What is the Federal Reserve Actually Doing?

Evan Kurzman (EHK73), Kelvin Sun (KWS67), Payton Martin (PCM65)

Introduction

The Federal Reserve ('Fed') has gained more power and influence than arguably ever before. Unless you live under a rock, you might see the Chairman, Jerome Powell's face multiple times a day as he gets lambasted by a CNBC talking head over 'transitory inflation'. If you visit the Federal Reserve website, it is not supposed to be a flashy institution and it historically adheres to a strict dual mandate and mission of "stable prices and maximum sustainable employment". The Fed historically did this through the setting of the Federal Funds Rate, which is the rate Banks can lend to each other overnight, and by communicating their thoughts to the public in Press Releases. Over the last two decades and especially post the 2008 Financial Crisis, the Federal Funds rate has been relatively low and the Federal Reserve has gained the power to buy assets on the open market, known as 'quantitative easing'. The question many will find politicians, media, and business leaders pondering today is whether prices and unemployment are the only parts of the economy the 'Fed' is affecting. To understand if the Fed is influencing more than what they set out to do, we must turn to data analysis.

To find a solution to our problem we compiled data which represented the 'Fed' actions and will be our independent variables. We also assembled various asset price indexes and macroeconomic indicators as dependent variables. We then look to see if the Federal Reserve's actions have predictive power for these dependent variables, and if so which actions. To do this we will model each dependent variable separately using the independent data features. This will give insight into whether the Federal Reserve is causing the next big stock market bubble and subsequent crash, exacerbating wealth inequality or whether they are just keeping the economy humming along.

Data Analysis

Data Description

We collected Federal Reserve Data from Fred (https://fred.stlouisfed.org/), a website run by the Chicago Fed which has thousands of economic indicators and index prices available for download dating back to1903. This data comes seasonally adjusted. The S&P 500 index ETF, 'SPY', data came from Yahoo Finance. The press release language was downloaded directly from the Federal Reserve's website. The complete list of variables appears on the next page. The dataset starts in January 2003 as that is the year the Federal Reserve started recording balance sheet data, giving us 226 samples.

We picked 17 independent variables which we believe encompass the entirety of the Fed's ability to influence the economy and markets. We then picked 17 (just coincidentally) asset price indexes, and Macroeconomic indicators which we think aptly represent the areas the Fed should be influencing and other areas in the United States they could be potentially, but should not. Next we describe how we combined these many different data types, which came in many different scales and frequencies.

<table>
<tr><th>Independent Variables - Labels</th><th>Dependent Variables - Labels</th></tr>
<tr><td>

FedFunds Effective Rate - FEDFUNDS
Federal Reserve Balance Sheet liabilities

Total - RESPPNTNWW
Treasuries

All - TREAST
10 year+ maturity - TREAS10Y
5-10 year maturity - TREAS5T10
1-5 year maturity - TREAS1T5
91 day - 1 year maturity - TREAS911Y

Mortgage Backed Securities

Over 10 year maturity - MBS10Y

Federal Agency debt

All - FEDDT
10+ year maturity - FEDD10Y

Press Release Language

Positive - Positive
Neutral - Neutral
Negative - Negative
Sentiment - Sentiment
Polarity - Polarity
Subjectivity - Subjectivity

M1 Money Supply - M1SL

</td><td>

Asset Prices

SPY ETF S&P 500 index - SPY
US Home Prices - CSUSHPINSA
Wilshire 5000 company index - WILL5000INDFC
Wilshire small cap index - WILLSMLCAP
US Coporate Bond Total Return Index - BAMLCC0ACMTRIV
US Corporate Bond High Yield Total Return Index - BAMLHY0A0HYM2TRIV

Macroeconomic Indicators

Personal Consumption Expenditures - PCEPI
Personal Consumption Expenditures excluding food and energy - PCEPILFE
Unemployment Rate - UNRATE
Gross Domestic Product - GDP
Share of total US wealth held by the top 1% - WFRBST01134

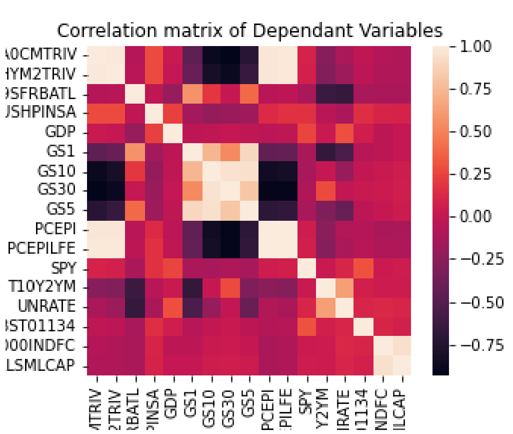
Interest Rates

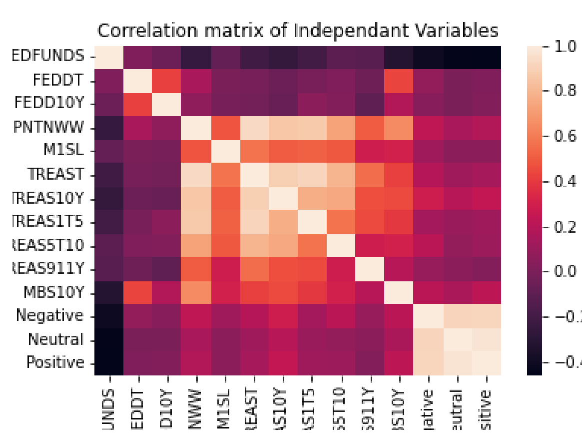
1 Year Treasury yield- GS1
5 Year Treasury yield- GS5
10 year treasury yield -GS10
30 year treasury yield - GS30
2 vs 10 Treasury yield - T202YM

</td></tr>
</table>

Data Collinearity Concerns

Overfitting is a major problem for financial time series and we want to be careful to avoid this issue and create valid models. To get a first peak out of our variables we used Pearson correlation





matrices. For our dependent variables the correlation just indicates the 2 will likely be predicted with the same model and thus are affected by the same fed tools, but does not mean the models are invalid or overfit.

We can see that our yields all move together and there will not be much difference in our models for those variables. Same goes for our inflation indicators and our bond and stock indexes. Outside of each grouping it appears they are not highly correlated, so we will see different actions affecting these variables in different ways. The independent features do not have high inverse correlation in any cases. It is also worth mentioning that we see high correlation in the different sentiment indicators and the balance sheet total with its components, which is to be expected. We will be careful to regularize our data and use cross validation to avoid the pitfalls of collinearity.

Data Cleaning & Organization

Our data was available at various different frequencies, daily, weekly, monthly, and quarterly. We wanted all our data to be at a monthly frequency. This presented a complicated problem as some data needed to be differenced and weekly and quarterly data was not available for the first of every month. For our data where we decided to take the arithmetic change which had weekly Wednesday data, we forward filled all the daily data with the prior totals, took the monthly totals and then took the difference. For data where we wanted the percentage change which either came at a quarterly or monthly interval, we took the difference from the original data and forward filled it for the missing months, such as the case for GDP.

Sentiment Analysis

As mentioned above, the press releases of the Federal Reserve have always been a part of the toolkits available to affect the market. With the prevalence of Natural Language Processing used in investment analysis by asset management firms and hedge funds, the press releases play an increasingly significant role in affecting asset prices.

We capture this effect by first collecting textual data from the Federal Reserve Board website (https://www.federalreserve.gov/). We then use Natural Language Toolkit (nltk), TextBlob and Flair available in Python to process the text and transform it into vectorized numerical data. The package nltk and TextBlob use bag of words, while Flair uses character level neural networks. Both nltk and Flair consider heuristics while TextBlob has the advantage of measuring subjectivity of the speech. Flair, using neural networks, can measure the sentiment of words that were not seen before.

Nltk gives a set of three elements - "Negative", "Positive" and "Neutral" - adding up to one. Textblob gives "Polarity" that captures positivity of the sentiment and "Subjectivity" that captures subjectivity. Flair gives one single measure of sentiment between zero and one.

Missing Data Imputation

One unfortunate fact of the press releases is that it only started publishing at a satisfactory frequency before January 2007. However, other independent variables have complete data points as early as January 2003. We do not wish to discard a large amount of data due to missing data in one feature. Therefore, we use Matrix Completion method implemented in Scikit-learn package to impute missing values in sentiment data between 2003 and 2007.

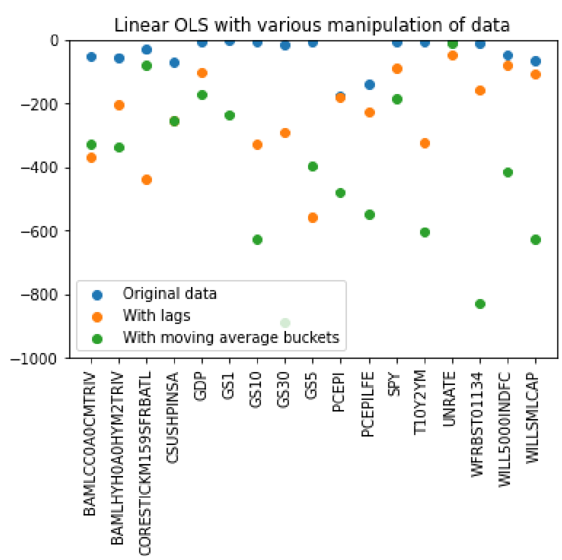
Feature Engineering

The past actions by the Fed are likely to have as much if not more impact than the current actions. We also want the ability to split our data into training, validation, and test sets. We are not especially interested in our ability to predict changes in our dependent variables in the next period, but instead looking to analyze past effectiveness of Fed policies, so we are okay with randomizing our dataset. In order to encapsulate past data in each of our samples, we added moving averages, lags, and lag buckets, averaging various lags into biannual groupings. We also added multiplicative combinations of variables for our linear regression based models such as OLS, Ridge, and Lasso. For Lasso and Ridge, we scaled our independent variables between (-1,1) so that the regularization parameters could work correctly. We tested each model type with and without these added features and noted the tradeoffs between accuracy and interpretability. The smallest feature set being the original 17, and the largest which included all lags and combinations giving 3978 different features.

Model Analysis

OLS

We start with the most basic model of OLS linear regression. We intend to use this as the baseline model and expect other models to at least outperform OLS. To measure the performance of the model, we apply 5-fold cross validation on the data and take the average of the out-of-sample prediction score (which is the out-of-sample R squared). We then tested the model with the original dataset and different sets of added features described in the feature engineering section to see whether using certain added features improves out-of-sample performance. We would be using the same methodology for performance testing of other models in this study.



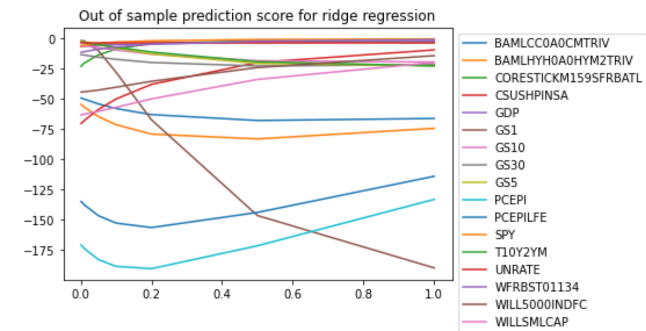
From the plot, we can see that there is no out-of-sample predictability for any dependent variable using any kind of input data.

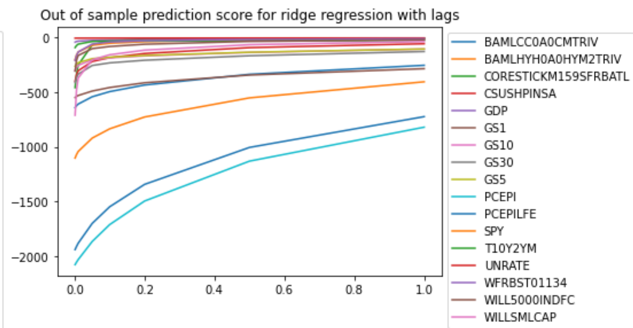
Ridge

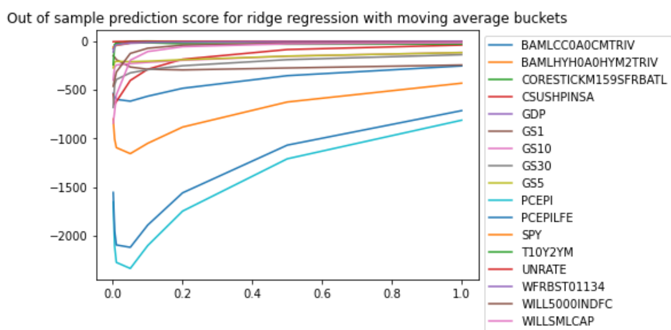
We then move on to test whether using L2 regularizer in ridge regression would improve the out-of-sample performance. With ridge regression we have the regularizer coefficient as a

$$= \underset{\beta \in \mathbf{R}^p}{\operatorname{argmin}} \underbrace{\|y - X\beta\|_2^2}_{\text{Loss}} + \lambda \underbrace{\|\beta\|_2^2}_{\text{Penalty}}$$

hyperparameter, and we would use cross-validation to find the best coefficient. The model evaluation is the same as above.







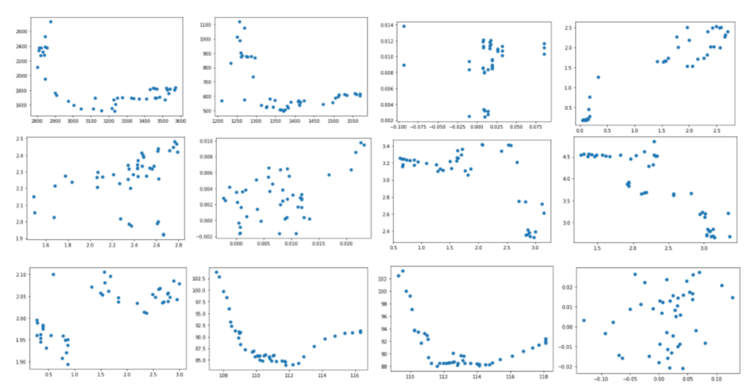
We see from the plots that although the choice of regularizer coefficient has some effect on the out-of-sample performance of the model. However, since none of the models have any out-of-sample predictive power, ridge regression is still ineffective. What is noticeable is that ridge regression has a significant overfitting problem, and the level of overfitting increases when using data with added features. Therefore, overfitting is a major problem and should be addressed in the following models.

Random Forest

Furthermore, we want to explore the possibility of non-linear models to model the relationship. At the same time, we also want to avoid overfitting as much as possible. Therefore, we decided to explore the Random Forest model. With this model, we have maximum depth of each tree (from 5 to 17), minimum ratio of the sample required to have a split (from 0.01 to 0.09) and also the type of loss function as the hyper variables that we can choose (squared or absolute). Through testing, we find that the choice of loss function does not make a material difference on the out-of-sample performance of the model.

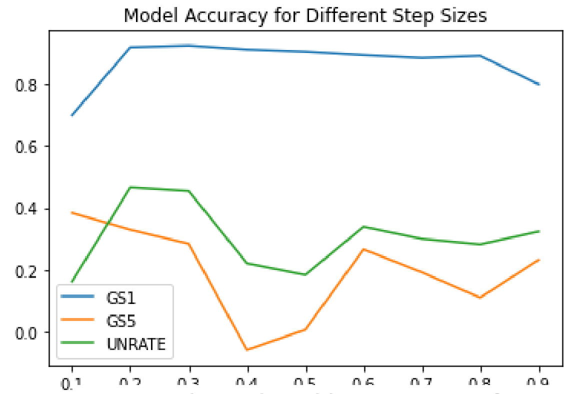
Observing our R-squared with different combinations of hyper parameters we see that the choice of hyper parameters does not affect the out-of-sample performance in any significant way.

When we then look at the plot of predicted versus actual below, we see that the random forest model still does not perform well out-of-sample as these plots do not show a linear relationship between predicted and actual results.

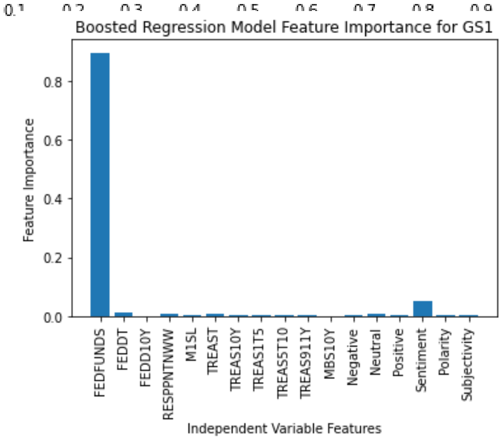


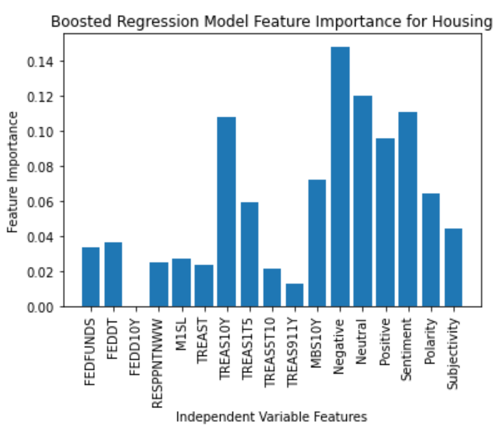
We still need to explore other models to see if we can train models with any predictability.

XGBoost



Another model we implemented for this problem was the XGBoost algorithm. XGBoost is a decision-tree-based ensemble Machine Learning algorithm that uses a gradient boosting framework. When tuning hyper-paramters for our original data set, the changes in number of estimators and max depth did not make much of a change in predictive scores. We then looked at different boosted models when changing the learning rate between 0 and 1. Only 1-Year Treasury, 5-Year Treasury, and Unemployment had positive predictive scores with the 1-Year being predicted quite well. We then looked at the model with a learning rate of 0.3 and examined the Independent feature importance for that model. We see that the most important feature for predicting GS1(1-year Treasury Yield) is the Federal Funds Rate. This makes intuitive sense, and is precisely what the Fed intends when setting the rate.





Next we included moving averages in the data set. After hyper-parameter tuning we found that our best model gave us mostly similar results except for one interesting difference: we achieved an accuracy score of 0.35 for US Housing Prices (CSUSHPINSA). This was the best score thus far for a dependent variable in which the Fed is not intending to affect with their actions. We then observed the feature importance for this model.
Most of the independent features have some weight in the model, negative, neutral and positive sentiment seem to be three of the most important to the model. This conclusion makes some sense in the context of 2008 when the sentiment of the Fed was heavily influenced by the collapse of the housing market. It will be interesting to see as more data becomes available post-2008, if this correlation continues to be present. Compared to the Random Forest models, XGBoost allowed us to predict the 1-Year Treasury Yield, 5-Year Treasury Yield, Unemployment, and Housing prices.

Lasso

The Lasso regression or L1 model seems like the ideal candidate for our problem. We wish to identify the main features in each model and disregard the insignificant ones. Lasso has the useful

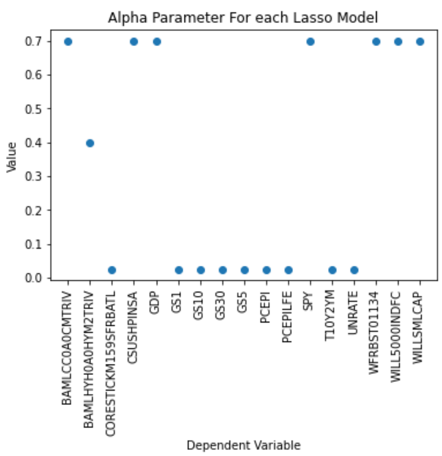
$$\sum_{i=1}^{M}(y_i-\hat{y}_i)^2=\sum_{i=1}^{M}\left(y_i-\sum_{j=0}^{p}w_j\times x_{ij}\right)^2+\lambda\sum_{j=0}^{p}|w_j|$$

Cost function for Lasso regression

property of setting coefficients zero for less predictive features in order to minimize its loss function.

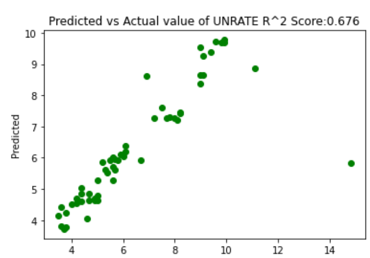
When using the full feature set, we want to avoid the pitfalls of multicollinearity and overfitting we saw in Ridge Regression. We use a 5-fold cross validation on our test set to find our regularization parameter, alpha, for each model. We then train the model on the entire training dataset and find the R-squared value on the test set. The R-Squared scores for the four asset prices, GDP, and wealth inequality were close to 0, but the other 11 had high scores. Lasso zeroed out the majority of the features in each model, but it is still too many. We will analyze the simplest model with predictive power which is the 24 month Moving Average model.

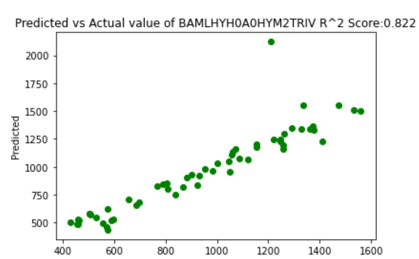
Index	Baseline	MA	MA_Lag_Buckets	ag_buckets_Cor	MA_Mthly_Lags	Mthly_lags_Corr
GS1	0.945655	0.952844	0.952844	0.952847	0.952519	0.952519
BAMLHYH0A0HYM2TRIV	0.421393	0.904185	0.822394	0.866361	0.98686	0.982302
BAMLCC0A0CMTRIV	-4.00273	0.900419	0.832596	0.880844	0.972316	0.984836
PCEPILFE	0.395907	0.864331	0.875074	0.831085	0.979532	0.966946
PCEPI	0.369589	0.847768	0.849033	0.804294	0.975064	0.960599
T10Y2YM	0.395339	0.836517	0.845615	0.836824	0.883761	0.869811
GS30	0.370257	0.807331	0.817096	0.799018	0.883304	0.877621
GS5	0.632962	0.78436	0.801873	0.793909	0.854265	0.849553
GS10	0.476631	0.775595	0.793038	0.780197	0.856977	0.852239
CORESTICKM159SFRBATL	0.277404	0.76513	0.786477	0.754594	0.812629	0.78276
UNRATE	0.464535	0.71262	0.676213	0.646983	0.711624	0.686077

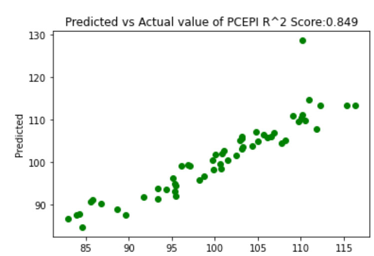


The alphas varied for each dependent model and by feature set, to the left are the Moving Average model alphas. The Lasso cross validation found the best models for lower alpha parameters for variables where prediction was possible, but for the variables Lasso fails to find a useful model, a higher alpha is used as no independent variables are of any use. The exception is in our two Bond indexes, despite their R-squared scores, rely on many correlated variables. We believe there is a variable outside our dataset that our regression model is somehow capturing through a correlation with our data.

We can further see the accuracy of our models by taking a look at a selection of prediction plots.



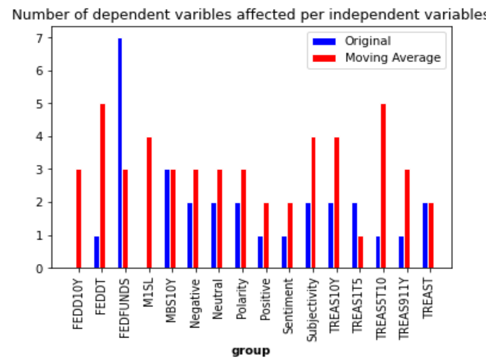




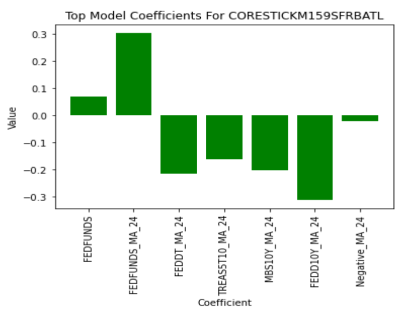
Conclusion/More Lasso Analysis

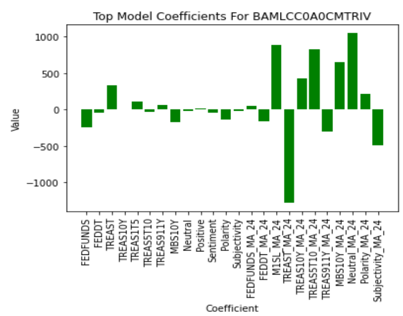
Since we have some accurate models now, let us see what Fed actions are most effective in carrying out their mission. We counted how many of the following dependent variables each independent variable influenced. 1. Unemployment Rate 2. Core Sticky Inflation 3. Personal Consumption Expenditure 4. 1 year yield 5. 10 year yield 6. US Corporate Bond Market 7. 10-2 year yield difference.

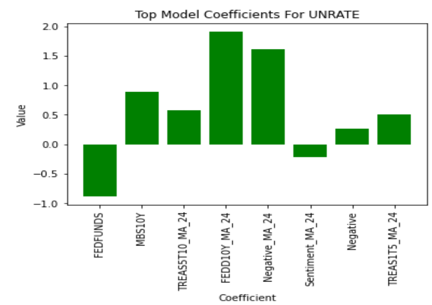
This was done to eliminate duplicative dependent variables, resulting in double counting. The current Fed Funds rate, the original tool, still reigns supreme. Purchases made in both treasuries and federal agency debt showed influence in 5 of the 7 areas the Fed seeks to impact.

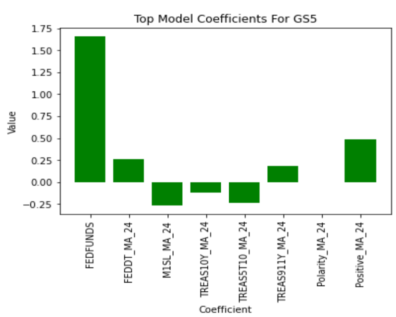


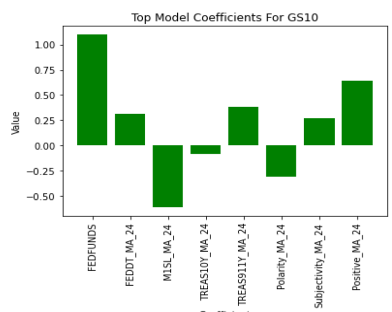
Next we Most exciting are the plots which show the Fed's effective tools for influencing important macroeconomic factors that we could not get from any of our other models.

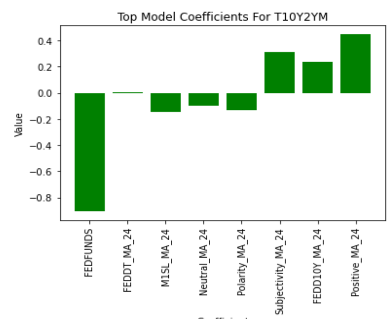












In all we can see that there is a high correlation between past Fed actions, and current macroeconomic factors and bond yields. This is evidenced by the overwhelming presence of moving averages rather than present values. This analysis would be useful for the Fed to know how each action affects the economy. We found no evidence of influencing other areas besides possibly housing prices. We do not see this analysis being used as a weapon of Math destruction as it is a backwards looking tool to help future decision making for the good of our country. This type of hindsight looking analysis could be helpful in a variety of institutions and not just for the Federal Reserve.

Further Research

We were unable to show predictive power using the Fed's tools for our stock market indexes. In the future we would like to investigate whether the direction can be predicted with some accuracy given the Fed's actions. We would like to try this analysis as a classification problem, which should be easier to predict. We also believe that there have been significant changes to the way the market and economy reacts to Fed actions and we might get different results if we split up regimes, but we do not have enough data to do that yet.