US FOMC Communication Interest Rate Forecaster

—

Shawn Mitchell MBA

Introduction

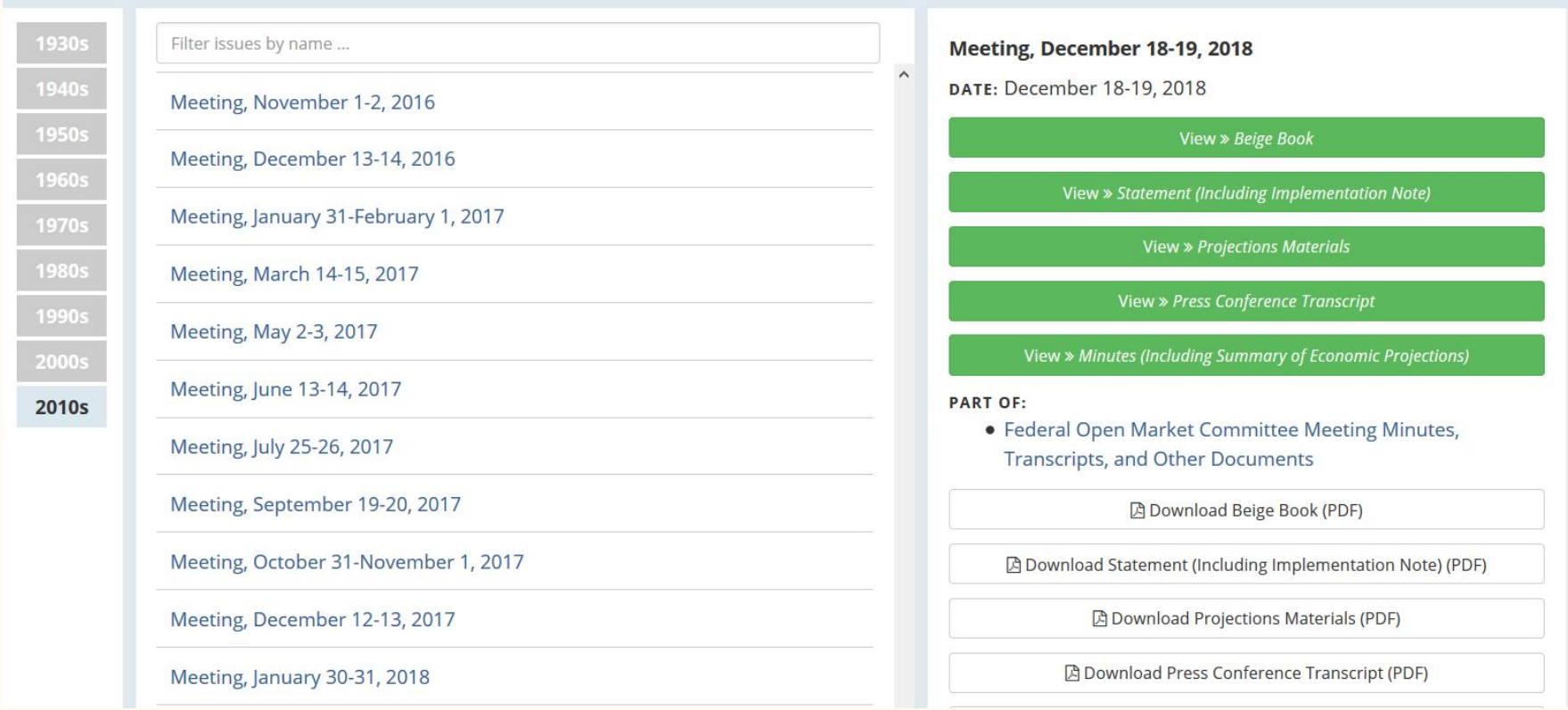
1. US interest rates follow Federal Reserve guidance, but vary around it
2. USD is the global reserve currency:
    a. Virtually all international trade is in USD
    b. US government treasuries are considered the most “safe” asset
    c. All global governments hold significant USD reserves
3. Predicting US interest rates is one of the most difficult financial activities
4. US interest rate changes are closely linked with economic cycles

Executive Summary

Using US Federal Reserve Open Market Committee public communications, we can forecast with approximately 65% accuracy the following 6-month interest rate sentiment. This is based on using FOMC communications since 1960, and allowing the model to be trained on a random 75% training sample. Removing portions of data based on date significantly reduces future model accuracy, as the features of the communications changes over time, and the economic/political environment changes over time. When used with caution, an NLP model using FOMC communications as features can be a useful supplemental tool for interest rate forecasting.

Data Science question:
Can NLP be used to forecast US rate changes with a useful level of accuracy?

Data Source - FOMC Communications



Data Source - FOMC Communications

Which documents to use?

1. Manual process of scripting specific documents types to use for all months
    a. Time consuming, possibly error-prone
2. Use all documents
    a. Data imbalance, poor quality data, false features
3. Use a consistent pattern to pull a sub-selection for each month
    a. Document type changes

Data Source - FOMC Communications

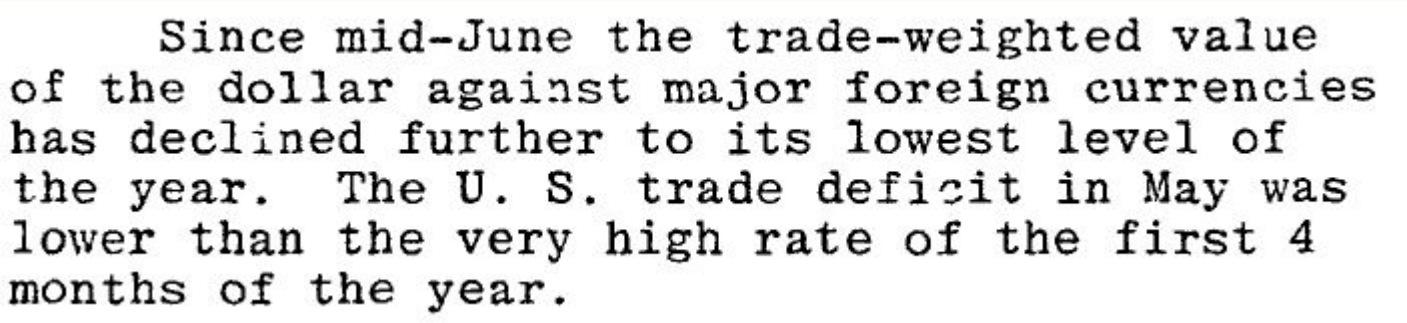
Problems with using the largest file:

1. Largest file may not have the most text, some files may be image-heavy
2. Largest file may be a section from a multi-part communication document intended to be read as a single whole
3. What the largest file is can vary month to month. Sometimes it may be a summary of current economic conditions, other months it may be a transcript of FOMC meeting minutes

Data Source - FOMC Communications

Problem: Text Extraction from PDF files

Typos, word combinations, non-communication text, document descriptions and other non-valuable text strings.



Since mid-June the trade-weighted value of the dollar against major foreign currencies has declined further to its lowest level of the year. The U. S. trade deficit in May was lower than the very high rate of the first 4 months of the year.

Data Source - Interest Rates

US Federal Reserve FRED API

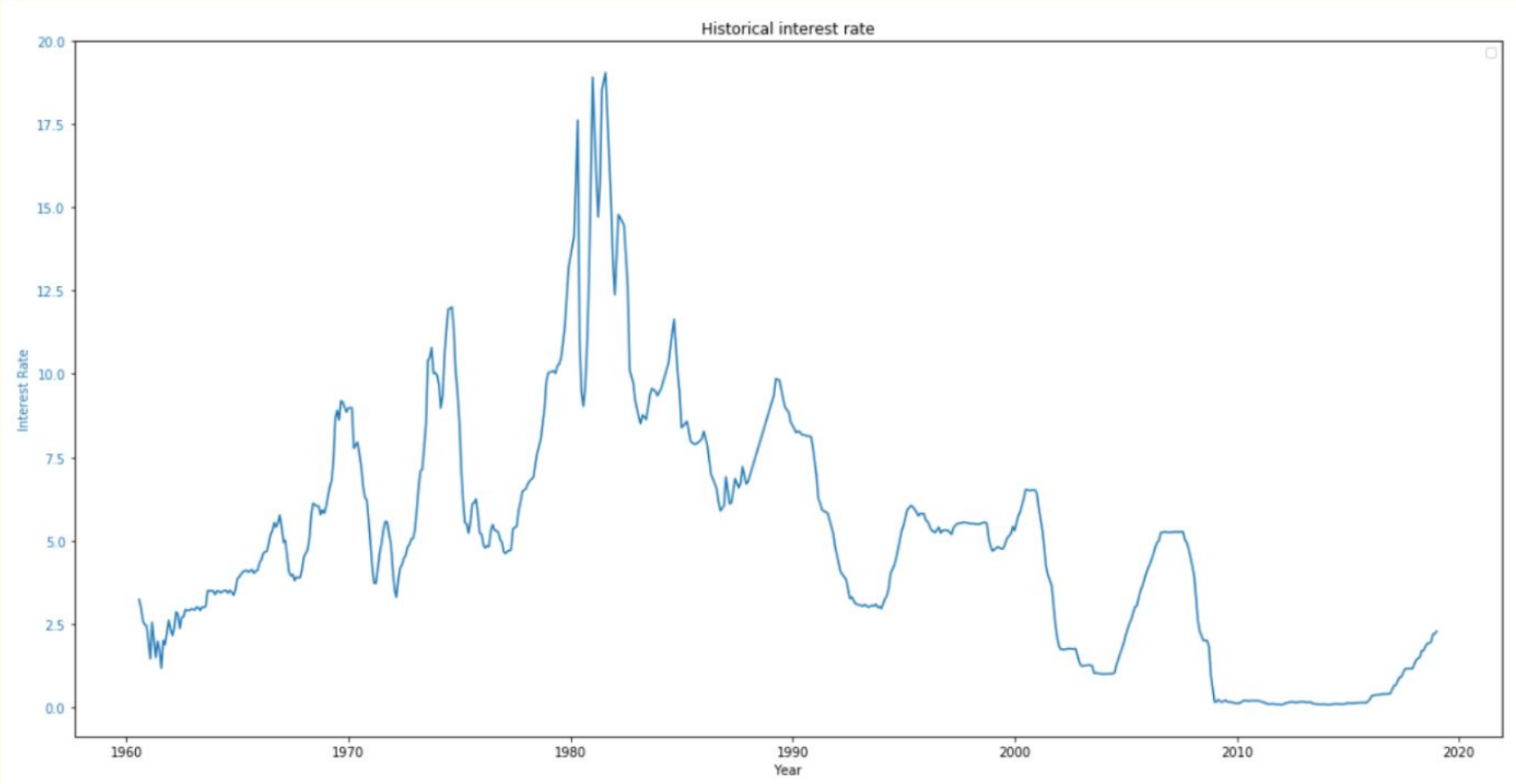
Large source of global information continually updated by the St. Louis Federal Reserve.

Interest Rate Target

1. Exact interest rate target too difficult to forecast
2. Target must be actionable
3. Sentiment analysis needs a reasonable threshold

date	text	rate	change_6
1960-02-01	A meeting of the Federal Open Market Committee...	3.991935	NaN
1960-03-01	A meeting of the Federal Open Market Committee...	3.965517	NaN
1960-04-01	A meeting of the Federal Open Market Committee...	3.838710	NaN
1960-05-01	A meeting of the Federal Open Market Committee...	3.916667	NaN
1960-06-01	A meeting of the Federal Open Market Committee...	3.846774	NaN
1960-07-01	A meeting of the Federal Open Market Committee...	3.316667	NaN
1960-08-01	A meeting of the Federal Open Market Committee...	3.225806	-0.766129

date	change_6
1960-02-01	-1.0
1960-03-01	-1.0
1960-04-01	-1.0
1960-05-01	-1.0
1960-06-01	-1.0

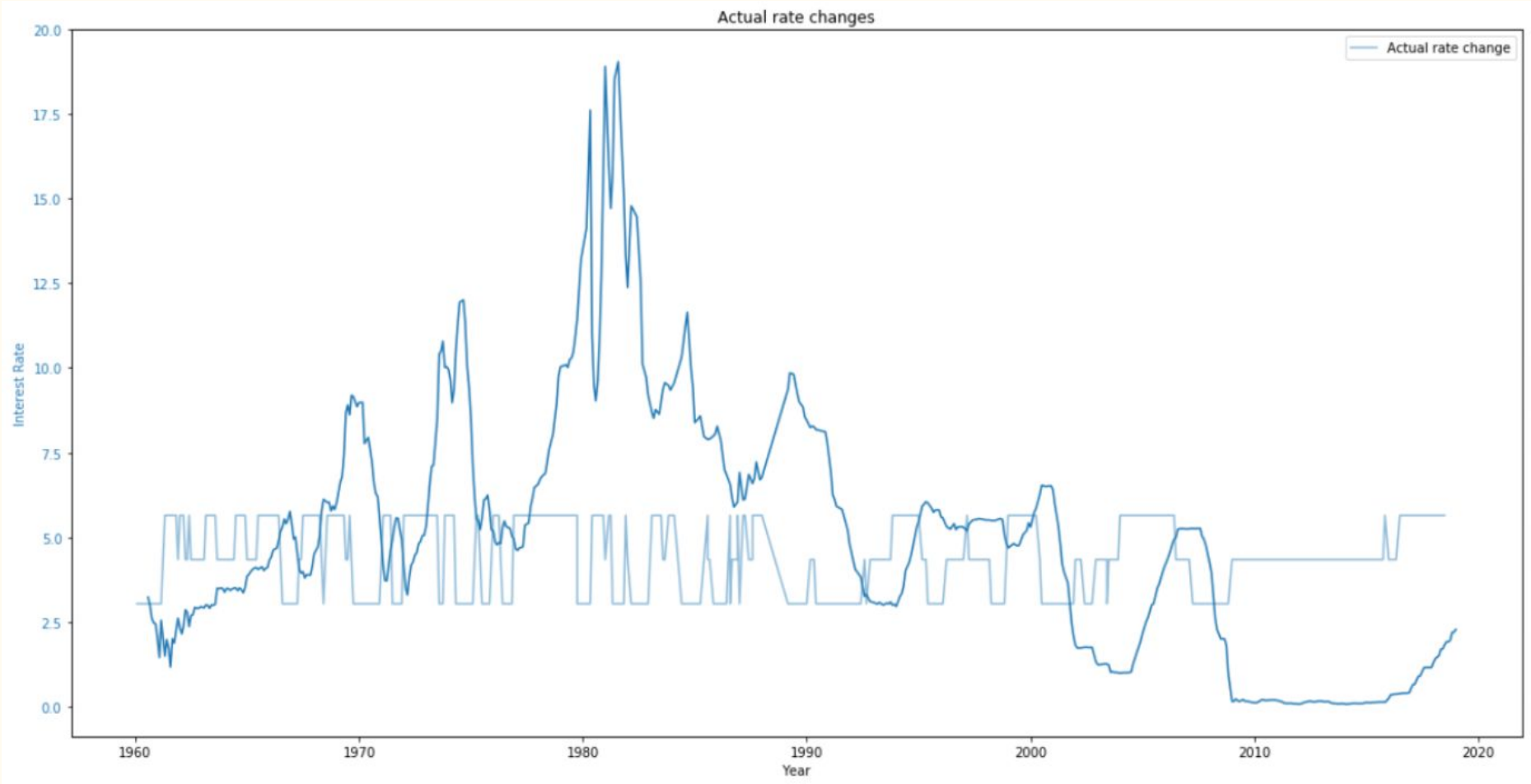

Historical interest rate
Interest Rate
20.0
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0.0
1960
1970
1980
1990
2000
2010
2020
Year

Modelling

Recap:

Observations - FOMC public communications from 1960 to present. The largest by file size communication per month will represent each month.

Target predicted value - 6-month sentiment. The sentiment following the communication will be positive, neutral, or negative depending if the interest rate 6 months from the communication was over/under/within the threshold (0.25%).


Actual rate changes
Actual rate change
Interest Rate
Year
20.0
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0.0
1960
1970
1980
1990
2000
2010
2020

Model Performance - Train/Test Score

Overfitting is a problem

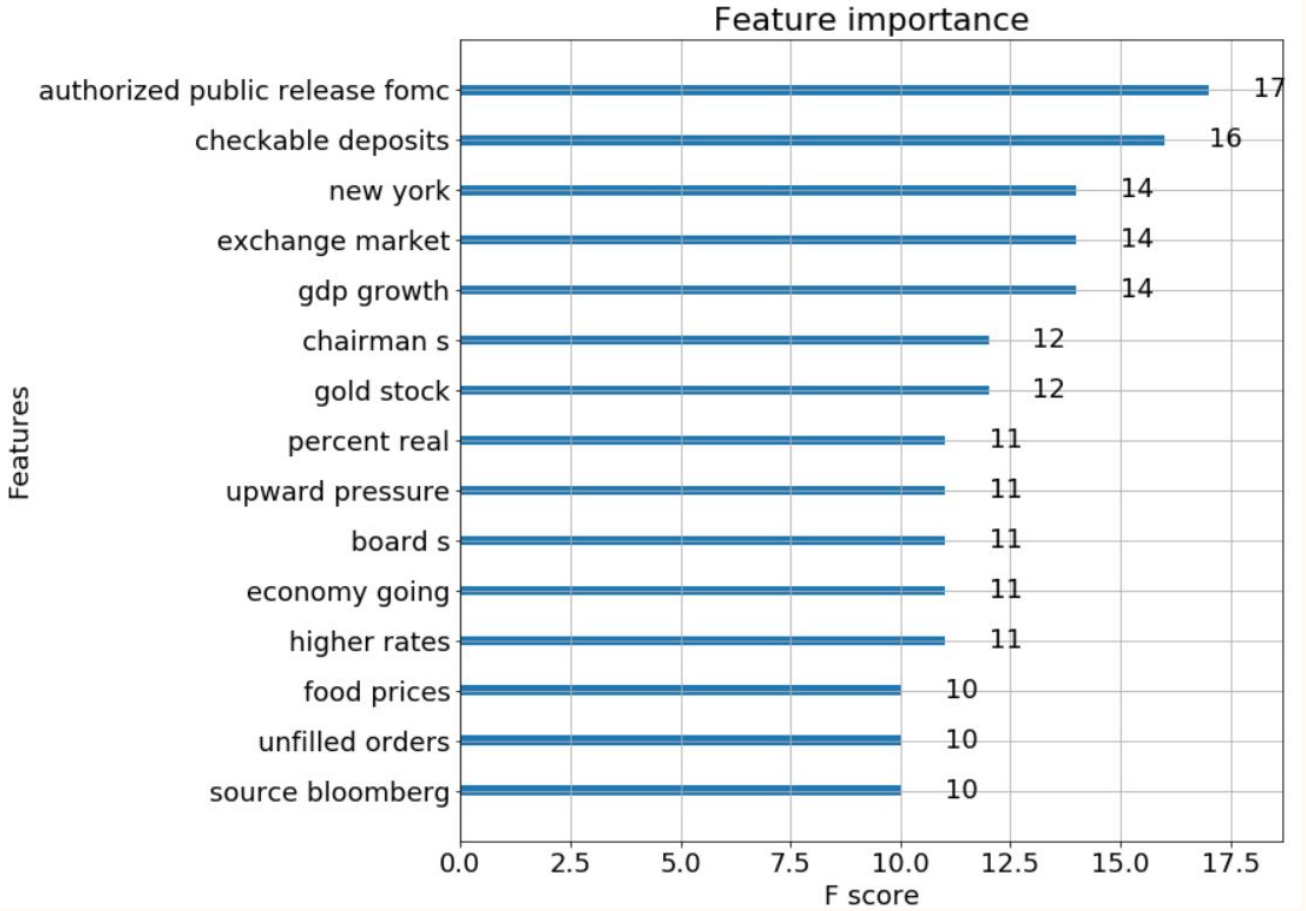
65% accuracy on the test set

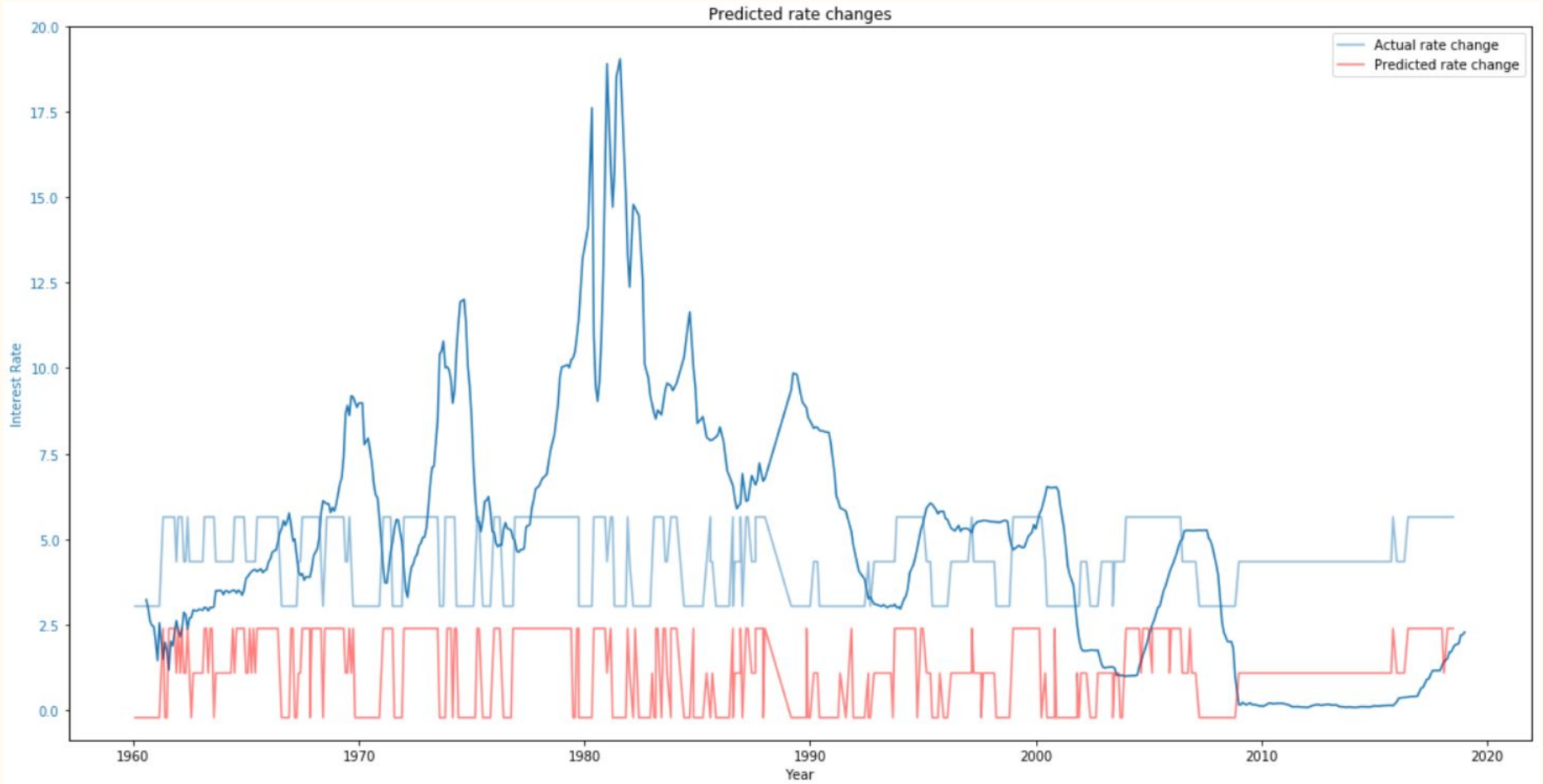
38% accuracy is the baseline (largest class of the data)

```
Run 1... train: 1.0, test: 0.66
Run 2... train: 1.0, test: 0.68
Run 3... train: 1.0, test: 0.62
Run 4... train: 1.0, test: 0.66
Train average: 1.0
Test average: 0.655
```

Mostly logical features:

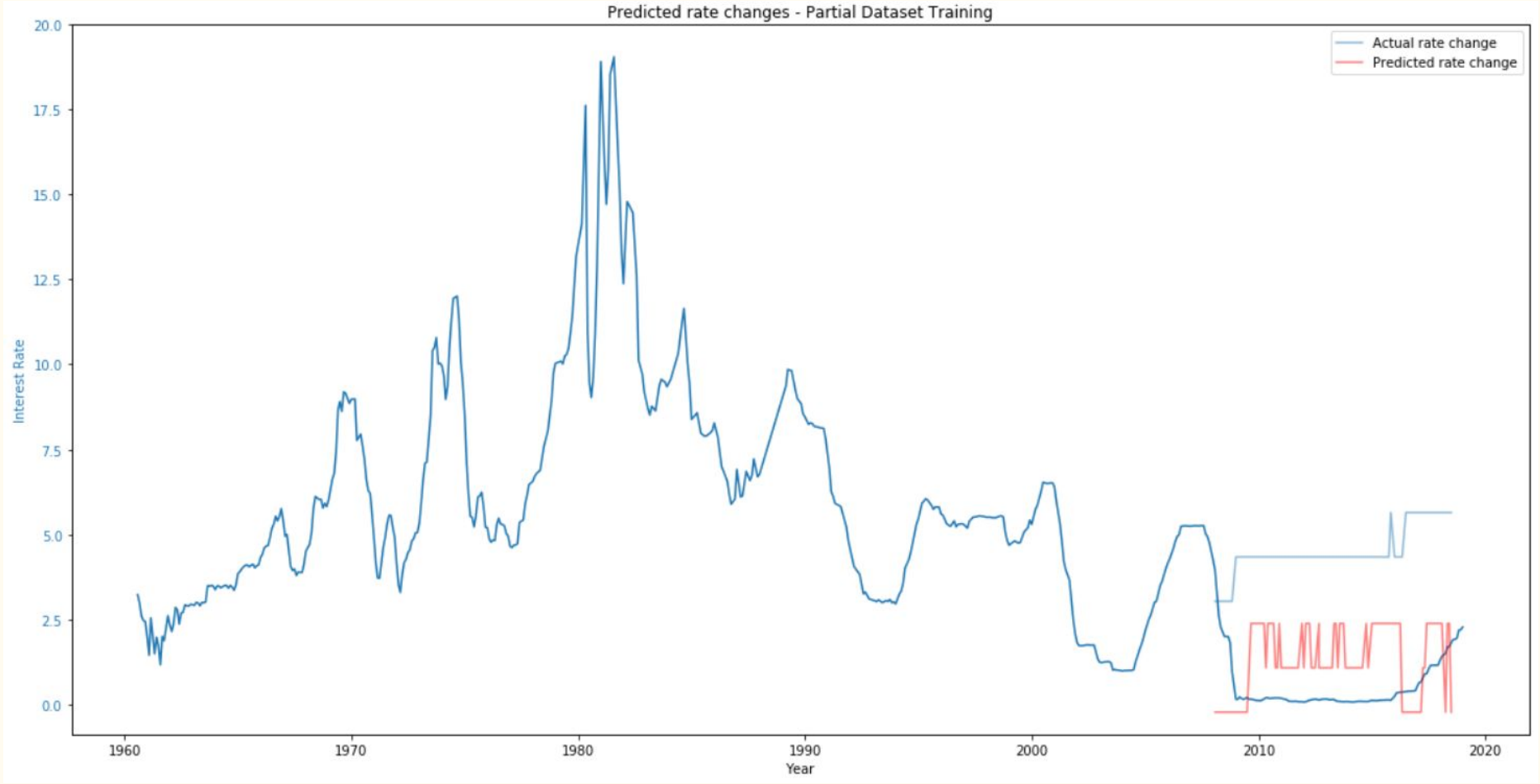
1. Fitting to proper nouns
2. Fitting to sentence fragments
3. Features will vary significantly based on train/test split
4. **Mostly logical and reasonable features**

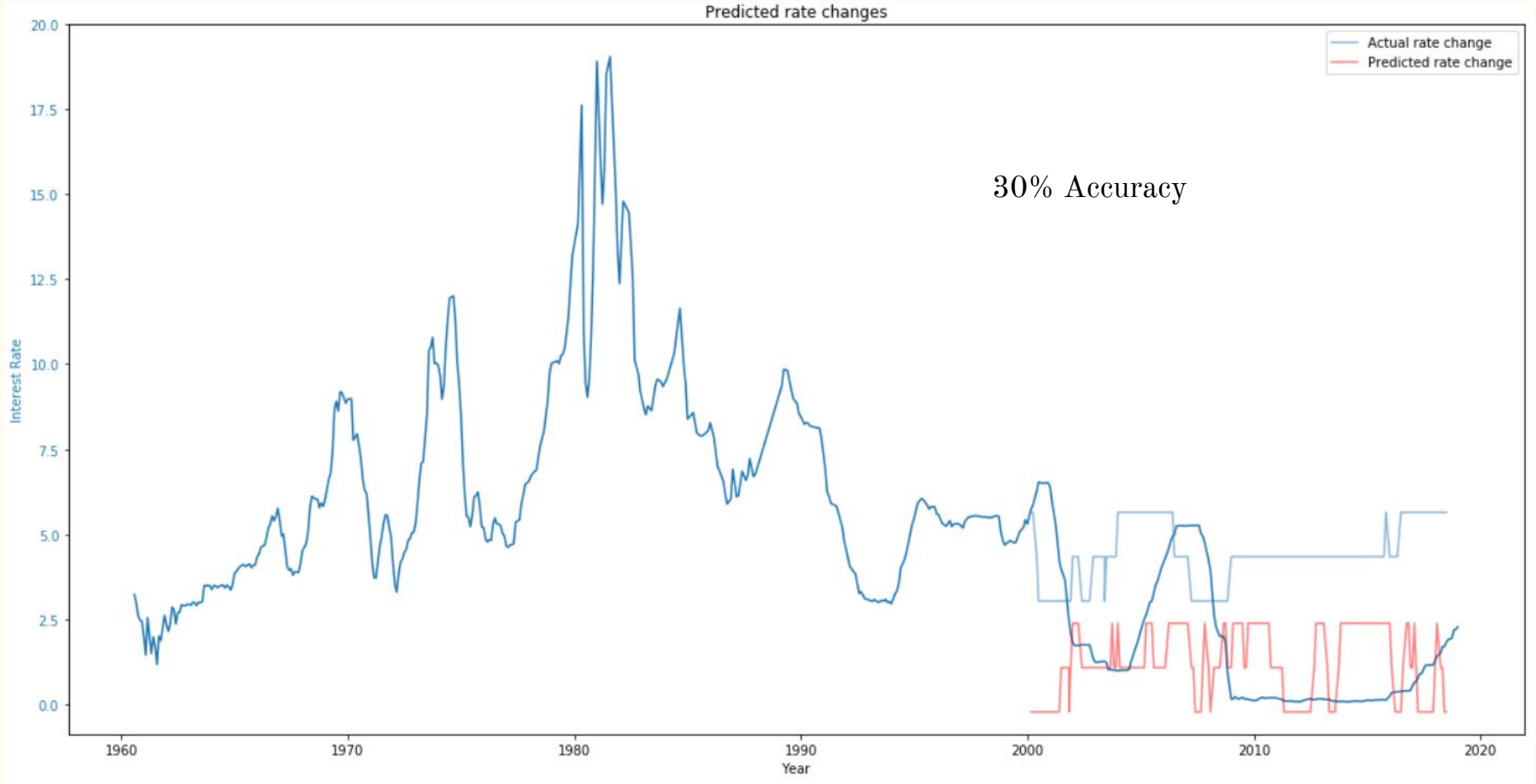



Predicted rate changes
Actual rate change
Predicted rate change
Interest Rate
Year
20.0
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0.0
1960
1970
1980
1990
2000
2010
2020

Model Limitations

1. Overfitting (100% accuracy on train, 65% on test)
2. Fitting to proper nouns
3. Fitting to characteristics of communication documents, not the meaning they're trying to communicate
4. Performs extremely poorly when creating long-term projections, without training on new communications


Predicted rate changes - Partial Dataset Training
Actual rate change
Predicted rate change
Interest Rate
Year
20.0
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0.0
1960
1970
1980
1990
2000
2010
2020


Predicted rate changes
Actual rate change
Predicted rate change
Interest Rate
Year
20.0
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0.0
1960
1970
1980
1990
2000
2010
2020
30% Accuracy

Use Case - Actual Performance

How would the model actually be used?

Latest confirmed data available would be provided, what was the latest change following the communication 6 months ago.

Latest information would always be provided, the model would be trained to the current interest rate environment.

1. Communication type
2. Economic situation
3. Fed recent actions (example current unprecedented low interest environment)

Use Case - Actual Performance

Test sample range: 1980 to 2018

One sample per year (January to June score)

Model provided 75% random sample of observations until the range date

Example: 1985 January prediction for 1985 June's rate change

Data only provided until 1984 June

Score: 65%

Conclusion - A Word of Warning

65% accuracy, on average, is a realistic expectation. Model performance is high enough to be considered a practical tool, but should **only be used in combination with other quantitative and qualitative tools.**

Current market situation:

1. Record length bull market
2. Unprecedented low interest rates - very little powder left at central banks
3. **Record amounts of capital invested in quant-only strategies**
4. Record amounts of capital in ETFs

'Rational' Expectations Science of Monetary Policy Under QE

ALHAMBRA INVESTMENTS