What is the Federal Reserve actually doing?

Evan Kurzman
Kelvin Sun
Payton Martin

Motivation & Goals for this Project

- Federal Reserve has gained the power to purchase assets in the open market in addition to its prior ability to set the Federal Funds rate.
- Members also speak often to the press in an effort to carry out their mission.
- Widely debated whether the Fed is fulfilling its dual mandate
    - Stable prices
    - maximum sustainable employment.
- We seek to examine the effects of the Federal Reserve’s specific actions on the macroeconomic indicators they seek to affect, such as unemployment, inflation, and bond yields as well as market prices, such as housing, stocks, and bond market which they do not.

Independent Variables

- FedFunds Effective Rate -FEDFUNDS
- Federal Reserve Balance Sheet liabilities
    - Treasuries
        - 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 - FEDDT
- Press Release Language
    - Positive
    - Negative
    - Polarity
    - Sentiment
    - Subjectivity
- M1 Money Supply - M1SL

Dependant Variables

- 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

Data Cleaning & Feature Engineering

- Data Cleaning
  - Take weekly, daily, and monthly totals from various data sources and transform to be monthly differences indexed by time series
  - Scaled features (-1,1)
- Feature Engineering
  - 24 month moving Averages
  - 24 month lagged variables
  - 6 month moving average buckets
  - Variables Combinations

Natural Language Processing (Sentiment analysis)

- Other than the numeric data points that were commonly used, we are also concerned about the effect of speech by the Federal Reserve.
- We uses the following packages to process text data and apply sentiment analysis:
  - NLTK: Natural Language Tool Kit
  - Textblob
  - Flair

Negative	Neutral	Positive
0.029	0.88	0.091
0.029	0.88	0.091
0.049	0.839	0.112
0.049	0.839	0.112
0.054	0.834	0.112
0.049	0.847	0.104
0.049	0.847	0.104
0.043	0.827	0.13
0.062	0.821	0.117
0.053	0.824	0.123
0.053	0.824	0.123
0.065	0.815	0.119
0.047	0.854	0.099
0.047	0.854	0.099
0.063	0.813	0.124
0.059	0.823	0.118
0.059	0.823	0.118
0.062	0.81	0.128

NLTK's sentiment analysis tool uses a bag of words approach with some simple heuristics

- Increases intensify for words such as "really", "so" and flip the sign for words like "not".

The sentiment is split into three numbers representing negative neutral and positive. They add up to 1.

Polarity	Subjectivity
0.0530024	0.351518
0.0530024	0.351518
0.0548236	0.381281
0.0548236	0.381281
0.0737571	0.407863
0.0619957	0.394408
0.0619957	0.394408
0.0551906	0.381322
0.0635065	0.391904
0.0598768	0.364245
0.0598768	0.364245
0.0587127	0.356272
0.0612666	0.336138
0.0612666	0.336138
0.0797143	0.372482
0.0575163	0.355922
0.0575163	0.355922
0.0738622	0.373302

Similarly, TextBlob also uses a bag of words but it did not contain heuristics.

It has the added advantage of measuring subjectivity.

Polarity measures the positivity of the speech, where 0 represents the highest level of negativity and 1 represents the highest level of positivity.

Subjectivity is also between 0 and 1. The higher the number, the more subjective the speech is.

Sentiment
0.672449
0.672449
0.729134
0.729134
0.560768
0.682006
0.682006
0.697426
0.706971
0.892929
0.892929
0.897204
0.938606
0.938606
0.897392
0.908461
0.908461
0.933176

Flair’s sentiment classifier is based on a character-level LSTM neural network, which takes sequences of letters and words into account when predicting

It also takes into account heuristics such as intensifiers and negations.

It has the greatest advantage in being able to predict sentiment that has never been seen before.

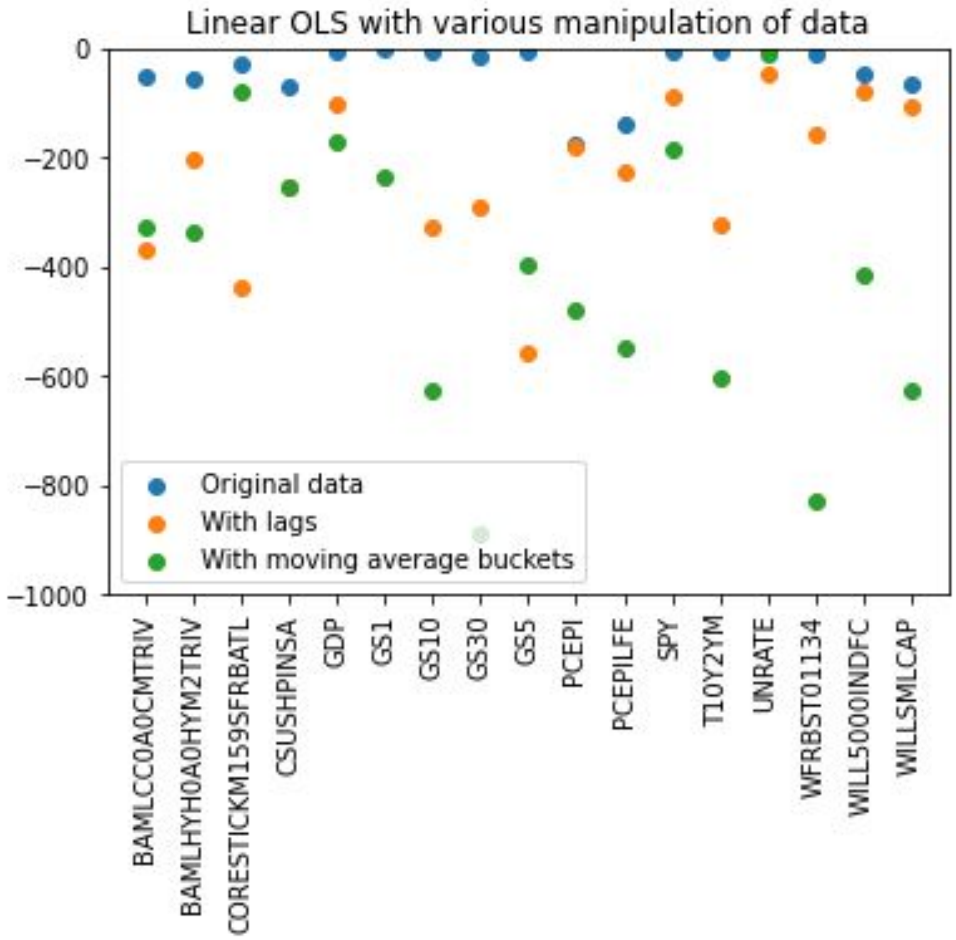
The score is between 0 and 1. The higher the score, the more positive the speech is.

Filling Missing Values - Matrix Completion

- For Federal Reserve press releases, the earliest we can obtain is from January, 2007.
- However, other independent variables goes as early as 2003.
- We do not want to discard many data points due to missing data in language sentiment.
- Therefore, we use matrix completion to impute the missing data points.

OLS Regression Models

- We start with the most generic model of OLS regression.
- To assess the out-of-sample performance of the model, we apply 5-fold cross validation and take the average of the average R-square score.
- We also apply the model to:
  - Original dataset
  - Dataset with lags and moving averages of independent variables
  - Dataset with buckets of moving averages
- We assess the model with the average of R-squared scores of each fold of cross validation.



None of the models for any dependent variable, with any manipulation of the dataset, has above 0 prediction score.

There is no predictive power for out-of-sample performance.

Lasso Regression Models

- Why Lasso?
  - Features have high Collinearity
  - Want to find most important features and ignore others
- Tried different subsets of features to improve interpretability.
- Used 5 fold Cross Validation on our training set to find the best alpha parameter for each dependent variable model.
- R^2 values were mostly better for all models with more available features.
- Did not get much worse by giving less available features, but made it easier to identify important features.

Alphas for each model

BAMLCC0A0CMTRIV	float64	1	0.7
BAMLHYH0A0HYM2TRIV	float64	1	0.4
CORESTICKM159SFRBATL	float64	1	0.025
CSUSHPINSA	float64	1	0.7
GDP	float64	1	0.7
GS1	float64	1	0.025
GS5	float64	1	0.025
GS10	float64	1	0.025
GS30	float64	1	0.025
PCEPI	float64	1	0.025
PCEPILFE	float64	1	0.025
SPY	float64	1	0.7
T10Y2YM	float64	1	0.025
UNRATE	float64	1	0.025
WFRBST01134	float64	1	0.7
WILL5000INDFC	float64	1	0.7
WILLSMLCAP	float64	1	0.7

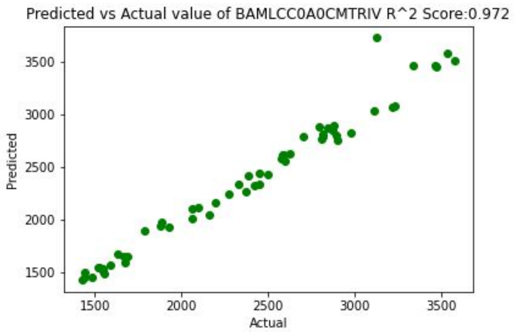
R^2 Values for each model

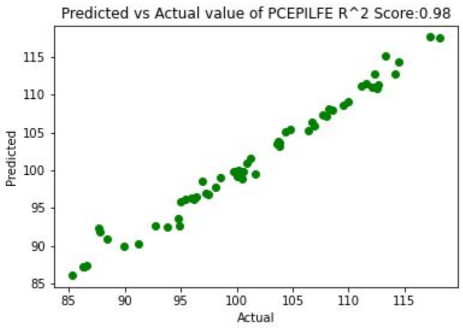
Index	Baseline	MA	MA_Lag_Buckets	ag_buckets_Cor	MA_Mthly_Lags	Mthly_lags_Corr
GS1	0.945655	0.952844	0.952844	0.952847	0.952519	0.952519
BAMLHYH0A0HYM2TRIV	0.421393	0.904185	0.822394	0.866361	0.98686	0.982302
BAMLCC0A0CMTRIV	-4.00273	0.900419	0.832596	0.880844	0.972316	0.984836
PCEPILFE	0.395907	0.864331	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
SPY	-3.06238e-05	-3.06238e-05	-3.06238e-05	-3.06238e-05	-3.06238e-05	-3.06238e-05
WILL5000INDFC	-6.66415e-05	-6.66415e-05	-6.66415e-05	-6.66415e-05	-6.66415e-05	-6.66415e-05
WILLSMLCAP	-0.00137921	-0.00137921	-0.00137921	-0.00137921	-0.00137921	-0.00137921
CSUSHPINSA	-0.00748421	-0.00748421	-0.00748421	-0.00748421	-0.00748421	-0.00748421
GDP	-0.0500182	-0.0500182	-0.0500182	-0.0500182	-0.0500182	-0.0500182
WFRBST01134	-0.0752394	-0.0752394	-0.0752394	-0.0752394	-0.0752394	-0.0752394

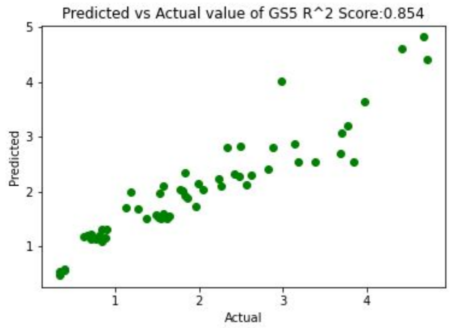
Most effective Fed tools

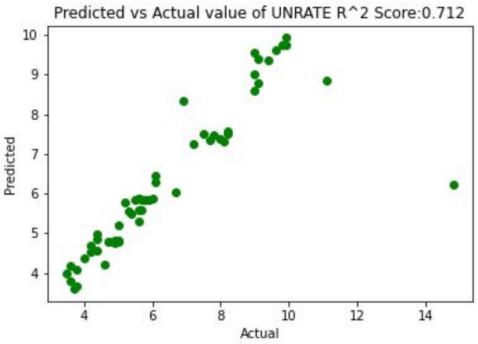
Index	LCC0A0CM	YH0A0HYI	TICKM159SF	GS1	GS10	GS30	GS5	PCEPI	PCEPILFE	T10Y2YM	UNRATE
M1SL_MA_24	954.467	502.292	nan	nan	-0.613162	-0.98648	-0.268289	17.2526	18.008	-0.172419	nan
Polarity_MA_24	244.944	123.121	nan	nan	-0.308284	-0.466792	-0.000221096	nan	nan	-0.202344	-0.809263
Negative_MA_24	-726.954	-467.791	-0.151643	nan	nan	0.0108929	nan	-12.184	-12.1603	nan	1.54749
FEDDT_MA_24	-288.423	-160.354	-0.338073	nan	0.315677	0.0766173	0.261357	-3.4583	-2.22646	0.141896	0.69354
Sentiment_MA_24	nan	nan	nan	0.0555289	nan	0.307756	nan	nan	nan	nan	-0.395815
TREAS911Y_MA_24	-648.653	-268.872	nan	nan	0.382973	0.484055	0.181308	-13.0414	-11.6745	nan	-0.159459
FEDFUNDS	-297.175	-125.862	0.0134343	2.35339	1.09805	0.539339	1.66165	-4.06077	-4.06505	-0.922218	-0.965844
Positive_MA_24	-656.906	-270.86	-0.0132044	nan	0.642757	0.652963	0.481519	-7.78564	-7.42733	0.61338	0.0713577
Subjectivity_MA_24	-550.043	-333.512	nan	nan	0.271741	0.700738	nan	-8.22276	-7.49124	0.24254	nan
TREAS10Y_MA_24	nan	3.89762	nan	-0.0690358	-0.0836411	nan	-0.114453	nan	nan	nan	-0.554401
FEDFUNDS_MA_24	135.447	87.9172	0.400209	nan	nan	nan	nan	1.68567	0.984491	nan	-0.502287
Negative	14.4861	nan	nan	nan	nan	nan	nan	0.535498	0.310364	nan	0.201799
MBS10Y_MA_24	822.889	498.97	-0.227649	nan	nan	nan	nan	10.9345	9.45012	nan	0.214829
TREAS1T5_MA_24	-578.795	-444.834	nan	nan	nan	nan	nan	-7.38757	-8.10819	nan	0.595631
MBS10Y	-185.331	-105.213	nan	nan	nan	nan	nan	-2.15454	-1.85879	nan	0.650571
TREAS5T10_MA_24	493.256	309.733	-0.01075	-0.0698587	nan	nan	-0.236758	5.93648	4.85787	nan	0.803151

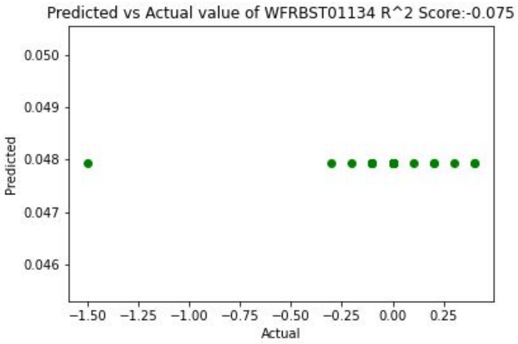
Predicted Vs Actual

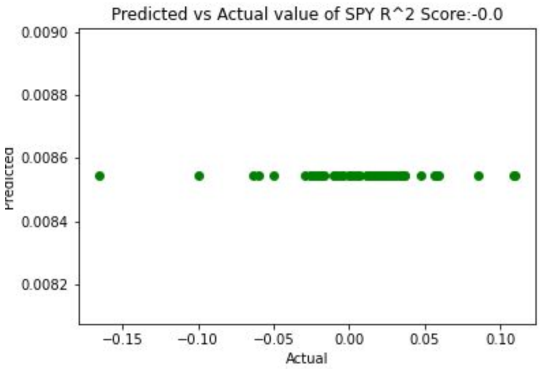






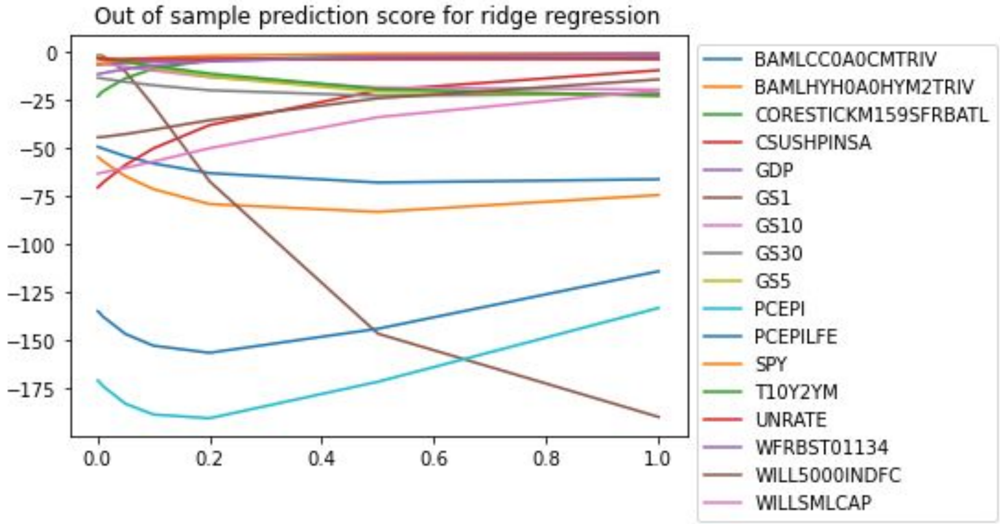


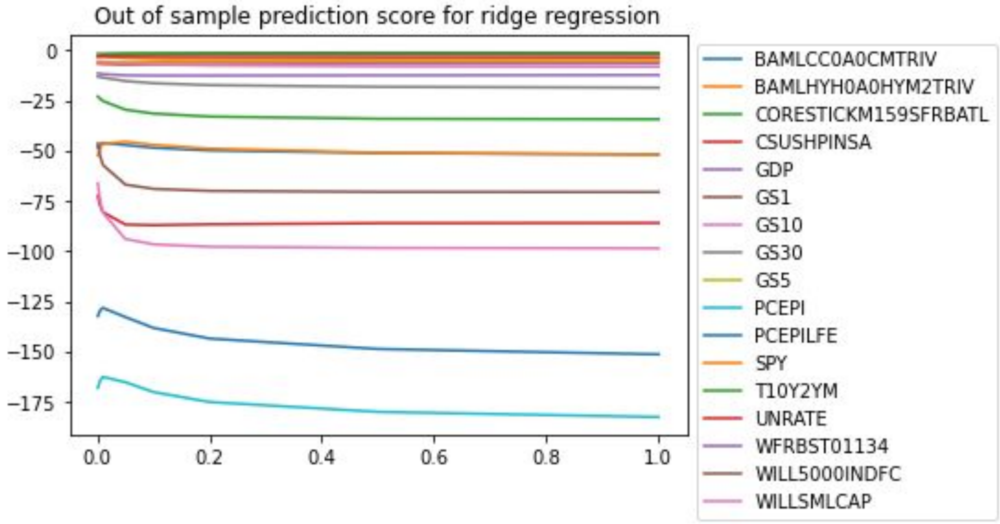




Ridge Regression Models

- With ridge regression, we have the choice of the regularizer coefficient as a hyper-parameter.
- We want to test different choice of the hyper-parameter and find the best model.
- We also compare original and regularized models, since scaling affects the outcome of ridge regression.
- The evaluation of the performance is the same, with cross-validation and average prediction score.

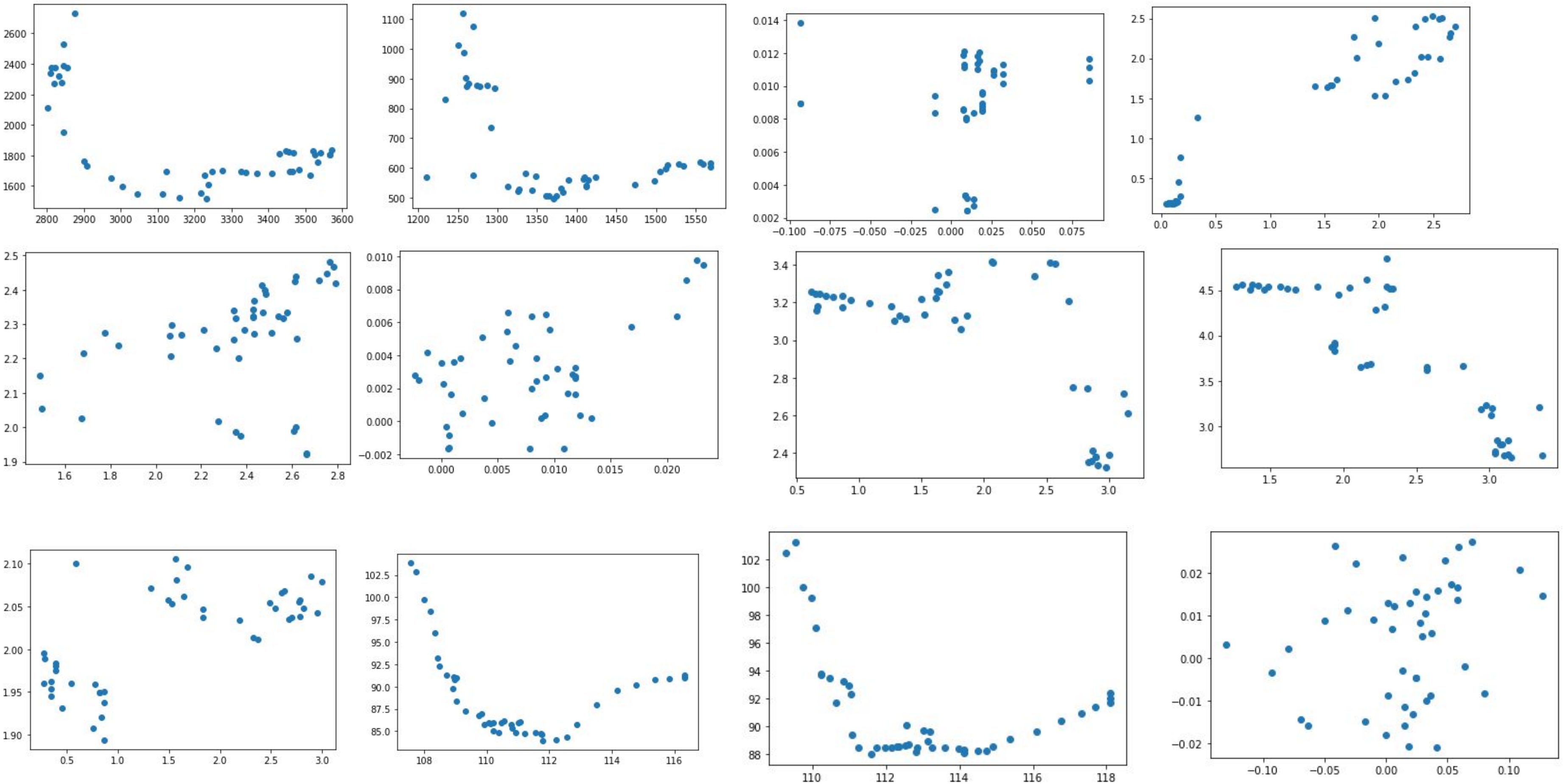


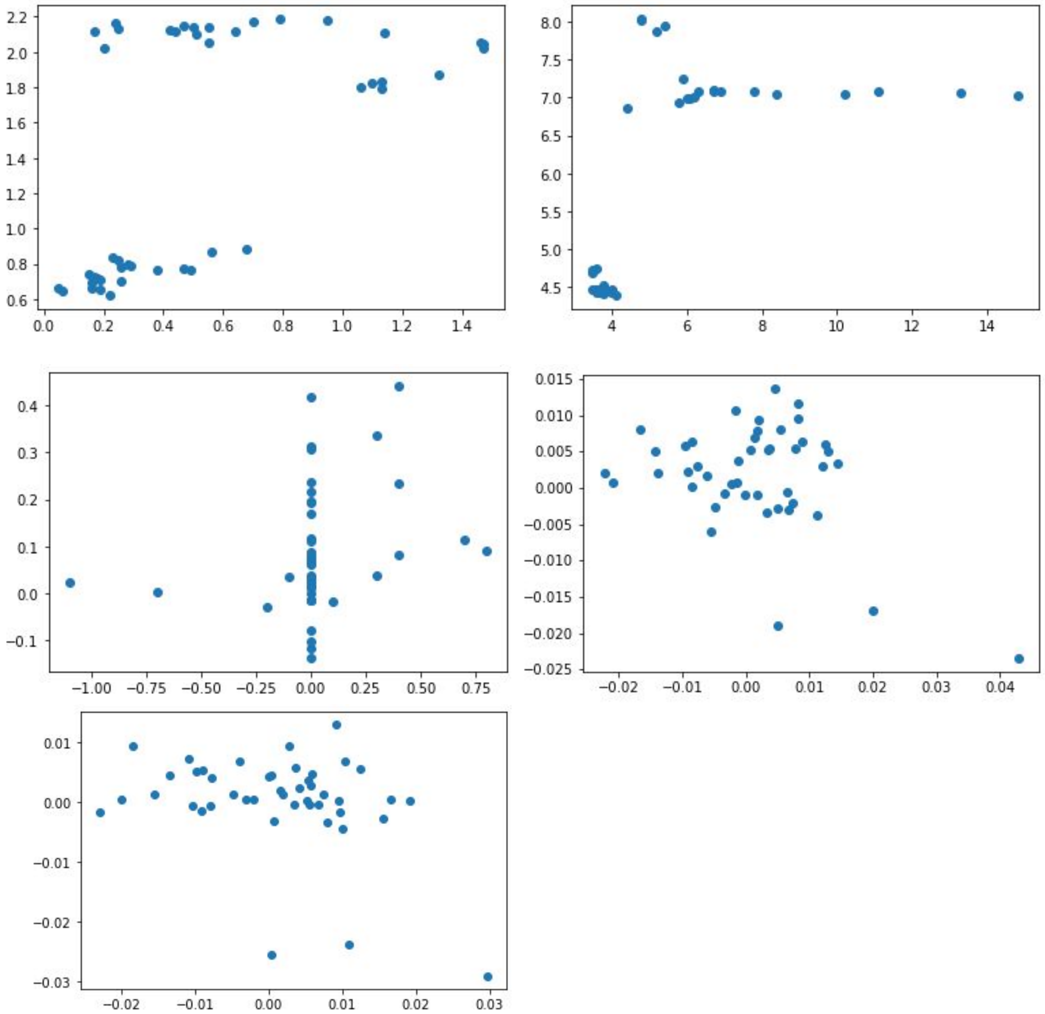


- The top plot consists of models with normalization, while the bottom plot has models without normalization.
- Models, just like OLS, has no out-of-sample predictive power.
- The performance is worse if the original dataset is manipulated with lags and moving averages.
    - In particular, the in-sample overfitting is worse.

Random Forest Model

- We also want to test models with decision trees and ensemble methods.
- With Random Forest, we have the following hyper-parameters:
  - Max depth of each tree (varying from 5 to 17)
  - Minimum ratio of samples required to have a split (varying from 0.01 to 0.1)
  - Loss function:
    - Squared error
    - Absolute error
- Again, we test with different manipulation of dataset, 5-fold cross validation and take the average prediction score.


2600 2400 2200 2000 1800 1600
2800 2900 3000 3100 3200 3300 3400 3500 3600
1100 1000 900 800 700 600 500
1200 1250 1300 1350 1400 1450 1500 1550
0.014 0.012 0.010 0.008 0.006 0.004 0.002
−0.100 −0.075 −0.050 −0.025 0.000 0.025 0.050 0.075
2.5 2.0 1.5 1.0 0.5
0.0 0.5 1.0 1.5 2.0 2.5
2.5 2.4 2.3 2.2 2.1 2.0 1.9
1.6 1.8 2.0 2.2 2.4 2.6 2.8
0.010 0.008 0.006 0.004 0.002 0.000 −0.002
0.000 0.005 0.010 0.015 0.020
3.4 3.2 3.0 2.8 2.6 2.4
0.5 1.0 1.5 2.0 2.5 3.0
4.5 4.0 3.5 3.0
1.5 2.0 2.5 3.0
2.10 2.05 2.00 1.95 1.90
0.5 1.0 1.5 2.0 2.5 3.0
102.5 100.0 97.5 95.0 92.5 90.0 87.5 85.0
108 110 112 114 116
102 100 98 96 94 92 90 88
110 112 114 116 118
0.02 0.01 0.00 −0.01 −0.02
−0.10 −0.05 0.00 0.05 0.10



- Above are plots of predicted dependent variables against actual.

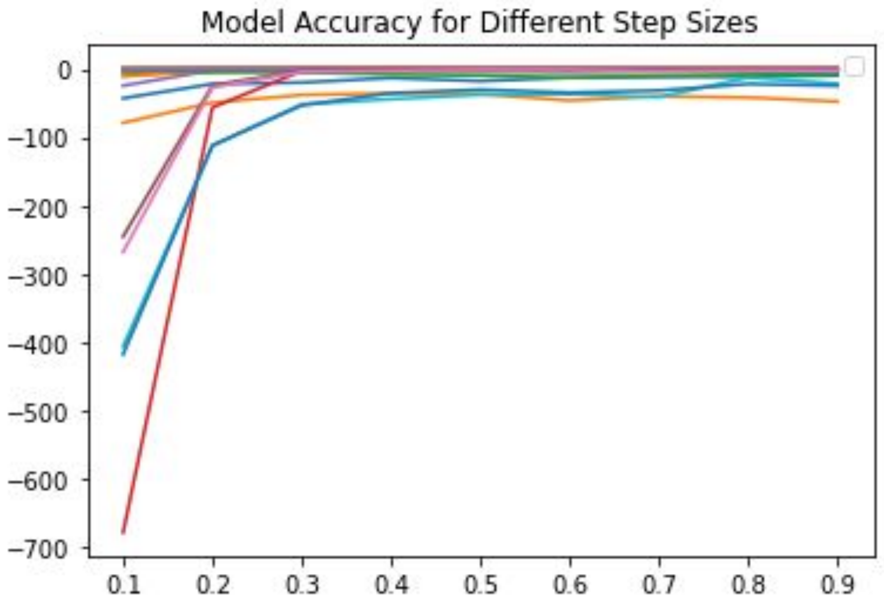
- We see that for most dependent variables there does not seem to have any consistent out-of-sample performance.

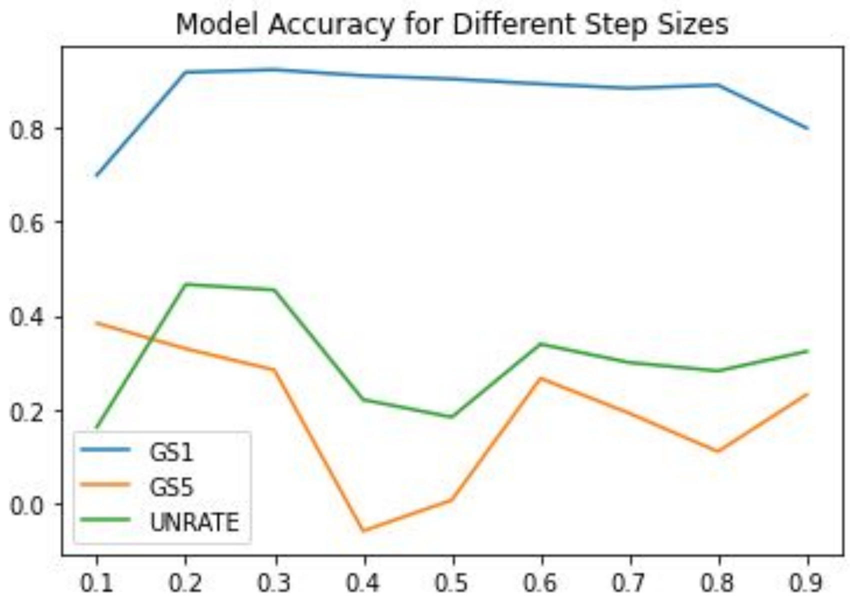
	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
5	-261.147	-269.096	-267.83	-274.296	-274.226	-260.588	-263.188	-269.134	-272.009
6	-265.342	-269.291	-276.719	-269.35	-270.428	-260.352	-263.128	-268.678	-270.714
7	-268.388	-261.415	-273.772	-270.069	-271.86	-256.662	-262.247	-268.186	-270.136
8	-266.073	-258.283	-280.27	-270.934	-273.086	-257.937	-262.427	-268.408	-270.15
9	-272.186	-261.144	-282.073	-272.5	-272.554	-257.862	-262.427	-268.408	-270.15
10	-271.873	-255.313	-281.494	-271.814	-272.554	-257.862	-262.427	-268.408	-270.15
11	-270.513	-256.334	-281.096	-271.814	-272.554	-257.862	-262.427	-268.408	-270.15
12	-269.763	-256.043	-281.096	-271.814	-272.554	-257.862	-262.427	-268.408	-270.15
13	-270.333	-256.043	-281.096	-271.814	-272.554	-257.862	-262.427	-268.408	-270.15
14	-270.286	-256.043	-281.096	-271.814	-272.554	-257.862	-262.427	-268.408	-270.15
15	-270.286	-256.043	-281.096	-271.814	-272.554	-257.862	-262.427	-268.408	-270.15
16	-270.286	-256.043	-281.096	-271.814	-272.554	-257.862	-262.427	-268.408	-270.15
17	-270.286	-256.043	-281.096	-271.814	-272.554	-257.862	-262.427	-268.408	-270.15

- The last slide is a table of predictive scores with varying hyper-parameters.
- We see that the performance of the model does not vary significantly with changes in hyper-parameters.
- This is consistent for all dependent variables.

Boosted Regression (XGBoost)

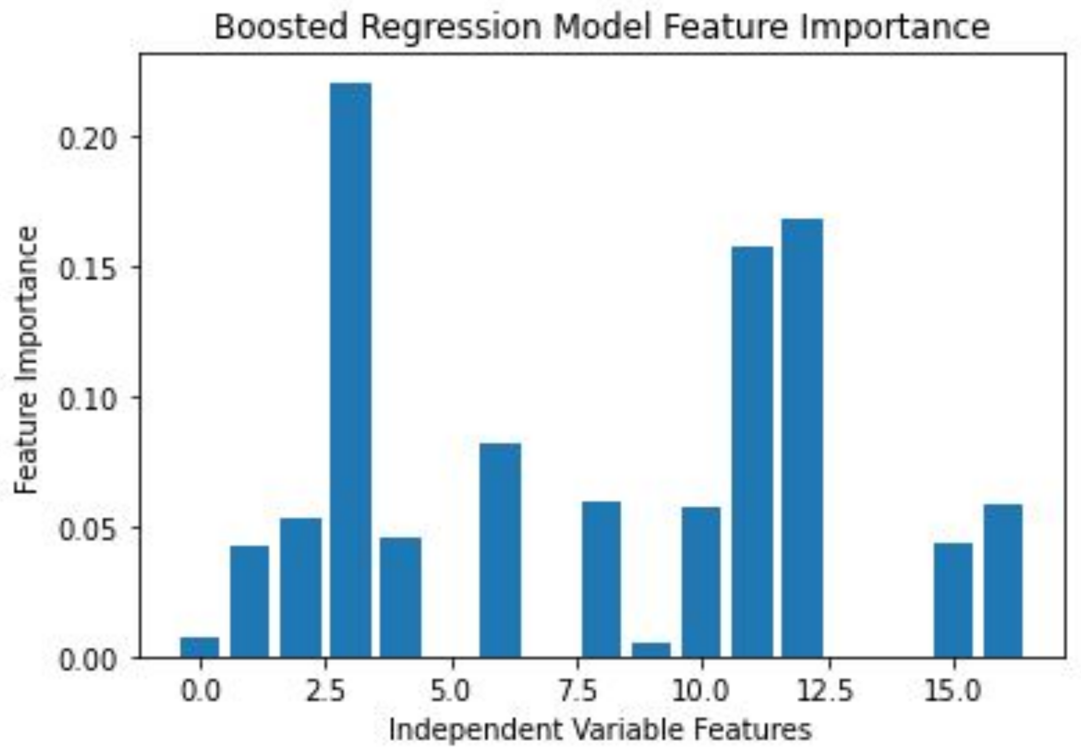
- We also wanted to fit a more flexible model to our data and settled on boosted regression.
- We first trained the model iterating over multiple values of the hyper-parameters max depth and number of estimators and found that the mean-squared error did not change very much between the different hyper-parameters.
- Given this and for simplicity we trained our model over 9 different learning rates to compare which model performed the best.





- Not very helpful
- Shows that the majority of the dependent variables were not well predicted by the models.

- Top three features
- 1-Year Treasury Yield (GS1)
- 5-Year Treasury Yield (GS5)
- Unemployment Rate (UNRATE)



- Best model was LR=0.3
- Top three Independent Variable Features:
  - Total Fed Balance Sheet (RESPPNTNWW)
  - Neutral Sentiment (Neutral)
  - Negative Sentiment (Negative)

Conclusions

- Federal Reserve is not affecting asset prices besides fixed income securities.
- Almost all of the Federal Reserve's tools appear to be having impact in the areas they are supposed to.
    - Unemployment
    - Inflation
    - Bond yields
- Past actions have more impact than current actions.
- The Fed appears to be fulfilling their dual mandate without collateral damage.
- Machine Learning can be effective in analyzing the success of opaque government entities.
- Lasso is far superior for questions of this type.

Future Work

- We believe there is still an abundance of future work available on the topic of the effects the Federal Reserve has on macroeconomic indicators.
- A specific area of research we did not cover in this project is the effects the Fed has on the global economy.
- We may also look like predictability of asset prices if we consider asset returns as up or down classifiers
  - This mean allow for better pre
- It may also be worth repeating this study in the future as more data becomes available and the scope of the Fed's actions widen.
  - Specifically, many believe that there are fundamental changes to the market and the actions of Federal Reserve after the Global Financial Crisis. We may be able to find models with more significant relationships looking at only post-GFC data. However, the time horizon is not long enough.