[1]Federal Election Data provided by the Federal Election Commission:

"Election Results and Voting Information." *FEC.Gov*, www.fec.gov/introduction-campaign-finance/election-results-and-voting-information/.

[2]Dow Jones Stock Price Data:

"Dow Jones Industrial Average Historical Rates (DJI)." *Investing.Com*, www.investing.com/indices/us-30-historical-data.

[3]NASDAQ Stock Price Data:

"Nasdaq 100 Jun 24 (NQ=F) Stock Price, News, Quote & History." *Yahoo! Finance*, Yahoo!, 4 May 2024, finance.yahoo.com/quote/NQ=F?.tsrc=fin-srch.

[4]Apple Stock Price Data:

Apple Inc. (AAPL) Stock Historical Prices & Data - Yahoo Finance, finance.yahoo.com/quote/AAPL/history/.

[5]Facebook/Meta Stock Price Data:

“Meta Platforms, Inc. (Meta) Stock Price, News, Quote & History.” *Yahoo! Finance*, Yahoo!, 4 May 2024, finance.yahoo.com/quote/META?.tsrc=fin-srch.

[6]SPY Stock Price Data:

“SPDR S&P 500 ETF Trust (Spy) Stock Price, News, Quote & History.” *Yahoo! Finance*, Yahoo!, 3 May 2024, finance.yahoo.com/quote/SPY?.tsrc=fin-srch.

[7]Google Stock Price Data:

“Alphabet Inc. (GOOG) Stock Price, News, Quote & History.” *Yahoo! Finance*, Yahoo!, 3 May 2024, finance.yahoo.com/quote/GOOG?.tsrc=fin-srch.

[8]Chat GPT Open AI:

Chatgpt, chat.openai.com.

[9]Gemini AI:

“Gemini - Chat to Supercharge Your Ideas.” *Google*, Google, gemini.google.com/app.

Signature Page

Preston Kouyoumdjian:

(Performed EDA and ran regression independent regression for NASDAQ and FACEBOOK stock datasets).

Jameson Adams: JA

(SPY and Google Analyst for EDA and linear regression, created future prediction model)

Fletcher Newman: Fletcher Newman

(Dow Jones and Apple Analyst for EDA and linear regression, created election data frame and growth period query code)