Uncertainty to Interest Rates: An Interdisciplinary and Cross-Industry Analysis of the Potential Factors That Influence Stock Prices

1.Introduction

In this project, we aimed to use a variety of indicators to predict the movement of the share price of the S&P 500, in order to discover degrees of correlation between the stock market and the events and statistics in question. This analysis focuses on five subcategories of events and statistics, analyzing the impact of each one individually on the S&P 500, before attempting to build larger models, in order to best approximate two attributes of the S&P 500: 1) whether the share price will decrease or increase in a specified time frame and 2) the high of the share price in the specified time frame. We aimed to predict each of these attributes on two timeframes, daily and monthly.

2. Data Standardization for Daily Timeframe: Modeling the Daily Activity of the S&P 500

Throughout this research analysis, both a linear regression and logistic regression were used to develop the models. The data was initially standardized through R, specifically filtering the indicator datasets to only include respective daily and monthly dates that hold historical data from the S&P 500; this removed all weekends and holidays. Three different subcategories of data were used to build the daily prediction models: 1) Exchange Rates, 2) Treasure Securities, and 3) Search Engine Trends. Throughout the paper, the specific variables used in the models will be referenced using abbreviations and acronyms - full descriptions of the variables can be found in figure 1 in the appendix.

In order to determine which specific input data to use in a full predictive model, this project analyzes each subcategory to independently determine whether each subcategory genuinely correlates with the stock market.

3.Exchange Rates
3a) Introduction and Context

Much discussion has been made of how the exchange rates (the value of the American dollar in terms of other currencies) affect the stock market. Conventional theory suggests that stock prices affect the exchange rate rather than the other way around. However, some studies have also argued that the causal chain runs the other way. Some research has shown that exchange rates have the capacity to increase the volatility of the stock prices.

Six exchange rates were chosen for this analysis, from countries with varying levels of economic development, geography and trade with the United States. The last variable is the Broad US Dollar Index which measures the value of the US dollar relative to other global currencies.

3b) Collinearity and VIF

To check for collinearity, a correlation matrix was computed and Variance Inflation Factor (VIF) analysis was conducted. It is clear from the correlation matrix that there is a strong relationship between some variables, particularly between DTWEXBGS and DEXBZUS and between BEXSLUS and DEXINUS (see Apendix 1). As all US exchange rates rely on the strength of the US economy, it is not surprising to see correlations between exchange rates.
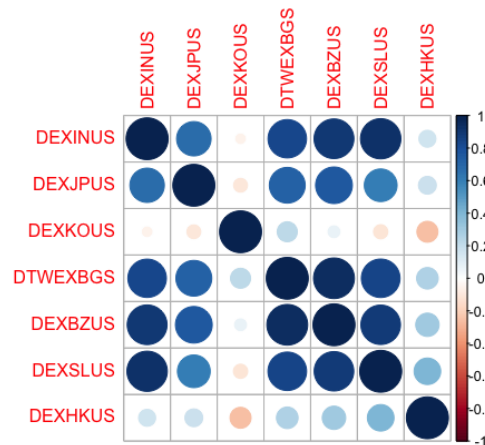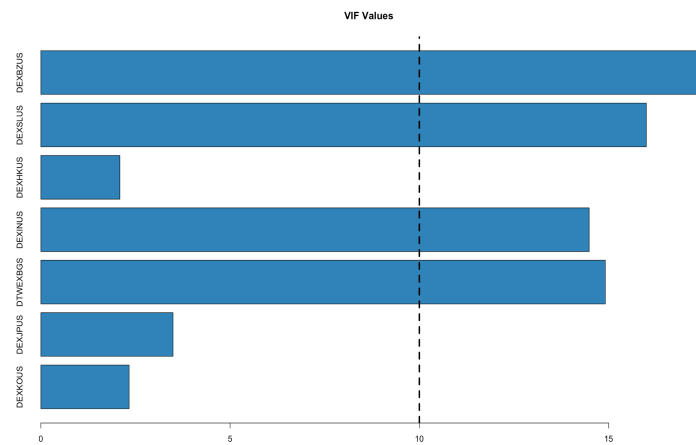
Figure 3a: Coefficient Matrix for variables   *Figure 3b*: VIF of variables

After running an initial linear regression with all seven variables, VIF values were computed (*Figure 3b*). Using the benchmark that any variables with a VIF greater than 10 are considered too collinear to use in a regression,[1] only three variables (DEXHKUS, DEXJPUS and DEXKOUS) will be used in the regression.

3c) Logistic regression, modeling whether the share price went up or down during the day:

Intercept	-0.392
DEXKOUS	-0.000316
DEXHKUS	-1.8576
DEXJPUS	-0.00150

Based on the logistic regression, it seems that DEXHKUS holds the greater influence in increasing the odds of the daily S&P 500 increasing and decreasing. This can be attributed to Hong Kong's larger role in the global economy compared to South Korea and Japan. Hong Kong is largely seen as an international financial center and how it is doing may have ripple effects on the S&P 500.

When the prediction threshold is set at 0.5, the accuracy, precision, and recall of the models are as follows:

$$accuracy = 0.533 \quad precision = 0.536 \quad recall = 0.973$$

When we set the threshold such that the model predicts an "increase" in proportion to the number of true "increases" in the dataset (this achieved at a threshold of 0.54), we get:

$$accuracy = 0.518 \quad precision = 0.556 \quad recall = 0.50$$

---

[1] (Statistics How To, n.d.)

This logistic regression results in a relatively low accuracy and precision. It is unsurprising that three currency exchange rates, on their own, are strong determinants of whether the S&P 500 increases or decreases.

3d) Linear regression, modeling the daily high share price:

$$High = 5017 - 0.1284\,(DEXKOUS) + 2.614\,(DEXJPUS) + 652.7(DEXHKUS)$$

$$\textit{P-value}: 2.2 * 10^{-16} \quad \textit{R-squared}: 0.6659$$

An r-squared value of 0.6659 indicates a relatively high proportion of the variability is explained by the linear regression. Furthermore, the fact that the p-value is significantly lower than 0.05 indicates that the model is statistically significant. This suggests that exchange rates do, indeed, correlate with the stock market, and should be included in the larger model.

4. Treasury Securities

4a) Introduction and Context

Another factor that potentially affects stock prices are treasury securities. Treasury securities are fixed-interest, low-risk US government debt securities that pay back interest payments periodically. Treasury securities are "cornerstones" of the US and global economy (Cussen, 2022) because they are safe investments backed by the US Government. For this reason, it could be helpful to understand what effect the market yield of treasury securities, and their increase or decrease, has on stock prices and financial markets. To that end, five datasets were used in this model to analyze what, if any, effect there is. There are three main types of securities depending on the length of maturity: treasury bills, bonds and notes. This model uses treasury securities at 1-Year (bills), 2-Year, 5-Year, 10-Year and 20-Year (bonds) constant maturity.

4b) Collinearity and VIF

As we can see in the collinearity matrix, the variables are pretty collinear. It makes sense that these securities move in similar directions, especially those with similar maturities. For example, treasury bills and 2-Year treasury notes are extremely collinear (X3.17 and X2.88), but treasury bills and bonds are only very slightly correlated (X3.17 and X4.39). For this reason, only treasury bills, 5-Year treasury notes, and bonds were used in the final linear regression (X3.17, X3.28 and X4.39). However, these three variables still have VIF values of slightly over 10. Consequently, this model serves as a good example of what happens when variables are too collinear (i.e. overfitting). This will be further discussed below.
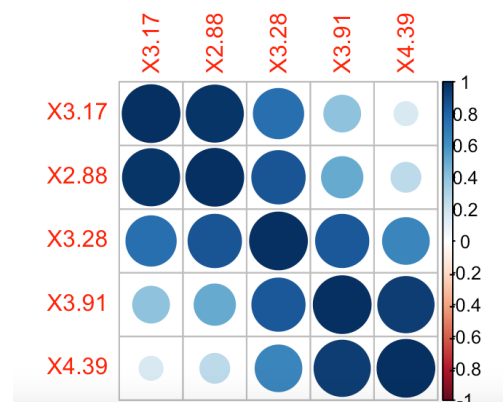
4c) Logistic regression, modeling whether the share price went up or down during the day:

Intercept	0.3759
X3.17	-0.2206
X3.28	0.13799
X4.39	-0.09984

Accuracy: 54.2%

Precision: 54.6%

Recall: 86.2%

The model shows that treasury bills have the greatest influence on changes in stock prices. This could be because treasury bills (1-Year constant maturity) are more common than longer term bonds. However, the accuracy and precision of the model are fairly low. Once again, the precision could have been decreased due to the multicollinearity of the variables.

4d) Linear regression, modeling the daily high share price:
> *Before checking for linearity:*
> $$High = 0.833 + 55.376\,(X3.17) - 71.055\,(X2.88) + 5.178(X3.28)$$
> $$+ 229.081\,(X3.91) - 227.091\,(X4.39)$$
> *P-value*: $2.2 * 10^{-16}$ *R-squared*: 0.8694

> *After checking for linearity (and removing variables):*
> $$High = 0.747 - 5.516\,(X3.17) + 80.921\,(X3.28) - 88.413(X3.91)$$
> *P-value*: $2.2 * 10^{-16}$ *R-squared*: 0.807

After removing the two variables the R-squared value decreased by about 0.06, but it is still relatively high. The p-values are also very low, indicating statistical significance and correlation between treasury securities and stock prices. However, it is important to talk about linearity here. One of the dangers of using collinear variables in regression analysis is overfitting the model. This makes the model a worse predictor of future data because it does not reflect the relationship between the variables. Thus, the model has low error in reference to its training data but it may have high error on test data. Moreover, collinear variables decrease the precision of the estimates and the model, which means the P-values and R-squared values are unreliable. This is not to say that the market yield of treasury securities has no effect on stock prices, but it is important to mention these limitations.

5. Search Engine Trends
5a) Introduction and Context
Measuring patterns in search engine activity has become possible in the past two decades as companies offering the service have amassed the following and capacity necessary to collect and publish search engine trends. For Google, this service is Google Trends, a tool that generates reports on the popularity of search queries in Google by geographic location and time. Google Trends attributes an individual popularity rank to each search term. This rank measures "relative popularity" (at each point in time during a selected time frame) against a term's maximum popularity during that same period and

specified location. If the search term "clarify" was searched most (being 50 searches) on June 1, 2018, out of the entire period for which data regarding searches of the term "clarify" exist (1/1/2004 - Present), Google Trends will return that "clarify" has a popularity score of 100 for June 1, 2018. If 25 searches occurred for "clarify" any other day during this period, the popularity score that day = 50 (Shivers).

5b) Collinearity and VIF
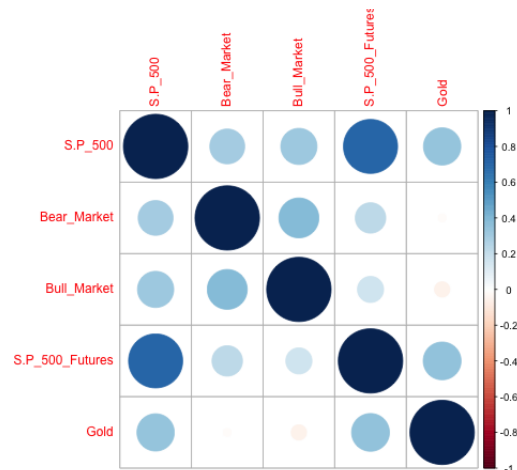
From the correlation matrix, we can recognize three major relationship trends which arise between our covariates:

    1) *S.P_500* **with** *S.P_500_Futures*
    2) *Bull_Market* **with** *Bear_Market*
    3) *Gold* **with** *S.P_500* **&** *S.P_500_Futures*

These relationships appear somewhat intuitive given the nature of user interaction with a search engine query: increasing searches for the phrase "S&P 500" puts more users in proximity to information relevant to the search phrase "S&P 500 Futures".

After running an initial linear regression with all five variables, VIF values were computed. All variables had VIF values equal to or below 2.275. Thus, all five variables were included in the linear and logistic regression models.

5c) Logistic regression, modeling whether the share price went up or down during the day:

Intercept	0.166386
S.P_500	0.015115
Bear_Market	-0.036347
Bull_Market	0.007526
S.P_500_Futures	-0.020854
Gold	0.001341

Accuracy: 55.4%
Precision: 55.5%
Recall: 85.7%

5d) Linear regression, modeling the daily high share price:

$$High = 87.51 + 4.67\,(S.P\_500) - 1.93\,(Bear\_Market) - 0.80(Bull\_Market)$$
$$- 0.84\,(S.P\_500\_Futures) + 1.17\,(Gold)$$
$$P\text{-value}: (< 2.2e - 16) \quad R\text{-squared}: 0.4629$$

These values suggest that search engine terms may correlate with the stock market, at least to a minimal degree, and should be included in the larger model.

6. Monthly Time Frame: Modeling the Monthly Activity of the S&P 500

Two different subcategories of data were used to build the monthly prediction models: 1) Financial Indicator Rates, 2) Economic Uncertainty. The specific variables used in the creation of these models and their measurement is featured in Figure 1.

7) Economic Uncertainty - Global and Domestically

7a) Introduction and Context

"Markets tank over new questions about where the economy is heading," a recent headline by "The Washington Post" in April 2022 – and a sentiment commonly seen for the last few months. This understanding of economic uncertainty has demonstrated substantial effects on the stock market's performance and this portion of the analysis aims to quantify this to predict a monthly's increase or decrease in the S&P 500. Initial literature has suggested that greater economic policy uncertainty has contributed to an increase in stock volatility. Other research (Tromler, Cody (2022)) has found that the stock market's performance and volatility shifts personal saving habits – however little research has shown how personal savings and confidence in the market has impacted the stock market.

With uncertainty tying into various aspects of the stock market, these seven covariants were selected as proxies to the multi-faceted idea that is economic uncertainty. This includes various uncertainty indexes, including monetary policy, fiscal policies, and financial regulation; uncertainty indexes for various markets; and, a proxy to consumer sentiment regarding the stock market. These data sets are selected from 1993 to 2018 on a quarterly basis, sourced from Yahoo Finance and FRED.

7b) Colinearlity and VIF
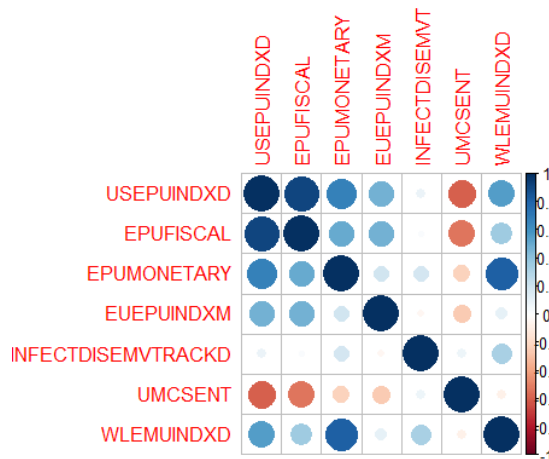
Based on this correlation plot across the variables, USEPUINDXD (U.S. Uncertainty) and EPUFISCAL (Fiscal Policy) have a strong positive correlation, followed by EPUMONETARY and EUEPUINDXM (Europe Uncertainty). This indicates that there is some connection between European and U.S. economic conditions, as well as demonstrating fiscal policy impacting uncertainty to a higher degree than monterary or financial regulation policies. Additionally, USEPUINDXD carries a negative correlation with UMCSENT (Consumer Sentiment). This means that with a stronger economic uncertainty in the U.S., the consumer sentiment becomes more negative.

These factors were analyzed for their VIF values, of which, determined no factor needs to be removed from analysis. Further calculation can be found in Figure 2 in the Appendix.

7d) Linear Regression Model

$$High = 54.46 - 0.13\,(USEPUINDXD) + 0.13\,(EPUFISCAL) - 0.02(EPUMONETARY)$$
$$+ 0.65\,(EUEPUINDXM) + 15.911\,(INFECTDISEMVTRACKD) + UMCSENT\,(0.49)$$
$$- 0.18(WLEMUINDXD)$$

$$\textit{P-value}: 2.2 * 10^{-16} \quad \textit{R-squared}: 0.4844$$

The linear model demonstrates that INFECTDISEMVTRACKD, UMCSENT, and EUEPUINDXM having a positive correlation to the predicted stock price – indicating that higher equity volatility with infectious diseases, positive consumer sentiment, and European policy uncertainty contribute to higher monthly prices. On the contrary, greater domestic uncertainty such as USEPUINDXD, WLEMUINDXD contribute to a lower stock price.

7e) Logistic Regression Model

Intercept	-0.3922012850	Based on the logistic regression model, INFECTDISEMVTRACKD holds the greater influence in increasing the odds of a specific month's increase or decrease in stock price. This means during increased equity market volatility measured with infectious diseases, the model expects a ~1.55 increase in odds per unit increase. This may be explained by the limited exposure to infectious diseases during the period from 1993 - 2018 combined with large downfalls (inspired by infectious diseases) in the market that are often followed by large rallies. It also seems a stronger positive consumer sentiment (UMCSENT) increases the odds of a positive change in the S&P 500. On the other hand, slight negative coefficients are shown for WLEMUINDXD and USEPUINDXD, indicating a decrease in economic uncertainty is associated with higher odds of a positive month.
USEPUINDXD	-0.0039855096	
EPUFISCAL	0.0053183054	
EPUMONETARY	0.0005064827	
EUEPUINDXM	0.0003242229	
INFECTDISEMVTRACKD	0.4473854919	
UMCSENT	0.0116606573	
WLEMUINDXD	-0.0088768462	

When the prediction threshold is set at 0.5, we witness the following:

Accuracy: 62.9%   Precision (63.6%)   Recall (92.67%)   Predicting Rate: 89%

However, when we set the threshold to be at an amount where the model's positive predictions are proportionally approximate to the actual data set's positive months (a threshold of 0.622), we witness the following:

*Accuracy: 61.6%   Precision: 68.9%   Recall: 68.5%   Predicting Rate: 61%*

8) Financial Indicator Rates
8a) Introduction and Context
        Interest rates are a topic of great importance for those looking to understand the inner workings of the economy. Governments play a key role in trying to control interest rates in order to prevent rampant economic inequality. Part of the way that the United States government attempts to control interest rates is by setting targets for the federal funds rate, or the rate at which banks loan to each other. It is widely believed that altering the target federal funds rate has a long-term impact on the health of the economy,

and a shorter-term impact on the stock market. Other indicators of economic health include the unemployment rate and the inflation rate, each of which are measured and reported monthly by the Bureau of Labor Statistics. This section will examine the relationship between these three rates and the change in stock price of the S&P 500. The data used for each statistic was collected on the first day of each month, starting in 1985.

8b) Collinearity and VIF
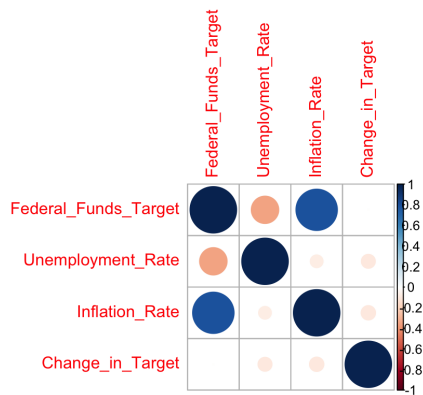
This correlation plot demonstrates that the only input variables with a strong positive correlation are the Federal_Funds_Target and the Inflation_Rate. This correlation is expected, as raising interest rates is a commonly suggested and implemented policy in response to rising inflation rates. There is also a weak negative correlation between FFT and Unemployment_Rate.

None of the variables have a VIF value above 10, and we will therefore use all four in constructing our models. Further analysis can be seen in Figure 6 in the appendix.

8c) Linear Regression Model

$$High = 3538.74 - 190.87 \ (Federal\_Funds\_Target) - 270.13 \ (Unemployment\_Rate)$$
$$- \ 50.25 \ (Inflation\_Rate) + 64.27 \ (Change\_in\_Target)$$
$$P\text{-}value: 2.2 * 10^{-16} \quad R\text{-}squared: 0.764$$

8d) Logistic Regression Model

Across the 407 months between 1985 and 2018, 266 out of the 407 months (or 65.36%) saw a net increase in the S&P 500 stock price. This means that a "model" which always predicts an increase in the stock price from month to month would expect to have an accuracy of 65.36%. With a threshold of 0.5, our logistic regression performs almost exactly the same as an always-positive model, predicting a net increase in the stock market for 406 of the 407 months. This yields:

Accuracy: 65.11%   **Precision: 65.27%**   **Recall: 99.62%**   **Predicting Rate: 99%**

This model is slightly less accurate than the always-positive model, whereas it is significantly more accurate than a random-guessing model.

With a threshold (in this case, 0.6403) such that the model predicts an increase in proportion to reality (i.e. it predicts an increase 65.36% of the time), the model becomes slightly less accurate:

Accuracy: 60.69%   **Precision: 70.70%**   **Recall: 68.05%**   **Predicting Rate: 65%**

This model performs considerably worse than an always-positive model. This yet again suggests that the impacts of the federal funds rate, interest rate, and unemployment rate on the monthly trajectory of the stock market are negligible.

Coefficient values of the logistic regression:

Intercept	0.12780666
Federal_Funds_Target	0.09775204
Inflation_Rate	-0.22808935
Change_in_Target	-0.64771043
Unemployment_Rate	0.12911277

As can be seen from the coefficient table above, the Change_in_Target seems to have the largest impact on the S&P 500 share price, in comparison to the other economic indicator rates. This suggests that, when the Federal Reserve changes their target inflation rate, the stock market is genuinely impacted.

9. Combined Models – Monthly

After completing sub categorically-divided models, all datasets were combined in effort to create a move effective model. Following a similar level of analysis from before, the following linear, logistic regressions were created. In addition, L1 Regularization (Lasso) was used to prevent overfitting to the training data – producing a final model for future dates. Due to the appearing lack of accuracy from daily datasets combined with limited factor datasets, only monthly was combined for further analysis.

9a. Combined Model: Monthly Data Factors

Linear Regression, predicting the monthly highs of the S&P 500:

$$High = 4523.5418 + 0.8892 \ (Federal\_Funds\_Target) - 306.1387 \ (Unemployment\_Rate) - 405.3842$$
$$(Inflation\_Rate) - 1.3868 \ (Change\_in\_Target) + 0.6210 \ (EPUFISCAL)$$
$$- 2.3265 \ (EPUMONETARY) - 1.7929 \ (WLEMUINDXD) - 11.0223 \ (UMCSENT)$$
$$+ 5.2209 \ (EUEPUINDXM) - 15.6904 \ (INFECTDISEMVTRACKD)$$
$$P\text{-}value: < 2.2e\text{-}16 \qquad R\text{-}Squared: 0.707$$

Logistic Regression, predicting the monthly increase or decrease of the S&P 500:

$$Binary = -.227 - 0.001 \ (Federal\_Funds\_Target) + 0.197 \ (Unemployment\_Rate) - 0.215 \ (Inflation\_Rate)$$
$$+ 0.003 \ (Change\_in\_Target) + 0.0035 \ (EPUFISCAL)$$
$$+ 0.00022 \ (EPUMONETARY) - 0.0127 \ (WLEMUINDXD) + 0.032 \ (UMCSENT)$$
$$-0.002 \ (EUEPUINDXM) - 0.155 \ (INFECTDISEMVTRACKD)$$
$$P\text{-}value: < 2.2e\text{-}16 \qquad R\text{-}Squared: 0.707$$

Accuracy: 69.67%	*Precision: 76.38%*	*Recall: 76.38%*	*Predicting Rate: 65%*	*Threshold: 64.78%*
Accuracy: 70.00%	*Precision: 70.22%*	*Recall: 92.46%*	*Predicting Rate: 65%*	*Threshold: 50%*

Both the regression and linear models indicate positive influential factors include inflation rate and EUEPUINDXM. On the other hand, the model also witnesses several negative factors including unemployment rate, INFECTDISEMVTRACKD, and UMCSENT.

10. Combined Models - Daily

All three subcategories of exchange rates, treasury securities, and search engine trends have demonstrated correlations with stock market price, and will therefore be used in the combined model.
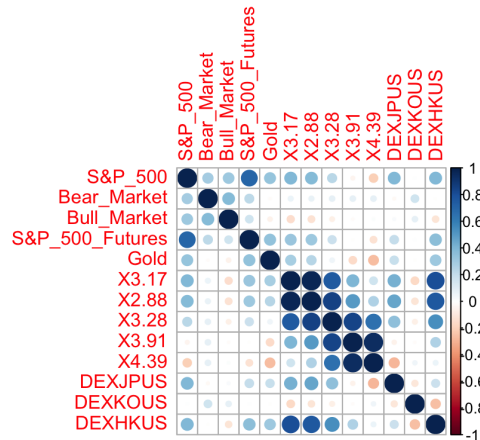
The correlation matrix of all daily variables, reveals two specific trends of the relationships between covariates originating from separate individual models:

1) Generally, variables from the one origin model tend to not correlate with variables from other origin models.
2) There is some correlation between variables from the Exchange Rates model and the Treasury Securities model when in this Combined Model

9a. Combined Model: Daily Data Factors

Linear Regression, predicting the monthly highs of the S&P 500:

$$High = -4302 + 1.611 (S.P\_500) - 0.2527 (Bear\_Market) - 0.2587 (Bull\_Market) - 0.2908 (S.P\_500\_Futures) + 0.5851 (Gold) - 22.34 (X3.17) + 69.04 (X3.28) - 78.98 (X3.91) + 1.291 (DEXJPUS) - 0.1567 (DEXKOUS) + 554.4 (DEXHKUS)$$

P-value: < 2.2e-16   R-Squared: 0.8961

The linear model demonstrates that, between exchange rates, treasury securities, and search engine terms, a pretty successful linear regression model can be built to predict the high of the S&P 500 share price on a given day. Specifically, it is interesting to further consider the magnitudes separating variables from a given individual model when combined as above.

Logistic Regression, predicting the monthly increase or decrease of the S&P 500:

$$Binary = -9.61 + 0.023 (S.P\_500) - 0.035 (Bear\_Market) + 0.004 (Bull\_Market) - 0.0221 (S.P\_500\_Futures) + 0.0033 (Gold) - -0.3507 (X3.17) + 0.362 (X3.28) - 0.1504 (X3.19) - 0.005 (DEXJPUS) - 0.00001 (DEXKOUS) + 1.32 (DEXHKUS)$$

P-value: < 2.2e-16   R-Squared: 0.8961

Based on the logistic regression model, *DEXHKUS* holds the greater influence in increasing the odds of a specific month's increase or decrease in stock price across all daily variables in the combined model. Alternatively, the coefficients for *Bear_Market*, *S.P_500_Futures*, *X3.17,* and the majority of the other variable's coefficients are negative, and no other positive coefficient comes near *DEXHKUS's* influence.

When the prediction threshold is set at 0.5, the accuracy, precision, and recall of the models are as follows:

$$accuracy = 0.5580 \quad precision = 0.5593 \quad recall = 0.830$$

When we set the threshold such that the model predicts an "increase" in proportion to the number of true "increases" in the dataset (this achieved at a threshold of 0.54), we get:

$$accuracy = 0.5518 \quad precision = 0.5824 \quad recall = 0.5824$$

L1 Regularization (Lasso) – Generalizing Model

In addition to combining all factors into models, each model was redefined using L1 regularization to introduce an additional parameter regulating the magnitude of covarient coefficients as part of the objective function. However, given the context of the data we are plotting, the entire data set is used to regularize and determine the most effective lambda for future predictions beyond the 2018 dataset. The initial AUC calculated for models include 0.6776 and 0.567 for monthly and daily, respectively. Daily was included in regularization to see if this would increase the performance of the models created beforehand.

AUC Curves (Pre-Regularization), Monthly (Left); Daily (Right)
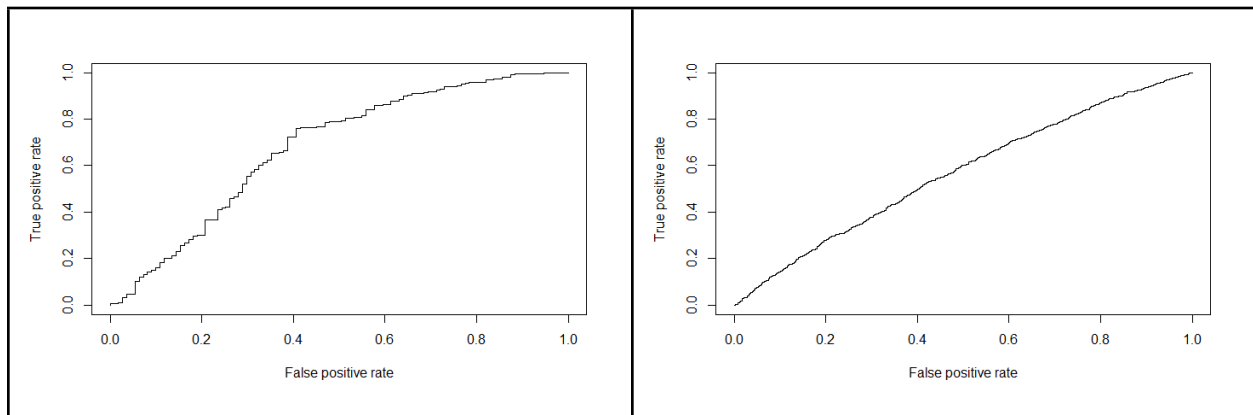
After regularization and testing ten lambda values, the most effective lambda was inputted into the new monthly and daily models. Figures 3 and 4 demonstrate the AUC changing in relation to the various lambda values as well as the coefficients post regularization. As a result, several coefficients were zeroed out, revealing several factors most influential including inflation rate, unemployment rate, UMCSENT, and WLEMUINDXD for monthly. In terms of daily, we see all of the factors not zeroed out, meaning they all contribute relatively similar influences to the model's prediction. With the new lambda values and regularization, the AUC for each model changed to 0.6758 and 0.5675, for monthly and daily respectively.

Conclusion

This paper highlights the roles different variables play in the stock market. When combined, the inflation rate, unemployment rate along with other economic indicators such as the economic policy uncertainty index can predict at approximately 70% whether the stock market increases or decreases on a monthly scale. On the other hand, on a daily time frame variables such as the exchange rates of different

countries, treasury securities, and search engines all indicate a much lower accuracy of around 55%. Stock markets are known to fluctuate on a daily basis based on a plethora of factors and hence, the three factors alone cannot account does not account for the entire change,

Appendix.

Figure 1. Table of all Variables Analyzed

| Variable	Description	Year	Source
High	The daily/month;y high of the S&P 500 Index	2008 -2018	FRED[2]
Binary	Whether the S&P500 share price increased or decreased over the course of the previous month/day.	2008 -2018	FRED
Exchange Rates			
DEXKOUS	South Korean Won to U.S. Dollar Exchange Rate	2008 -2018	FRED
DEXJPUS	Japanese Yen to U.S. Dollar Exchange Rate	2008 -2018	FRED
DEXINUS	Indian Rupees to U.S. Dollar Exchange Rate	2008 -2018	FRED
DEXHKUS	Hong Kong Dollars to U.S. Dollar Exchange Rate	2008 -2018	FRED
DEXSLUS	Sri Lankan Rupees to U.S. Dollar Exchange Rate	2008 -2018	FRED
DEXBZUS	Brazilian Reals to U.S. Dollar Exchange Rate	2008 -2018	FRED
DTWEXBGS	Broad US Dollar Index	2008 -2018	FRED
U.S. Treasury Securities			
X3.17	Market Yield on U.S. Treasury Securities at 1-Year Constant Maturity	2008 - 2018	FRED
X2.88	Market Yield on U.S. Treasury Securities at 2-Year Constant Maturity	2008 - 2018	FRED
X3.28	Market Yield on U.S. Treasury Securities at 5-Year Constant Maturity	2008 - 2018	FRED
X3.91	Market Yield on U.S. Treasury Securities at 10-Year Constant Maturity	2008 - 2018	FRED
X4.39	Market Yield on U.S. Treasury Securities at 20-Year Constant Maturity	2008 - 2018	FRED
Search Term Trends			
S.P_500	Relative popularity of search term "S&P 500"	2008 -2018	Google Trends
Bear_Market	Relative popularity of search term "Bear Market"	2008 -2018	Google Trends
Bull_Market	Relative popularity of search term "Bull Market"	2008 -2018	Google Trends
S.P_500_Futures	Relative popularity of search term "S&P 500 Futures"	2008 -2018	Google Trends
Gold	Relative popularity of search term "Gold"	2008 -2018	Google Trends
Economic Uncertainty			
USEPUINDXD	Economic Policy Uncertainty Index for United States	1993 - 2018	FRED
WLEMUINDXD	Equity Market-related Economic Uncertainty Index	1993 - 2018	FRED
EPUMONETARY	Economic Policy Uncertainty Index: Categorical Index: Monetary policy	1993 - 2018	FRED
EPUFISCAL	Economic Policy Uncertainty Index: Categorical Index: Fiscal Policy (Taxes OR Spending)	1993 - 2018	FRED

---

[2] Federal Reserve Economic Data (FRED)

|			
INFECTDISEMVTRACKD	Equity Market Volatility: Infectious Disease Tracker	1993 - 2018	FRED
EUEPUINDXM	Economic Policy Uncertainty Index for Europe	1993 - 2018	FRED
UMCSENT	University of Michigan: Consumer Sentiment	1993 - 2018	FRED
Interest, Inflation, and Unemployment			
Federal_Funds_Target	The target federal funds rate that the Federal Reserve has in effect on the first of the month.	1985 - 2018	FRED
Change_in_Target	The difference between the TFFR currently in place and the TFFR in place on the first of the previous month.	1985 - 2018	FRED
Unemployment_Rate	The unemployment rate released on the first of the month.	1985 - 2018	FRED
Inflation_Rate	The inflation rate released on the first of the month.	1985 - 2018	FRED
High	The highest price of the S&P500 over the course of the month.	1985 - 2018	Yahoo Finance

Figure 2.

| FACTOR / COVARIANT	VIF Value
USEPUINDXD	9.67
EPUFISCAL	8.002625
EPUMONETARY	3.964965
WLEMUINDXD	3.961392
UMCSENT	2.019779
EUEPUINDXM	1.380133
INFECTDISEMVTRACKD	1.157546

Figure 3.

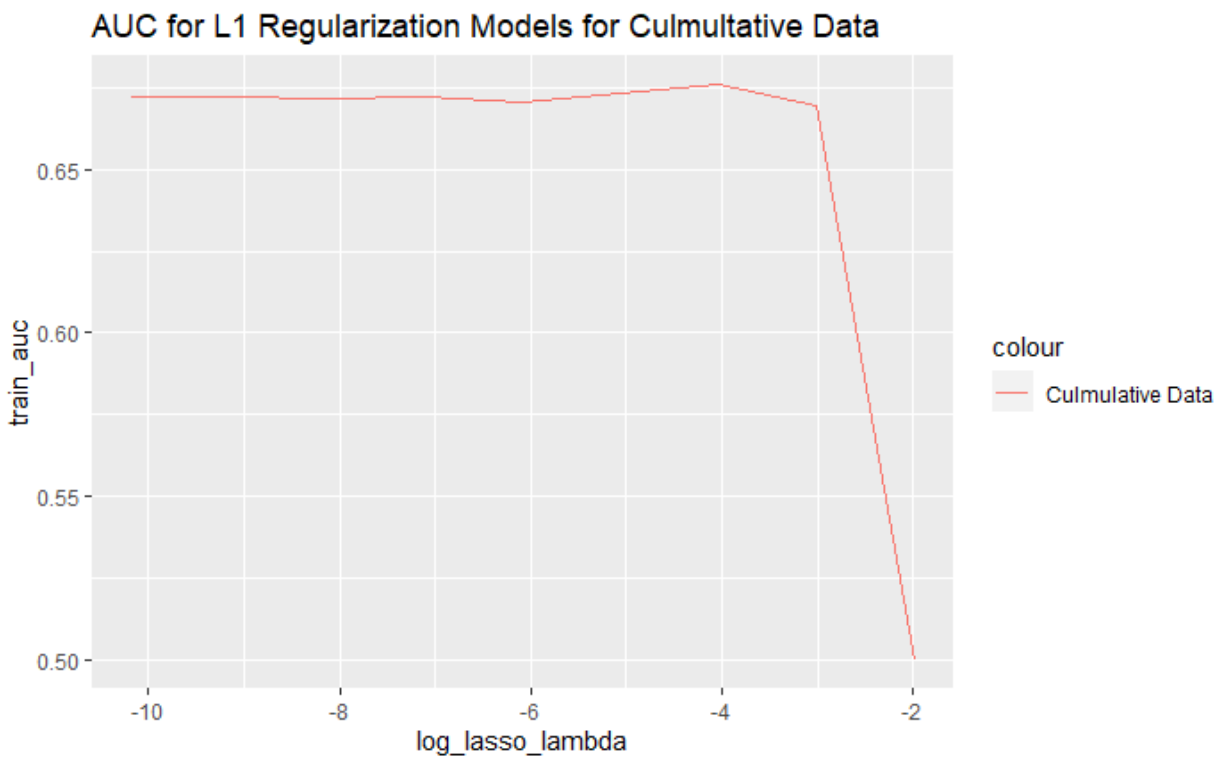Regularized Coefficients for Monthly
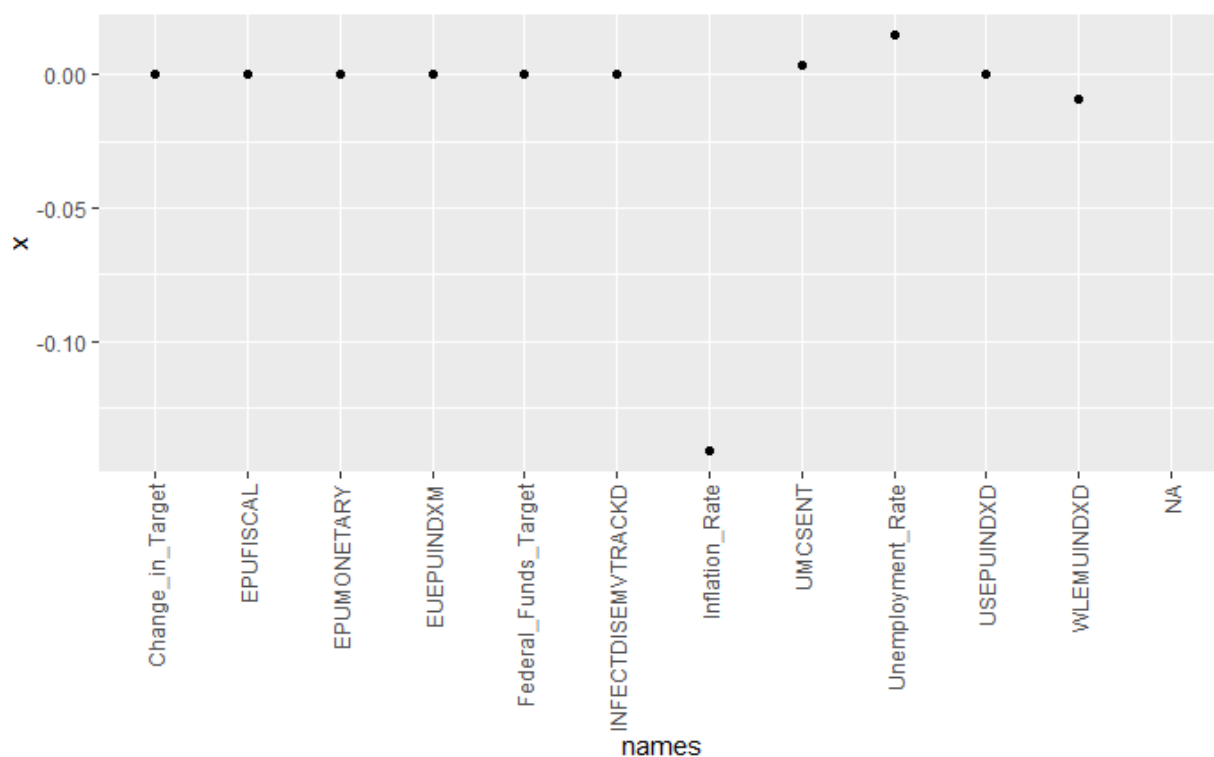
Regularized Coefficients for Daily
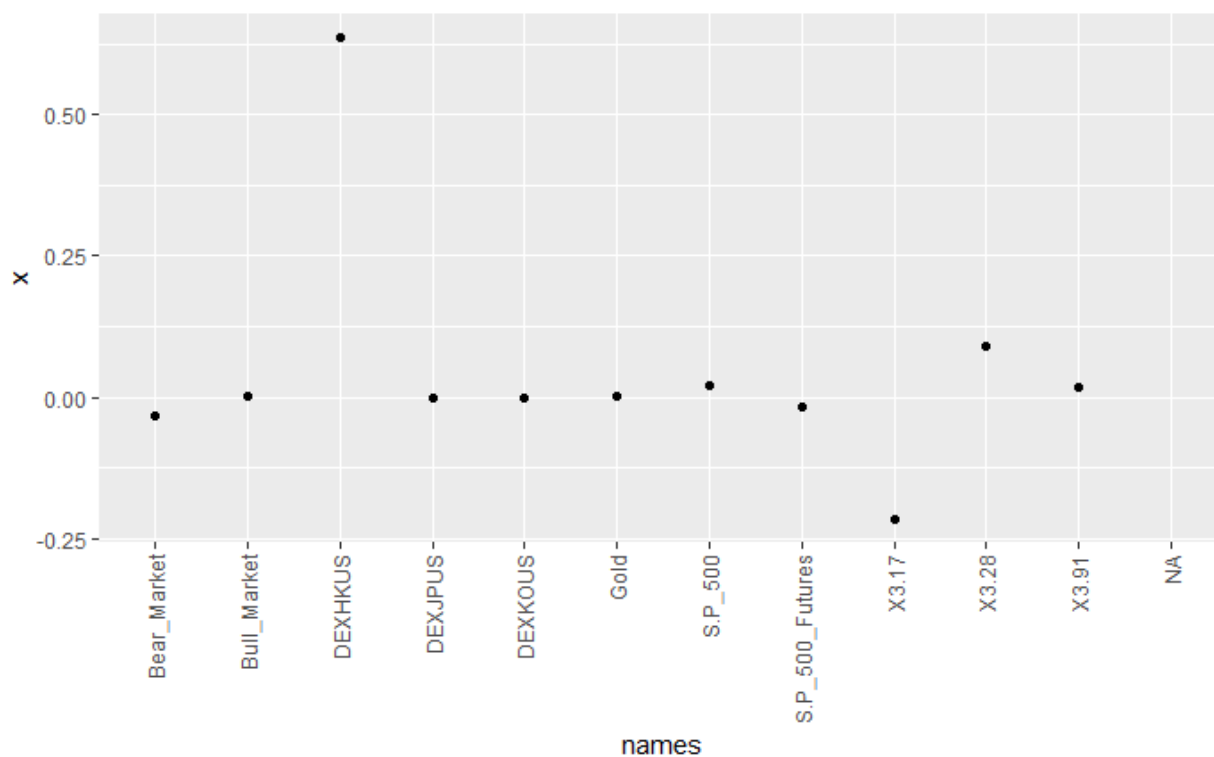
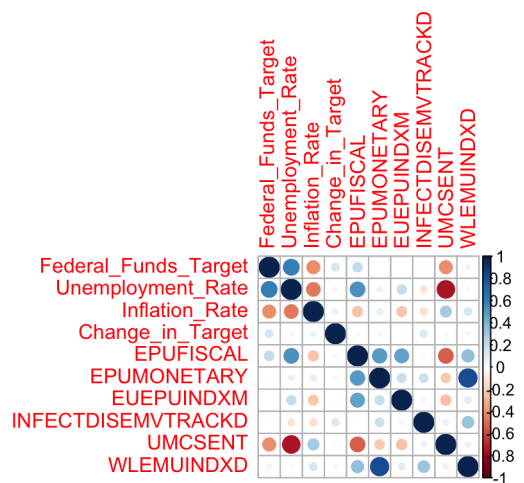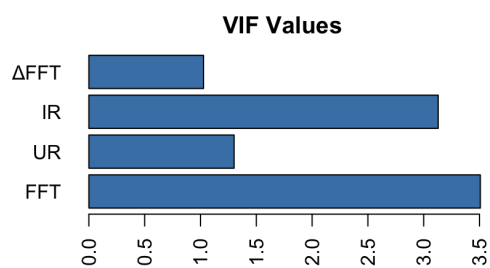**Figure 4. AUC for L1 Regularization Models for Monthly and Daily**

Monthly

AUC for L1 Regularization Models for Culmultative Data
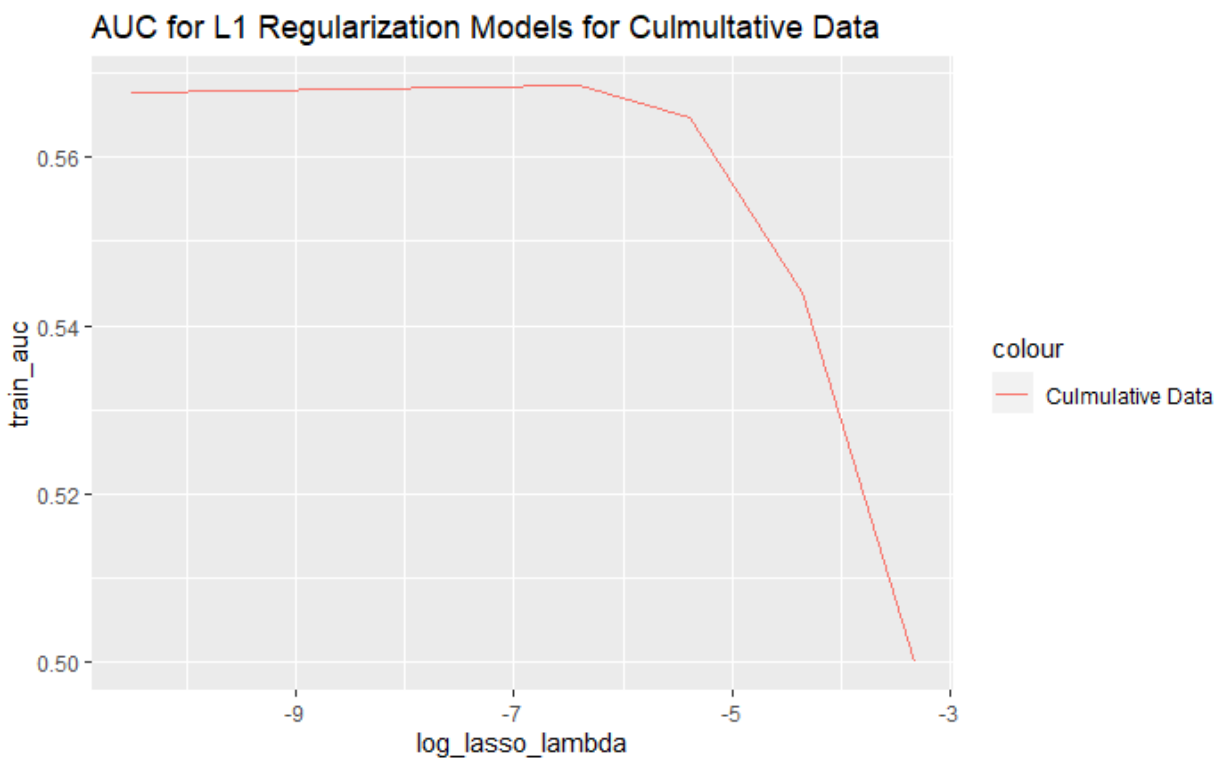
Daily

Figure 5. Correlation Between Monthly Factor Types

Figure 6. VIF values for economic indicator rates

Bibliography

Cussen, Mark P. "Introduction To Treasury Securities". *Investopedia*, 2022,
    https://www.investopedia.com/articles/investing/073113/introduction-treasury-securities.asp.

Hall, M. (2022, May 16). *How Do Interest Rates Affect the Stock Market?* Investopedia.
    Retrieved May 30, 2022, from
    https://www.investopedia.com/investing/how-interest-rates-affect-stock-market/#:~:text=As%20a
    %20general%20rule%20of,stock%20market%20to%20go%20down

Rasure, E. (2022, January 24). *What Is the Trade-Weighted US Dollar Index?* The Balance. Retrieved
    May 25, 2022, from
    https://www.thebalance.com/what-is-the-trade-weighted-us-dollar-index-5216766

Statistics How To. (n.d.). *//. //* - Wikipedia. Retrieved May 25, 2022, from
    https://www.google.com/search?q=what+vif+value+is+too+high&oq=what+vif+value+is+too+hi
    gh&aqs=chrome..69i57j69i60.5734j0j7&sourceid=chrome&ie=UTF-8

Shivers, Dartanyon. Exploring the Relationship between Google ... - Ccom.ucsd.edu. Chris Deotte PhD,
    Self Published, June 2017, https://www.ccom.ucsd.edu/~cdeotte/papers/GoogleTrends.pdf.