**Sentiment analysis of Federal Open Market Committee (FOMC) minutes: are they leading indicators of a recession similar to traditional Metrics?**

**Patrick Long** - MPS Data Science Program Student Department of Computer Science and Electrical Engineering University of Maryland, Baltimore County, 1000 Hilltop Circle, Baltimore, MD 21250 plong1@umbc.edu

**Abstract**

Monetary policy decision-making of the Federal Reserve Open Market Committee (FOMC) reacts to macroeconomic events occurring in the US economy. Using various methods developed for textual analysis, I assess whether the policymakers’ comments are closely aligned with the future direction of the US economy. Further I assess some related potential uses of textual analysis of FOMC minutes.

**KEYWORDS**

**FOMC, NLP, Text-analysis, Classification.**

**I. Motivation, Background and Problem Definition**

**Background**

By law the FOMC must meet at least four times a year and, since 1981, eight meetings have been held each year at regular intervals. The FOMC meetings are held to discuss economic and financial conditions and vote on potential changes to monetary policy. This project explores whether a sentiment analysis of FOMC meeting minutes are predictive of an impending recession, in the next 18 months, similar to traditional leading indicators of a recession similar to metrics like the University of Michigan Consumer Sentiment Index or the Conference Board Leading Economic Index [1].

**Dataset**

The FOMC minutes are available in .htm files through the Federal Reserve Public Website. I processed the .htm files using python Beautiful soup and pandas libraries and added a few other data points such as the meeting date. I saved these data in a .json file to avoid repeating this step. This single .json file, scraped from the Federal Reserve Public Website, serves as the input dataset for this project.

**Tools**

I utilized the python scikit-learn and vader-analysis modules for Exploratory Data Analysis (EDA). I exposed the raw text to vader-analysis python library, which assigned scores assessing the percentage of a meeting's minutes that were positive, negative and neutral. These scores served as the sentiment scores that are the features used for model development. Model development was exclusively done with the scikit-learn library. I used a jupyter notebook, inside Anaconda navigator, as my python environment and stored all source code in one .pynb file.

**Motivation**

There are three principal reasons to explore developing a model using FOMC minutes and assessing if such a model can serve as a leading indicator of a recession. Firstly, it would bereassuringto know that the monetary policy decision makers are planning accordingly for changes in the business cycle. Second, if a model is developed that shows these data are leading indicators then this will be a useful metric to show downturn and indeed an upturn in the economy. Similar models can be developed using central bank meeting minutes in other nations.

**II. Existing approaches**

There was no study, of which I was aware, using the sentiment of FOMC meeting minutes as a predictive measure of an upcoming recession. I did review two relevant articles that benefitted

The first article used 2017 FOMC minutes and focuses on sentiment analysis of each member of the board’s statements inside these minutes [3]. These data represent a small subset of the minutes I use. The use of such a small subset is likely due to pre-processing complexities of attributing statements to the individual board members. Ultimately each member of the board is ranked in a correlation matrix to the sentiment of the board. That is, which members most closely represent the median view of the board in a ranking of 1 to 10.

The second article uses FOMC statements, which are more timely and abbreviated summaries of the meetings than the FOMC minutes [4]. This study examines the semantic similarity from meeting to meeting over time. The cosine similarity of word persistence does appear to trend with the business cycle of the economy [4]. The authors use a moving average for smoothing out a lot of the noise in the cosine similarity charts which help us identify the trend.

**Pre-Processing of Data**

The minutes are available in .htm files through the Federal Reserve Public Website. Using the Beautiful Soup python library, I collected links to minutes from all federal reserve meetings between 1996 and 2019. I then looped through each one of these URLs collecting the text from each page and storing this in a data frame along with the meeting date, the year and the length of the meeting minutes text. I saved the final data frame as a .json in order to avoid re-running this step. I visually spot checked several rows to verify that the text matched the date it said on the public website.

**Exploratory Data Analysis**

Once the data were processed, the Vader Sentiment analysis python library was used to determine neutral, positive and negative scores of each of the minutes. I created a table of summary statistics based on the sentiment scores of the minutes. I also plotted the scores over time using matplotlib, and also did the same with length of the meeting minutes. And finally I measured the correlation between the scores and employment level, the job openings rate and the hires rate as these metrics tend to trend with the business cycle and in the case of the employment level are used to determine recession dates [4].

One of the three principal reasons to explore developing a model using FOMC minutes and assessing if such a model could serve as a leading indicator of a recession. Moving forward, using moving averages smooth out some of the seasonal volatility in the meeting minutes sentiment [3].

**III. Proposed solution and exploratory data analysis**

**Developing a Target Variable**

The Dataset of interest are sentiment scores that I generated using the python VADER Sentiment Analysis library to assign scores based on the minutes of Federal Open Market Committee (FOMC) meetings. The FOMC Meeting minutes were retrieved from as they are publicly available on the Federal Reserve’s Public Website. Following on from data collection and development of sentiment scores I’d like to assess whether these scores are predictive of a recession, either as a leading indicator or a coincident one. An option would have been to use a binary classification problem of whether there is a recession or not. This would be a fairly straightforward problem to model and would likely not be that illustrative of a recession and recession intensity. I extended this to a multiclass classification problem that provides a sense of what exactly happens in a recession and the points at which it is most intense [1].

The methodology I’ve used to create the target variable is based largely on the factors that the National Bureau of Economic Research (NBER) Business Cycle committee uses to determine whether there is a recession or not. NBER principally consider changes in the Real Gross Domestic Product (GDP), the unemployment rate and the employment level as factors for determining a recession [5]. Essentially I develop a target variable that is reflective of the magnitude of a recession, that is when all these indicators are negative the classification will be at its highest and when they are all positive it will be 0. This is based largely on the concept of recession magnitude, where certain points of a particular recession show coincident indicators used by NBER are more lower than at other points [6]. Further it also serves to not treat all recessions the same.

I developed two target variables using the aforementioned data used by NBER to determine if there is a recession. For each month there is a classification ranging from 0 to 5 based on the factors: GDP, unemployment rate, employment level and whether there is a recession [6]. The first of the target variables will be on the same dates as the data and one will be offset to 18 months prior. The concept of offsetting one of the target variables 18 months prior is used as a proxy for leading indication of a recession. The other target variable will be on the same dates as the data and this will assess whether the sentiment scores of the FOMC minutes are coincident indicators of a recession.

**IV. Details of the implementation, results, and conclusion.**

**Initial Model Evaluation**

As discussed, this is a Multiclass classification problem and I am attempting to develop two separate models on two different target variables.

**1. Leading Indicator** - Firstly I’ll look at the leading target variable. This serves to assess whether I can develop a model using FOMC minutes and assessing as a leading indicator of a recession. I split the test and training data on a four to 1 ratio or alternatively described as an 80/20 split. I start our evaluation using some of the more straightforward classification algorithms such as Naive Bayes, K-nearest neighbors and logistic regression. I then progressed to assess Random Forest Classifiers and Gradient Boosting Classifiers. Despite a great deal of parameter tuning on both Random Forest Classifiers and Gradient Boosting Classifiers, both models exhibited overfitting. The most promising predictive model on both the training and test set for leading indicator target variable was the Gaussian Naive Bayes Classifier which showed an accuracy score on the training set of 0.53 and an accuracy score on the test set of 0.56.

**2. Coincident Indicator** - Secondly I’ll look at the leading target variable. This assesses whether a model developed using FOMC minutes is coincident to a recession. I again split the test and training data on a four to 1 ratio. I used the same classification algorithms as with the leading indicator. So these included Naive Bayes, K-nearest neighbors, logistic regression, Random Forest Classifiers and Gradient Boosting Classifiers. Random Forest Classifiers and Gradient Boosting Classifiers both models exhibited overfitting, as with the leading indicators. Most of the more straightforward classification algorithms did not exhibit excessive overfitting. There were several models that were more promising with coincident indicators than with the lagging indicator. The most predictive of which was the K Nearest Neighbours Classifier which showed an accuracy score on the training set of 0.59 and an accuracy score on the test set of 0.57.

**Possible extensions**

The Data set showed some predictive power. I decided to explore alternate methodological approaches to see if I could uncover a more predictive model. Two of these alternate methodological approaches used different means of assessing sentiment and one redefined the problem as a regression problem with the target column being the unemployment rate rather than a recession. A summary of each of these four extensions along with their results is detailed below.

1. Using a Bagging classifier, which is an ensemble estimator used to fit base classifiers on subsets of the original data and then aggregating their individual predictions to form a prediction classifier [2]. The base estimator I used for this was logistic regression and I assessed it’s performance versus logistic regression’s non-ensemble algorithm as measured by their respective accuracy scores. The performance difference between the two algorithms was not trivial at 0.41 for the bagging classifier versus 0.48 for the non-ensemble logistic regression algorithm. So the using a bagging classifier failed to improve upon our initial model.
2. The first extension that didn’t use vader Sentiment Analysis scores used term frequency–inverse document frequency (Tfidf) [2]. From each of the raw text of a particular FOMC meeting minutes, I used Tfidf to generate a statistic designed to reflect how important a word is to a document. I then compared the performance between. I compared the performance between tfidf and vader sentiment analysis scores using logistic regression.
3. I used scikit learn Countvectorizer() method to convert a collection of text documents (FOMC Minutes) to a matrix of token counts. I then evaluated a model and compared the initial Naive Bayes results.
4. Finally, I explored an alternate potential use of using FOMC minutes to project the Unemployment Rate (a regression problem rather than Classification). The unemployment rate is one of the criteria used to

**Conclusions**

I was unable to develop a useful, predictive model that showed FOMC minutes to be a leading indicator of a recession similar to traditional metrics. The best performing model, with the leading indicator of a recession as the target variable, was the Gaussian Naive Bayes Classifier that showed an accuracy score on the training set of 0.53 and an accuracy score on the test set of 0.56. It is to an extent reassuring that there is a relationship between the sentiment of the FOMC minutes and a recession, but it is not a strong, usable or predictive one. Indeed based on my extension using regression analysis to attempt to project the unemployment, the minutes are likely not predictive of common timely economic indicators either.

**References:**

[1] V. Zarnowitz, “What is a Business Cycle?”. NBER Working Paper No. 3863, Issued in October 1991. <https://www.nber.org/papers/w3863>. [Accessed March 18, 2020].

[2] SciKit Learn, SciKit Learn - Supervised Learning, SciKit Learn Documentation. <https://scikit-learn.org/stable/supervised_learning.html>. [Accessed March 18, 2020].

[3] H. Ramachandran and D. DeRose Jr., “A Text Analysis of Federal Reserve meeting minutes” arXiv preprint arXiv:1805.07851, 2018, Available: https://arxiv.org/ftp/arxiv/papers/1805/1805.07851.pdf. [Accessed March 18, 2020].

[4] M. Acosta and E. Meade, “Hanging on every word: Semantic analysis of the FOMC's post meeting statement” FEDS Notes, September 30, 2015. Available: https://www.federalreserve.gov/econresdata/notes/feds-notes/2015/semantic-analysis-of-the-FOMCs-postmeeting-statement-20150930.html. [Accessed March 18, 2020].

[5] S. Ng and J. Wright, “Facts and Challenges from the Great Recession for Forecasting and Macroeconomic Modeling” Journal of Economic Literature 2013, 51(4), 1120–1154 http://dx.doi.org/10.1257/jel.51.4.1120. [Accessed March 24, 2020].

[6] J. Mazurek, The Evaluation of Recession Magnitudes in EU Countries during the Great Recession 2008–2010, Review of Economic Perspectives, 16(3), 231-244, 2016. <https://doi.org/10.1515/revecp-2016-0014> [Accessed March 24, 2020].

[7] S. Guido and A. Muller, “Introduction to Machine Learning with Python: A Guide for Data Scientists” Boston: O’Reilly, pp 90-94, 2017.