**Investment Advice from the FOMC**

Stephanie Langeland[1]

Columbia University

Quantitative Methods in the Social Sciences: Master's Thesis

Thesis Advisor: Benjamin Goodrich

May 2018

## ABSTRACT

This study explores how artificial intelligence, specifically natural language processing, can be used to examine meeting materials released by the Federal Open Market Committee (FOMC). Latent Dirichlet allocation and Bayesian posterior prediction models are employed to calculate the distribution of words in each FOMC text and use those probabilities to predict changes in the ten-year constant maturity rate, S&P 500 index, CBOE interest rate on the ten-year Treasury Note, and VIX index. Using FOMC post-meeting announcements and meeting minutes from 1995 – 2018 and open/close securities data, the model successfully predicts changes in these securities within one standard deviation of their normal fluctuations on FOMC days. These results suggest that Fed communications influence markets so much that trading can be carried out based on the model results and trades can be executed immediately after FOMC materials are released and before closing that day to preemptively capture the market's future reaction.

[1] E-mail: stephanielangeland@gmail.com
Please refer to the following GitHub link for the code supporting this paper: https://github.com/Slangeland1/Investment_Advice_from_the_FOMC

## Table of Contents

# 1 Introduction

This study uses latent Dirichlet allocation (LDA) on the words contained in Federal Open Market Committee (FOMC) meeting materials and a posterior prediction model of stock market data to predict movements in various securities on days when FOMC post-meeting announcements and meeting minutes are released.

## 1.1 The Role of The Fed

The Federal Reserve ("Fed") is the central bank of the United States. At all times, the Fed must uphold its dual mandate of maximum employment, stable prices, and moderate long-term interest rates. To do so, the Fed has three tools of monetary policy: the discount rate, reserve requirements, and open market operations. "The discount rate is the interest rate charged to commercial banks and other depository institutions on loans they receive from their regional Federal Reserve Bank's[2] lending facility – the discount window" (Board of Governors of the Federal Reserve System). "Reserve requirements are the amount of funds that a depository institution must hold in reserve against specified deposit liabilities" (Board of Governors of the Federal Reserve System). Finally, open market operations are "the purchase and sale of securities in the open market made by a central bank" (Board of Governors of the Federal Reserve System). Through these tools, the Fed influences the supply and demand of balances held by depository institutions at the regional Federal Reserve Banks, thereby altering the federal funds rate. The federal funds rate is the rate at which depository institutions lend to each other overnight. Changes in the federal funds rate affect other short-term interest rates, foreign exchange rates, long-term interest rates, the amount of money and credit, as well as a range of economic variables, including employment, output, and prices of goods and services (Board of

[2] There are twelve regional Federal Reserve Banks: Boston, New York, Chicago, Saint Louis, Kansas City, San Francisco, Minneapolis, Dallas, Philadelphia, Richmond, Cleveland, and Atlanta.

Governors of the Federal Reserve System).

The FOMC consists of twelve members: the seven members of the Board of Governors of the Federal Reserve System, the president of the Federal Reserve Bank of New York, and four of the seven remaining regional Federal Reserve Bank presidents who serve one-year rotating terms.

The FOMC holds eight regularly scheduled meetings each year in order to review economic and financial conditions, determine the appropriate monetary policy stance, and contribute to economic and policy options (Board of Governors of the Federal Reserve System). In addition to its eight regularly scheduled meetings, the FOMC may also add meetings in times of economic stress or crisis. At each meeting, the FOMC faces three mutually exclusive choices: tighten the monetary policy stance, loosen, or keep it unchanged (Jung 15). Tight monetary policy aims to curb spending in the economy through actions such as, but not limited to, restricting the money supply and raising the federal funds rate. Conversely, loose monetary policy seeks to encourage spending through actions such as, but not limited to, increasing the money supply and lowering the federal funds rate. The third option is to leave the monetary policy stance unchanged since the previous meeting.

Since its inception in 1913, the Fed has evolved to foster transparency through open communication of its current monetary policy stance as well as provide forward guidance on the future path of monetary policy. This approach allows the Fed to anchor expectations and establish credibility, thereby affecting investment decisions of global market participants.

Before 1994, the FOMC did not announce the decisions made at each meeting. However, in February 1994, the FOMC released its first statement explaining the outcomes of the meeting. In February 1995, the FOMC decided that all changes to the monetary policy stance would be announced after every meeting. Until May 1999, the FOMC only released a statement when there was a change to monetary policy. Thereafter, the FOMC began releasing a statement (also referred to as a "post-meeting announcement") after every meeting, detailing the current and future stance of the economy and monetary policy, irrespective of whether a change to the current monetary policy stance was made (Federal Reserve Bank of Dallas).

Each scheduled FOMC meeting spans two days. Typically, the FOMC policy statement is released on day two at 2 PM EDT. Until November 2004, the meeting minutes were released two days after the subsequent meeting; which changed in December 2004, when the Fed started releasing the minutes three weeks after the corresponding meeting (Wynne). Additionally, five years after each meeting, complete transcripts of the meetings are published (Federal Reserve Bank of Richmond).

## 2 Literature Review

As the Fed has become more transparent through its communication tools, market participants have placed more emphasis on the outright and subtle messages portrayed in the FOMC meeting materials. Recent studies have used natural language processing in order to discern sentiment, topics, and signals communicated in the text of FOMC post-meeting announcements and minutes to predict market movements.

### 2.1 Natural Language Processing

"Natural Language Processing (NLP) is a branch of artificial intelligence that helps computers understand, interpret and manipulate human language … to fill the gap between human communication and computer understanding" (SAS). NLP helps programmers "organize and structure knowledge to perform tasks such as automatic summarization, translation, named entity recognition, relationship extraction, sentiment analysis, speech recognition, and topic segmentation" (Kiser).

As the FOMC meeting materials have evolved with increased Fed transparency, they have become longer and more complex, which researchers have studied using NLP. In particular, the FOMC post-meeting announcements and minutes have been analyzed to uncover stylistic changes as well as common topics and themes to gauge market expectations.

In an exploratory phase of their research of FOMC minutes, Boukus and Rosenberg noted that simple text analysis revealed language patterns associated with particular periods. For example, the word "price" spiked in mid-1990 during a recession, which may have reflected volatile oil prices at the time, whereas the significance of the word "growth" tended to diminish during recessionary periods and hit a trough at the end of the 2001 recession (Boukus and Rosenberg 18). These preliminary results inspired them to apply a more structured approach to extract conceptual patterns from FOMC minutes.

Boukus and Rosenberg used Latent Semantic Analysis (LSA) to parse out themes of FOMC minutes from 1987 – 2005. First, a term-frequency vector was constructed, which is a list of the words and their relative frequencies in each document. Each term-frequency vector was arranged into a term-document matrix such that each row corresponded to a term and each column represented one document with their relative frequencies. Themes were then extracted using singular value decomposition of the term-document matrix whereby themes were derived from correlations between term frequencies across documents (Boukus and Rosenberg 3). Finally, the themes were arranged in order of descending importance such that the first theme accounted for the most variance, i.e. the most dominant concept underlying the text (Boukus and

Rosenberg 9). This matrix acted as a training data set while new documents in the testing data set or corpus were folded in.

> Folding-in is a procedure that characterizes new texts in terms of the themes derived from an existing latent semantic structure. Thus, folding-in can be used to test the explanatory power of themes extracted from a fixed corpus to new documents not used in the estimation procedure (Boukus and Rosenberg 10).

Therefore, in testing, terms are categorized by the same methodology used in training. However, new terms that have not been identified in training are deleted in testing. Therefore, LSA can miss conceptual patterns from previously unheard-of words that occur in new documents, which would deteriorate predictive accuracy in testing. However, the FOMC typically uses consistent language in meeting materials so the sample of minutes studied by Boukus and Rosenberg did not seem to suffer from this potential disadvantage since they were able to discern various themes associated with financial market and macroeconomic indicators. Furthermore, they were able to demonstrate that these themes were correlated with current and future economic conditions. Their work suggests that market participants discern signals from the minutes, i.e. Treasury yield changes around the time of the minutes release depend on specific themes expressed, monetary policy uncertainty, and economic outlook (Boukus and Rosenberg 1).

Bosagh Zadeh and Zollmann also used NLP on FOMC minutes to predict realized volatility in stock market prices via S&P daily returns (Bosagh Zadeh and Zollmann 2). They used all 398 meeting minutes from 1967 – March 2008. Minutes from 1967 – 2000 were used as training data while minutes from 2001 – 2008 were used as testing data.

To perform the text analysis of the minutes, Bosagh Zadeh and Zollmann, like Boukus and Rosenberg, stemmed all words and removed stop words. Thereafter, they used an advanced bag-of-words approach.

A bag-of-words model, or BoW for short, is a way of extracting features from text for use

> in modeling, such as with machine learning algorithms … A bag-of-words is a representation of text that describes the occurrence of words within a document. It involves two things: (1) A vocabulary of known words. (2) A measure of the presence of known words. It is called a "bag" of words, because any information about the order or structure of words in the document is discarded. The model is only concerned with whether known words occur in the document, not where in the document (Brownlee).

Using the BoW approach, they employed a financial expert to select all words, sorted by decreasing frequency, from the minutes deemed most relevant to changes in volatility. These words were used as "'seeds' in the dependency tree fragment model: all subtrees of the form 'child' → 'parent', 'child' → 'parent' ← 'child', and 'child' → 'parent' → 'grandparent' that had at least one seed word in them were extracted" (Bosagh Zadeh and Zollmann 3).

As a result, the BoW model was used for classification and regression. Classification was used to predict whether volatility would rise or fall whereas regression was used to predict volatility itself (Bosagh Zadeh and Zollmann 3). Notably, the classifiers of the trained models learned that simple trees could predict the direction of expected volatility with sixty-two percent accuracy (Bosagh Zadeh and Zollmann 6).

Other than for prediction purposes, researchers have also used NLP to study the post-meeting announcements to understand their evolution and effect on the market. The FOMC post-meeting announcements have become a major form of Fed communication as they have evolved to support the Fed's commitment to transparency.

> The first statement issued in February 1994 was a mere 99 words in 4 sentences, compared with 564 words in 22 sentences in December 2014 ... [Also,] while the early statements are written at a reading grade level of 9 to 14 years of schooling, the more recent statements are written at a reading grade level of three years beyond a 4-year college degree (Acosta and Meade).

To gauge how the announcements changed over time, Acosta and Meade used a "tracked changes" approach to monitor words that were added or dropped across statements. This approach identified clues about the changes in the FOMC's views on the economy or policy intensions, specifically the correlation between words used in two consecutive post-meeting announcements (Acosta and Meade).

To examine these correlations, they calculated the cosine similarity between the vector-space models of consecutive FOMC statements to describe their consistency over time (Acosta and Meade). To do so, words and their relative frequencies for each document were noted in a document-term matrix. Next, the cosine similarity was calculated based on the similarity level of consecutive statements. Unrelated or orthogonal documents had a cosine similarity of zero because they shared no words while documents that used the same words in nearly the same proportion had a cosine similarity near unity. Therefore, cosine similarity in NLP is very similar to a correlation coefficient. Furthermore, the documents with cosine similarity near unity represented high semantic similarity and persistence or little “tracked changes” (Acosta and Meade).

Similar to Boukus and Rosenberg’s findings of conceptual patterns during certain periods in the FOMC minutes, Acosta and Meade noted persistence trends during certain periods.

> Between the start of the sample in May 1999 and mid-2003, persistence averaged 0.86, before increasing to 0.94 between mid-2003 and mid-2007. Average persistence declined during the financial crisis and then rose to a very high level between 2009 and 2014. In addition, linguistic persistence was quite variable until about mid-2009, but has shown little meeting to meeting variation since then (Acosta and Meade).

These trends were uncovered while analyzing raw documents that had not been pre-processed (e.g., removal of punctuation, white space, stop words, etc.). Therefore, the correlations between FOMC statements may not have accurately represented semantic content over time. The documents were then pre-processed to stem all words down to a root so that the cosine similarity could be recalculated. Thereafter, the document-term matrix, with the stem words and their term frequency-inverse document frequency (TFIDF) scores, was reconstructed. Each TFIDF score assigns a lower weight to terms that occur in many documents and thus contribute little meaning to the overall consistency of the statement. Words such as “inflation” or “committee” had a TFIDF score of zero because they did not distinguish semantic content between documents. Thus, rare words had higher TFIDF weights (Acosta and Meade).

By simply reading the statements, Acosta and Meade found that FOMC post-meeting statements have become consistently more similar in content from meeting to meeting, especially since the financial crisis. However, their analysis beyond the raw text identified that “semantic content of the statements from one FOMC meeting to the next is less similar and more variable than would appear at first glance” (Acosta and Meade). Therefore, market expectations may be altered given the subtle heterogeneity of each post-meeting statement, which was explored by Rohlfs, Chakraborty, and Subramanian.

Rohlfs, Chakraborty, and Subramanian also used NLP to classify text content of FOMC post-meeting announcements to estimate the statements’ causal effects (Rohlfs, Chakraborty and Subramanian 2097). They analyzed the post-meeting announcements rather than the minutes because the announcements have key market-moving content rather than the minutes, which are lengthier and released several weeks after each meeting (Rohlfs, Chakraborty and Subramanian 2097). Using statements from May 1999 – May 2016, they pre-processed the texts and applied a maximum entropy discrimination latent Dirichlet allocation (MedLDA) model. During pre-processing, non-alphanumeric characters were removed and all texts were converted to lower case so that each document could be separated into a BoW, such that one letter, common, and stop words were deleted and the remaining words were stemmed (Rohlfs, Chakraborty and Subramanian 2098).

Thereafter, they applied a MedLDA model to the FOMC statements and set interest rates as the response variables to compute topics that were closely related to variations in interest rates (Rohlfs, Chakraborty and Subramanian 2097).

> LDA (Latent Dirichlet Allocation) (Blei et al., 2003) is an unsupervised model, whereas supervised topic model (sLDA) (Blei and McAuliffe, 2007) introduces a response variable to LDA for each document. Max-Entropy Discrimination LDA (MedLDA) (Zhu et al, 2009) is [a] max-margin variant of the supervised topic models. MedLDA can be built for both regression and classification prediction tasks. In [their] study [Rohlfs, Chakraborty, and Subramanian] employed the model built for [the] classification task. For classification, the response variables *y* are discrete having values {1, 0, -1} denoting the movements of the interest rates. Hence, [Rohlfs, Chakraborty, and Subramanian] consider the multi-class classification version of the MedLDA. It is defined based on a Support Vector Machine (SVM), which integrates the max-margin principle with an underlying LDA model for topics (Rohlfs, Chakraborty and Subramanian 2098).

The main takeaway from their MedLDA classification model is that it used maximum entropy to find which words were associated with movements in short- and medium-term interest rates. The principle of maximum entropy is key because the most information is extracted from rare events with a low probability of occurring, i.e., rare words in FOMC statements signal changes in monetary policy and economic conditions which affects interest rates. This corresponds with the work of Acosta and Meade such that rare words distinguish semantic content between documents. Therefore, the most information comes from rare terms in FOMC post-meeting announcements. Furthermore, MedLDA classifies each document, or BoW, as associated with a response (e.g., an increase, decrease, or no change in interest rates), so that one can predict future interest rate movements based on these FOMC statement classifications.

In summary, NLP of FOMC minutes and post-meeting statements uncovers complex signals to the market about the future path of monetary policy and economic outlook. The next section discusses how NLP of these FOMC materials can be used to predict movements in interest rates as well as the bond and stock markets.

### 2.2 Predicting Market Expectations Based on NLP

The NLP analysis of FOMC minutes and post-meeting announcements by Boukus and Rosenberg, Bosagh Zadeh and Zollmann, Acosta and Meade, as well as Rohlfs, Chakraborty, and Subramanian showed that conceptual patterns in the FOMC materials were correlated and/or had statistically significant relationships with various financial market and macroeconomic indicators, and predictive power for effective federal funds rates.

Boukus and Rosenberg as well as Bosagh Zadeh and Zollmann used LSA and a BoW model, respectively, to predict Treasury yields and daily volatility changes based on FOMC minutes.

Boukus and Rosenberg examined absolute yield changes in five-minute intervals from 1:30 PM – 3 PM on release and non-release days from 1997 – 2005 using tick-by-tick bids, offers, and transaction prices from five major interdealer brokers. They found that three-month, two-year, and ten-year Treasury yields were more volatile on FOMC minute release days than other days.

> On minutes dates, prior to 2 PM, volatility remains near its mean, but then spikes upward within fifteen minutes following the news release. This heightened volatility persists anywhere from 15 minutes for ten-year Treasuries up to an hour after the event for two-year instruments (Boukus and Rosenberg 16).

Specifically, the ten-year yield, credit spread, and S&P 500 index were most significantly correlated with themes identified in their NLP analysis as well as the crude oil price and corporate bond yield (Boukus and Rosenberg 22). In summary, the strongest theme correlations were with short- and long-term interest rates, which was expected given the FOMC's role in setting rates. However, the theme relationships with macroeconomic indicators were not strong as many indicators are quarterly and more important for long-term trends rather than current monetary policy. Thus, themes did not have predictive power for macroeconomic indicators

(Boukus and Rosenberg 23). Therefore, yield changes depended on specific themes expressed in the minutes, the level of monetary policy uncertainty, and economic outlook, which suggests that market participants discern a complex, multifaceted signal from the minutes (Boukus and Rosenberg 30).

Minutes analysis was also conducted by Bosagh Zadeh and Zollmann, which showed that minutes had predictive power on future volatility in a classification setting. Particularly, in a classification setting such that a support-vector-machine model outperformed a random-guess and majority-based guessing baseline strategy to predict short-term S&P volatility as well as short- and long-term volatility of thirteen-week Treasury Notes. The thirteen-week Treasury predictions were not consistent and unachievable for ten-year Treasury Notes. Thus, FOMC meetings influence the bond market more than the stock market most notably on the days following the meeting (Bosagh Zadeh and Zollmann 6). As a recommendation for future research, they suggested using the VIX as a measure of volatility.

Additionally, Rohlfs, Chakraborty, and Subramanian used FOMC post-meeting announcements rather than minutes to predict market movements. Using May 1999 – May 2016 FOMC statements, they successfully predicted target and effective federal funds rates with ninety-three percent and sixty-four percent accuracy, respectively, as well as Median and Average Treasury rates with forty-two percent and thirty-eight percent accuracy, respectively, when no controls were used in a MedLDA model.

Rohlfs, Chakraborty, and Subramanian also used controls to supplement FOMC statement data to account for recent releases of macroeconomic data. Interest rate movements were first regressed against the unemployment, GDP growth, and CPI inflation rates with their corresponding changes, daily time trend, as well as year, month, and day-of-the week dummy variables (Rohlfs, Chakraborty and Subramanian 2098-2099). Using these variables alone and these variables plus a full set of the two-way interactions, regression models were run for all non-FOMC dates from May 1999 – May 2016. These regressions were run to estimate the relationship between interest rate movements and these macroeconomic indicators on non-FOMC dates (when no post-meeting announcement or minutes were released). Using the coefficients from these regressions, they generated residuals of interest rate movements for FOMC dates in order to create indicators that denoted whether the residual was positive or negative for that interest rate movement on that FOMC date (Rohlfs, Chakraborty and Subramanian 2099). However, this approach yielded lower prediction accuracy rates than the MedLDA model without controls.

Thus, Rohlfs, Chakraborty, and Subramanian concluded that the MedLDA model was most effective in associating text contents of FOMC post-meeting statements with the target federal funds rate given its high prediction accuracy rate of ninety-three percent.

In summary, NLP of FOMC minutes can successfully predict changes in the ten-year yield, credit spread, S&P 500 Index, crude oil price, and corporate bond yield based on themes expressed in the minutes. Additionally, a BoW classification model can outperform baseline-guessing strategies in predicting short-term S&P volatility, and more so short- and long-term volatility of thirteen-week Treasury Notes. Furthermore, a MedLDA classification model can successfully predict movements in the target federal funds rate based on FOMC post-meeting statements.

The next section discusses LDA in order to later compare the models previously discussed to the LDA model constructed in this paper.

### 2.3 Latent Dirichlet Allocation

In this paper, LDA was used on a text corpus made up of FOMC minutes and statements from 1995 – 2018 (refer to section 3.1 for details on the data collection). This section discusses the general LDA model and the next section compares the models used by Boukus and Rosenberg, Bosagh Zadeh and Zollmann, Acosta and Meade, as well as Rohlfs, Chakraborty, and Subramanian to LDA.

In order to perform LDA, each document in the corpus is reduced to a vector of real numbers, each of which represents ratios of counts (Blei, Ng and Jordan 993). This is represented in the form of a document-term matrix after the words have been pre-processed (as discussed in section 2.1).

> LDA is a three-level hierarchical Bayesian model, in which each item of a collection is modeled as a finite mixture over an underlying set of topics. Each topic is, in turn, modeled as an infinite mixture over an underlying set of topic probabilities. In the context of text modeling, the topic probabilities provide an explicit representation of a document (Blei, Ng and Jordan 993).

The goal of LDA is to identify short descriptions of each document within the corpus that allow for efficient processing of large collections of text while preserving vital statistical relationships for tasks such as classification (Blei, Ng and Jordan 993).

An important assumption for BoW models and LDA is exchangeability. Exchangeability is important when reducing the dimensionality of texts because it states that the order of words in a document as well as the specific ordering of the documents in a corpus can be neglected (Blei, Ng and Jordan 994). Words and documents are exchangeable when they are "'conditionally independent and identically distributed', where the conditioning is with respect to an underlying latent parameter of a probability distribution" (Blei, Ng and Jordan 995). Based on exchangeability, LDA captures "significant intra-document statistical structure via the mixing distribution" (Blei, Ng and Jordan 995).

LDA is a generative probabilistic model of a corpus such that "documents are represented as random mixtures over latent topics, where each topic is characterized by a distribution over words" (Blei, Ng and Jordan 996). The notation for LDA defines a word as a basic unit of discrete data that comes from a vocabulary (all words in the corpus) by indexing {1, … , V}. A document is a sequence of *N* words denoted by $w = (w_1, w_2, \ldots, w_N)$, where $w_N$ is the $n^{th}$ word in the sequence (Blei, Ng and Jordan 995). A corpus is a collection of *M* documents denoted by $D = \{w_1, w_2, \ldots, w_M\}$ (Blei, Ng and Jordan 995). Furthermore, LDA assumes the following generative process for each document *w* in a corpus *D* (Blei, Ng and Jordan 996):

1. Choose $N \sim$ Poisson(ξ), such that the number of possible words determined by a random variable $N_i$ is distributed Poisson(ξ).
2. Choose θ ~ Dirichlet(α), where θ is the probability of a topic occurring in a document that is distributed Dirichlet(α).
3. For each of the *N* words $w_N$:
    a. Choose a topic $z_N \sim$ Multinomial(θ).
    b. Choose a word $w_N$ from $p(w_N | z_N, \beta)$, a multinomial probability conditioned on the topic $z_N$.

In summary, each document is a mixture of topics such that each document has a set of topics that are assigned to it via LDA. LDA assigns topics to individual words in each document and the topic distribution is assumed to have a sparse Dirichlet prior. The intuition behind the sparse Dirichlet prior is that documents cover only a small set of topics and topics use only a small set of words frequently. Furthermore, a topic is not semantically or epistemologically defined, rather, it is identified by automatic detection of the likelihood of term co-occurrence. Thus, a word may occur in several topics with varying probabilities based on the distribution of words in each document and therefore topic.

The next section compares the models referenced in sections 2.1 and 2.2 to LDA.

## 2.4 NLP Model Comparison

This section compares LDA to the LSA, BoW, tracked changes, and MedLDA models discussed in sections 2.1 and 2.2.

LSA efficiently extracts "a topic representation of the associations between terms from a term-document co-occurrence matrix" (Po and Bergamaschi 7), which makes it difficult for LSA to deal with polysemous terms[3] because it does not identify different senses of a term. Unlike LSA, LDA is a generative probabilistic model that introduces a conditional independence assumption that a document and keyword are independent conditioned on a latent topic variable (Po and Bergamaschi 7). Therefore, "it has been shown that LDA outperforms LSA, in the representation of ambiguous words and in a variety of other linguistic processing and memory tasks" (Po and Bergamaschi 7). Thus, an LDA model was constructed for this paper rather than an LSA model, similar to the one created by Boukus and Rosenberg.

Additionally, the BoW model extracts features from text for modeling use. The BoW approach focuses on the occurrence of words in a document based on the vocabulary of known words and the measure of each word's presence. All information about the structure and order of words is discarded because the exchangeability assumption also applies to the BoW model. The BoW model employed by Bosagh Zadeh and Zollmann classified documents based on word frequencies within each document. The BoW model is the first step of an LDA model in that all

[3] Words that have multiple meanings.

words are pre-processed and put into a document-term matrix to convert the corpus to a BoW structure. At this point, documents are classified under the BoW model whereas LDA goes further to assign topics to individual words in each document and the topic distribution is assumed to have a sparse Dirichlet prior. Thus, LDA was used in this paper because it is more robust than a simple BoW classification model.

Similar to the BoW model, the tracked changes model did not go far enough to understand the distribution of words and topics within each document like LDA. Hence, the LDA model was used rather than the BoW and tracked changes models. The tracked changes model only explored how the FOMC post-meeting statements changed over time. This was not the objective of this paper, but rather to understand how NLP of FOMC materials could predict market movements. The tracked changes model was cited in this paper to note the existence of conceptual patterns in FOMC post-meeting announcements, which established the importance of the work completed by Rohlfs, Chakraborty, and Subramanian as they used a MedLDA model to analyze FOMC post-meeting statements rather than FOMC meeting minutes.

The MedLDA model used by Rohlfs, Chakraborty, and Subramanian is a supervised max margin variant of LDA that set various securities as response variables. LDA uses the most frequent words in each document to calculate the probability of each word occurring in each document given its topic. Whereas MedLDA uses rare terms to calculate the probability distribution of words in each document via maximum entropy. This MedLDA model inspired the LDA model used in this paper. However, rather than using rare terms like the MedLDA model, the LDA model in this paper takes a more traditional approach by using the most frequent words to model the distribution of words in each document.

# 3 Data & Methodology

This section describes the data used and methodology employed to obtain and pre-process FOMC texts, obtain and calculate changes in the response variables, and construct an LDA model on the words of FOMC materials and posterior prediction model that predicts market reactions on FOMC dates.

## 3.1 FOMC Post-Meeting Statements & Meeting Minutes

All 168 post-meeting statements and 185 meeting minutes released from January 1995 – February 2018 were downloaded from the Board of Governors of the Federal Reserve website. To pre-process the texts, the voting member information in the FOMC statements, punctuation, white space, dates, stop words, and irrelevant words (e.g., “Federal”, “Reserve”, Chairman”, etc.) were removed and all words were converted to lowercase and stemmed. Finally, the texts were converted to a BoW structure to understand how the distribution of words in each document correspond to market movements. That is, a document-term matrix was constructed such that each row represented the document release date (all documents released on the same day were concatenated) and each column was a word with word counts for each document. The document-term matrix was then fed into the posterior prediction model to understand how the distribution of words within each document predicts changes in various securities.

## 3.2 Response Variables

Based on the NLP techniques previously discussed, Boukus and Rosenberg, Bosagh Zadeh and Zollmann, as well as Rohlfs, Chakraborty, and Subramanian successfully predicted changes in the target and effective federal funds rates, ten-year Treasury yield, BAA credit spread[4], S&P 500 index, crude oil price, BAA corporate bond yield, and realized volatility in

[4] Based on the difference between the Moody’s BAA corporate bond yield and the ten-year constant maturity Treasury yield.

stock market prices via S&P daily returns. In this study, similar response variables were used: the ten-year constant maturity rate, S&P 500 index, Chicago Board Options Exchange (CBOE) interest rate on the ten-year Treasury Note, and CBOE VIX index[5]. The intension was to also include the three-month U.S. overnight indexed swap (OIS)[6] and crude oil prices as response variables because the effective federal funds rate is directly influenced by the target federal funds rate, which is set during FOMC meetings and noted in the meeting materials; and crude oil was successfully predicted in previous studies. However, many of the OIS and crude oil price data were missing for the sample period. Since the sample period only consists of 352 FOMC dates, it was decided that no observations could be dropped simply to include the OIS and crude oil data.

Additionally, the VIX was included in this study because Bosagh Zadeh and Zollmann noted that the VIX should be a better gauge of market volatility than realized volatility in stock market prices via S&P daily returns because it is the "fear index" (Bosagh Zadeh and Zollmann 7).

> VIX is the ticker symbol for the Chicago Board Options Exchange (CBOE) Volatility Index, which shows the market's expectation of 30-day volatility. It is constructed using the implied volatilities of a wide range of S&P 500 index options. This volatility is meant to be forward looking, is calculated from both calls and puts, and is a widely used measure of market risk, often referred to as the "investor fear gauge" (Investopedia).

The CBOE did not create the VIX as an academic exercise, but rather to profit from volatility (Harwood). The VIX can help inform investment strategies, particularly, using it as a gauge. For example, when the VIX ticks up, volatility is elevated in the market, usually signaling uncertainty, which would trigger investment in safe haven assets according to the flight to

[5] Refer to Figure 1 in the Appendix for data details.
[6] "An overnight indexed swap is a derivative contract based on the total return on a reference rate that is compounded daily over a specific time period. In the US, this reference rate is the effective federal funds rate, i.e. the weighted average of brokered trades between banks for overnight ownership of bank reserves. This rate is calculated and released daily by the Federal Reserve in its H.15 report" (Farid).

quality theory. Alternatively, a decreasing VIX signals subdued volatility, which is typically accompanied by investors reaching for yield. The VIX normally floats between ten and eighty-one (CBOE). Values greater than thirty reflect higher than normal levels of volatility while values below twenty signal subdued market volatility.

Historically, the VIX has displayed sensitivity to the Fed's communication about monetary policy. For example, on May 22, 2013 the Taper Tantrum occurred when "the Federal Reserve announced that it would begin tapering back its roughly $70 billion a month in bond and mortgage backed securities program" (Rapoza). As a result, the VIX jumped (Yahoo Finance) alongside a surge in U.S. Treasury yields (Investopedia). This exemplifies the impact of monetary policy announcements on changes in VIX and Treasury yields. CBOE also noted patterns of upward spikes in the ten-year Treasury Note together with large downswings in futures prices (CBOE), which should be concurrent with VIX levels since this index is made up of puts and calls[7] based on forward-looking expectations. Therefore, the VIX should be a stronger indicator of market volatility than S&P 500 returns.

The goal of the model was to predict the change in these response variables to understand the impact of Fed releases. To do so, the closing value was recorded as a percentage of the opening value for each response variable. Therefore, if the response variable is less than one, that security closed lower than it opened whereas if the response variable is greater than one, that security closed higher than it opened, and if the response variable equals one, there was no change. For all variables, increases represent market rallies or an "improvement" to market conditions. Therefore, the reciprocal of the transformed VIX response variable was recorded so

[7] "A put option is an option contract giving the owner the right, but not the obligation, to sell a specified amount of an underlying security at a specified price within a specified time. This is the opposite of a call option, which gives the holder the right to buy shares" (Investopedia).

that increases in the VIX represent subdued volatility and decreases in the VIX represent elevated volatility.

### 3.3 LDA Model and Posterior Prediction Model

The elements of the LDA model for FOMC materials were defined as objects in R[8] and fed into an LDA model and posterior prediction model coded in Stan. This LDA model calculated the distribution of words for each document and then predicted the distribution of words for new documents, which is the posterior predictive distribution of words in each document i.e. the probability of each word occurring. The main assumption is that the words and documents are exchangeable and the changes in the response variables were caused by the FOMC meeting materials release.

These probabilities informed the predictions of the response variables. The final output yielded predicted changes in the response variables, which was compared against the actual response variables.

## 4 Results

The results of the posterior prediction model yielded predicted changes in the response variables $\tilde{y}$. The $\tilde{y}$ parameter was an array of the number of iterations (1,000) by the number of chains (4) by the number of FOMC days times the number of response variables (352 * 4 = 1,408). Therefore, there was a matrix of posterior draws for each iteration and chain within $\tilde{y}$. These posterior draws were then averaged to obtain a single matrix, which was compared to the

[8] The LDA model inputs were the word stems, document IDs, prior weights of a word in a topic versus document, as well as the number of documents, words in each document, words in the vocabulary, and topics. The posterior prediction model inputs were the hyperparameters, number of response variables, observed data, probability of the daily change in each response variable, standardized influence of each word in the vocabulary, standard deviation in the coefficient across words, and standard deviation of the response variables.

observed data $y$[9]. The next sections discuss the prediction accuracy of the results and posterior predictive checks that were carried out to validate the model.

## 4.1 Prediction Accuracy

Two tests for prediction accuracy were completed to understand how closely the predictions were to the observed data (within a range) and to verify that the model predicted the correct direction of the change in the response variables.

### *4.1.1 Range Prediction Accuracy*

The absolute value of the average posterior draws matrix was subtracted from the absolute value of the observed data to discern how close the predictions $\tilde{y}$ were to the actual response variables $y$. For each variable, if these differences were less than or equal to one standard deviation for that response variable, the prediction was within an acceptable range. If the differences were greater than one standard deviation for that observed response variable value, the prediction was unacceptable since it was out of range. For example, one standard deviation for the observed S&P 500 index response variable $y$ was 0.0125. Since the difference between the average posterior draws and $y$ on July 6, 1995 was 0.004, which is less than the standard deviation for the S&P (0.0125), the prediction was within an acceptable range (refer to Figure 2 in the Appendix).

Using this range prediction accuracy methodology, the results showed that the model predicted within an acceptable range seventy-four percent of the time for the ten-year constant maturity rate, eighty-one percent of the time for the S&P 500 index, seventy-seven percent of the

[9] In order to use this model to trade, one should take each matrix of the posterior predictions and compare them to the observed data to evaluate the distribution of fit. This method captures the uncertainty and is more comparable to the marginal distribution of data on any particular FOMC release date.

time for the CBOE interest rate on the ten-year Treasury Note, and seventy-seven percent of the time for the VIX index (refer to Figure 3 in the Appendix).

#### *4.1.2 Directional Prediction Accuracy*

As previously discussed, the response variables are each security's closing value as a percentage of its opening value. For all variables, values greater than one represent market rallies or an "improvement" to market conditions, values less than one are market downturns, and values equal to one symbolize no change. It is crucial that the model correctly predicted the direction of the response variable {< 1, 1, > 1}. For example, if the observed value on a given day for the S&P 500 index is less than one, this signals that the S&P closed lower than it opened. Hence, the market experienced a downturn after the FOMC release (refer to Figure 4 in the Appendix). For this observation, if the model incorrectly predicted that the S&P closing value would be greater than one, implying that the S&P closed higher than it opened and a market rally occurred, investors using the model to trade would execute incorrect trading strategies (which will be discussed further in section 5.1).

Under this directional prediction accuracy methodology, the overall model with all response variables best predicted instances when the closing values were greater than the opening values (refer to Figure 5 in the Appendix). However, it was never able to predict when no changes to the response variables occurred, and poorly predicted when the closing values were less than the opening values. These trends were largely true for the individual response variables as well.

The model results for the ten-year constant maturity rate were the worst of all of the response variables as the prediction accuracy for all cases was very weak (refer to Figure 6 in the Appendix). Furthermore, the model results for the S&P 500 index best predicted instances when the closing values were greater than the opening values (refer to Figure 7 in the Appendix). Additionally, the model results for the CBOE interest rate on the ten-year Treasury Note and VIX index showed that the model best predicted when the closing values were greater than the opening values, nearly never predicted when there was no change to the closing values since opening, and poorly predicted when the closing values were less than the opening values (refer to Figures 8 – 9 in the Appendix).

Therefore, the model's strength was predicting the range of the response variables rather than the directional change.

## 4.2 Posterior Predictive Checks

This section discusses the posterior predictive checks that were carried out to test for chain convergence and model fit.

### *4.2.1 Chain Convergence*

To test how well the model's chains mixed, the $\hat{R}$ values were examined. Parameters of interest should have $\hat{R}$ values equal to one, which is convergence. Although $\alpha$ was specified to be a positive ordered vector, some chains appeared to be going to different places but implied the same distribution for all parameters that were not $\theta$ and $\alpha$ (even though within each chain, the $\hat{R}$ values for $\theta$ and $\alpha$ were acceptable). Most importantly, the $\hat{R}$ values for the predictions parameter of interest $\tilde{y}$ were equal to one, which supports the model results previously discussed since the chains for $\tilde{y}$ converged.

### *4.2.2 Model Fit*

Density overlay plots were created for each response variable to compare the observed distribution of data $y$ to the simulated distributions $\tilde{y}$ from the posterior predictive distribution.

In the density overlay plots for the ten-year constant maturity rate, S&P 500 index, and CBOE interest rate on the ten-year Treasury Note, it appeared that the observed data were more concentrated than the predictions. Thus, the model produced more extreme predictions than what actually occurred in the data. However, for the VIX index, it appeared that the predictions were more concentrated than the observed data. Consequently, the model produced less extreme predictions than what occurred in the data (refer to Figure 10 in the Appendix).

Additionally, marginal distribution plots supported by numerical fit assessments, that show the percentage of observed data that fit into the fifty percent predictive interval, were prepared to demonstrate whether the model under or over fit the predictions $\tilde{y}$. For the ten-year constant maturity rate, S&P 500 index, and CBOE interest rate on the ten-year Treasury Note, it appeared that the model over fit because over fifty percent of the observed data fit in the fifty percent predictive interval. However, for the VIX index, it appeared that the model under fit because under fifty percent of the observed data fit in the fifty percent predictive interval (refer to Figure 11 in the Appendix).

## 5 Discussion of Findings

In summary, the LDA model produced a posterior predictive distribution of words for each document, which fed the posterior prediction model that predicted changes in the ten-year constant maturity rate, S&P 500 index, CBOE interest rate on the ten-year Treasury Note, and VIX index, based on the distribution of words in the documents, i.e., the probability of each word occurring in each topic on a given FOMC date. Overall, the posterior prediction model successfully predicted changes in these securities within one standard deviation but struggled to predict the direction of the change in the securities.

NLP performed in this study, by Boukus and Rosenberg, Bosagh Zadeh and Zollmann, as well as Rohlfs, Chakraborty, and Subramanian, suggest that market participants act on messages communicated in FOMC meeting materials. Based on these findings, it is evident that a machine learning approach for analyzing Fed communications can highlight or predict consequent market movements. Based on the collective research performed and referenced in this paper, it is possible to construct a hypothetical trading strategy that is based on NLP analysis and predictive models for stock and bond market instruments.

## 5.1 Trading Implications

The results of the posterior prediction model described in this study successfully predicted movements in various securities within one standard deviation of their normal movements on FOMC dates. Under the range prediction accuracy methodology, the results showed that the model predicted within an acceptable range seventy-four percent of the time for the ten-year constant maturity rate, eighty-one percent of the time for the S&P 500 index, and seventy-seven percent of the time for the CBOE interest rate on the ten-year Treasury Note and VIX index. Therefore, upon future FOMC materials releases, this model could be run at that time with the new text to predict movements in these securities for that day. Thus, predicting the change in these securities at closing compared to opening. For example, investors could enter into put and call options, based on the closing "price" change predicted by the model, and whether they want to buy or sell based on their portfolio mix, for these securities within minutes of the FOMC materials release (refer to Figure 12 in the Appendix). Therefore, this model could be treated as a trading model.

## 5.2 Limitations & Future Research

Limitations that faced this study include lack of intraday data analysis, no control variables, and small sample size.

At the onset of this research project, the intension was to also include intraday data to forecast the magnitude of changes in the response variables on non-FOMC versus FOMC days before and after 2 PM to understand the impact of FOMC releases on the market. Exploratory data analysis of intraday data was conducted, but no meaningful differences in the response variables were noted on FOMC versus non-FOMC dates. This may be attributed to the small number of release dates (352 days) relative to the total number of trading days since 1995. In the future, using a bigger sample size of FOMC release dates, this study could be reproduced to include more FOMC release dates with post-meeting announcements and minutes, additional types of Fed press releases (e.g., speeches by Fed Governors), and other types of market moving news (e.g., corporate actions of valuable U.S. firms). This may be a future area of research that can be accomplished using the code, included in the GitHub repository for this paper, as a foundation.

Additionally, since Rohlfs, Chakraborty, and Subramanian had worse prediction results using control variables in their MedLDA model and better results in the model without controls, this study did not include control variables. However, the lack of controls may have caused weaker prediction results due to noise in the data.

## 6 Conclusion

Since its inception in 1913, the Fed has played a major and everchanging role in the U.S. economy and global financial markets. For most of the Fed's history, it was viewed as a black box, which changed in the 1990s when the FOMC began communicating its current and future monetary policy stance to the market. Since then, the role of "Fed Watchers" has become less focused on discerning exactly *what* the Fed is doing and more focused on the subtle and outright messages the Fed openly communicates to the market. Using artificial intelligence techniques, such as NLP, researchers have been able to dissect Fed communication materials in order to track the evolution of Fed announcements and predict how stock and bond markets will react to Fed announcements. By harnessing the power of machine learning, it is possible to teach algorithms to read and discern the meaning behind Fed communications and therefore predict market reactions.

## 7 Works Cited

Acosta, Miguel and Ellen Meade. *Hanging on every word: Semantic analysis of the FOMC's postmeeting statement*. 30 September 2015. 31 January 2018. <https://www.federalreserve.gov/econresdata/notes/feds-notes/2015/semantic-analysis-of-the-FOMCs-postmeeting-statement-20150930.html>.

Blei, David M, Andrew Y Ng and Michael I Jordan. "Latent Dirichlet Allocation." *Journal of Machine Learning Research* (2003): 993-1022.

Board of Governors of the Federal Reserve System. *Federal Open Market Committee*. 26 January 2016. 20 November 2016. <https://www.federalreserve.gov/monetarypolicy/fomc.htm>.

—. *Historical Materials by Year*. 5 January 2018. 20 January 2018. <https://www.federalreserve.gov/monetarypolicy/fomc_historical_year.htm>.

—. *Meeting calendars, statements, and minutes (2013-2018)* . 5 January 2018. 20 January 2018. <https://www.federalreserve.gov/monetarypolicy/fomccalendars.htm>.

—. *Open market operations*. 9 November 2016. 25 November 2016. <https://www.federalreserve.gov/monetarypolicy/bst_openmarketops.htm>.

—. *Reserve Requirements*. 27 October 2016. 25 November 2016. <https://www.federalreserve.gov/monetarypolicy/reservereq.htm>.

—. *The Discount Rate*. 18 October 2016. 25 November 2016. <https://www.federalreserve.gov/monetarypolicy/discountrate.htm>.

—. *Transcripts and other historical materials*. 29 January 2016. 23 November 2016. <https://www.federalreserve.gov/monetarypolicy/fomc_historical.htm#policy>.

Bosagh Zadeh, Reza and Andreas Zollmann. "Predicting Market-Volatility from Federal Reserve Board Meeting Minutes." 2009.

Boukus, Ellyn and Joshua V. Rosenberg. "The information content of FOMC minutes." 2006.

Brownlee, Jason. *A Gentle Introduction to the Bag-of-Words Model*. 9 October 2017. 4 February 2018. <https://machinelearningmastery.com/gentle-introduction-bag-words-model/>.

CBOE. *CBOE Volatility Index (VIX Index) FAQs*. 29 June 2016. 20 November 2016. <http://www.cboe.com/micro/vix/faq.aspx#1>.

—. *CBOE/CBOT 10-year U.S. Treasury Note Volatility IndexSM (TYVIXSM)* . n.d. 23 November 2016. <http://www.cboe.com/micro/volatility/tyvix/default.aspx>.

—. *VIX of VIX (VVIXSM) Whitepaper*. n.d. 23 November 2016. <http://www.cboe.com/micro/vvix/vvixwhitepaper.aspx>.

Farid, Jawwad. *Using OIS Swap valuation – Overnight Indexed Swap rates versus LIBOR*. 3 October 2012. 14 February 2018. <https://financetrainingcourse.com/education/2012/10/using-ois-overnight-indexed-swap-rates-versus-libor-for-irs-pricing/>.

Federal Reserve Bank of Dallas. *A Short History of FOMC Communication Vol. 8, No. 8, September 2013* . n.d. 19 November 2017. <https://www.dallasfed.org/research/eclett/2013/el1308.cfm>.

Federal Reserve Bank of Richmond. *FOMC*. n.d. 20 November 2016. <https://www.richmondfed.org/faqs/fomc>.

Federal Reserve Bank of Saint Louis. *Effective Federal Funds Rate*. 2 February 2018. 14 February 2018. <https://fred.stlouisfed.org/series/FEDFUNDS>.

—. *How Monetary Policy Works*. n.d. 19 November 2017. <https://www.stlouisfed.org/in-plain-english/how-monetary-policy-works>.

Grun, Bettina and Kurt Hornik. "topicmodels: An R Package for Fitting Topic Models." n.d.

Harwood, Vance. *How Does the CBOE's VIX Index Work?* 13 October 2016. 23 November 2016. <https://sixfigureinvesting.com/2014/07/how-does-the-vix-index-work/>.

Investopedia. *Flight to Quality*. n.d. 13 December 2016. <http://www.investopedia.com/terms/f/flighttoquality.asp>.

—. *Put Option*. n.d. 13 December 2016. <http://www.investopedia.com/terms/p/putoption.asp>.

—. *VIX - CBOE Volatility Index*. n.d. 23 November 2016. <http://www.investopedia.com/terms/v/vix.asp?lgl=no-infinite>.

Jung, Alexander. "Have FOMC minutes helped markets to predict FED funds rate changes?" *Working Paper Series* (2016).

Kiser, Matt. *Introduction to Natural Language Processing (NLP) 2016*. 11 August 2016. 31 January 2018. <http://blog.algorithmia.com/introduction-natural-language-processing-nlp/>.

Li, Cheng and Bingyu Wang. "Fisher Linear Discriminant Analysis." 2014.

Minitab 17. *What is quadratic discriminant analysis?* n.d. 23 November 2016. <http://support.minitab.com/en-us/minitab/17/topic-library/modeling-statistics/multivariate/discriminant-analysis/what-is-quadratic-discriminant-analysis/>.

Po, Laura and Sonia Bergamaschi. "Comparing LDA and LSA Topic Models for Content-Based Movie Recommendation Systems." n.d.

Rapoza, Kenneth. *Happy Anniversary To Market's Taper Tantrum*. 23 May 2014. 23 November 2016. <http://www.forbes.com/sites/kenrapoza/2014/05/23/happy-anniversary-to-markets-taper-tantrum/#7dac68f83aea>.

Rohlfs, Christopher, Sunandan Chakraborty and Lakshminarayanan Subramanian. "The Effects of the Content of FOMC Communications on US Treasury Rates." *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing* (2016): 2096–2102.

SAS. *Natural Language Processing: What it is and why it matters*. n.d. 31 January 2018. <https://www.sas.com/en_us/insights/analytics/what-is-natural-language-processing-nlp.html#>.

Wynne, Mark A. *Tools of Communication*. September 2013. March 2018. <https://www.dallasfed.org/research/eclett/2013/el1308.cfm>.

Yahoo Finance. *VOLATILITY S&P 500 (^VIX)*. 23 November 2016. 23 November 2016. <https://finance.yahoo.com/quote/%5EVIX/history?period1=1369281600&period2=1374552000&interval=1d&filter=history&frequency=1d>.

# 8 Appendix

### Figure 1: Data Details

<table>
<tr><th>Ticker</th><th>Description</th><th>Unit</th><th>Start</th><th>End</th><th>Source</th></tr>
<tr><td>^TNX</td><td>CBOE Interest Rate Ten-year Treasury Note</td><td rowspan="4">Open and Close</td><td>01/03/1995</td><td rowspan="4">03/01/2018</td><td>Yahoo Finance</td></tr>
<tr><td>H15T10Y Index</td><td>Ten-year Constant Maturity Rate</td><td rowspan="3">01/02/1995</td><td rowspan="3">Bloomberg Terminal</td></tr>
<tr><td>SPX Index</td><td>S&P 500 Index</td></tr>
<tr><td>VIX Index</td><td>CBOE VIX Index</td></tr>
</table>

### Figure 2: Range Prediction Accuracy

<table>
<tr><th>FOMC Date</th><th>Standard Deviation of the Observed S&P 500 Index Changes</th><th>Difference Between the Average Posterior Draws & Observed S&P 500 Index</th><th>Prediction Accuracy</th></tr>
<tr><td>1995-07-06</td><td rowspan="2">0.01246075</td><td>0.003953571</td><td>Within Range (Correct)</td></tr>
<tr><td>1996-07-05</td><td>0.02484074</td><td>Out of Range (Incorrect)</td></tr>
</table>

### Figure 3: Range Prediction Accuracy Results

| Security | Range Prediction Accuracy |
|---|---|
| Ten-year Constant Maturity Rate | 74% |
| S&P 500 Index | 81% |
| CBOE Interest Rate on the Ten-year Treasury Note | 77% |
| VIX Index | 77% |

### Figure 4: Directional Prediction Accuracy

| FOMC Date | Observed S&P 500 Index Change | Predicted S&P 500 Index (Average of Posterior Predictive Draws) | Prediction Accuracy |
|---|---|---|---|
| 1995-07-06 | 1.0122976 | 1.008344 | Correct |
| 2018-02-21 | 0.9929426 | 1.003418 | Incorrect |

### Figure 5: Overall Model Results (Directional Prediction Accuracy)

| | | Predicted | |
|---|---|---|---|
| | | closing < opening | closing > opening |
| Observed | closing < opening | 10 | 563 |
| | closing = opening | 0 | 138 |
| | closing > opening | 6 | 691 |

### Figure 6: Ten-year Constant Maturity Rate Model Results (Directional Prediction Accuracy)

| | | Predicted | |
|---|---|---|---|
| | | closing < opening | closing > opening |
| Observed | closing < opening | 3 | 127 |
| | closing = opening | 0 | 124 |
| | closing > opening | 1 | 97 |

### Figure 7: S&P 500 Index Model Results (Directional Prediction Accuracy)

| | | Predicted | |
|---|---|---|---|
| | | closing < opening | closing > opening |
| Observed | closing < opening | 2 | 157 |
| | closing > opening | 2 | 191 |

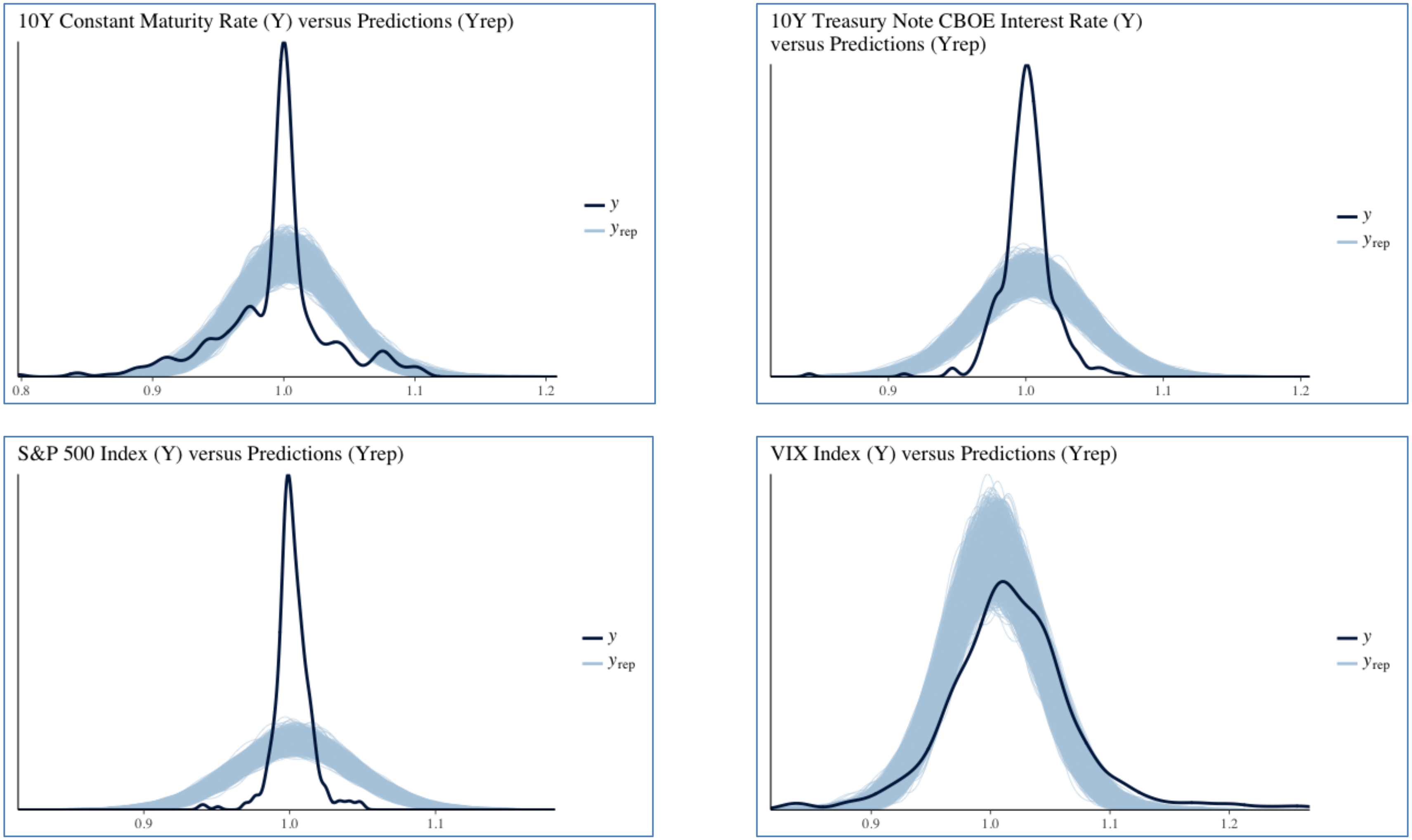
### Figure 8: CBOE Interest Rate on the Ten-year Treasury Note Model Results (Directional Prediction Accuracy)

| | | Predicted | |
|---|---|---|---|
| | | closing < opening | closing > opening |
| Observed | closing < opening | 1 | 162 |
| | closing = opening | 0 | 12 |
| | closing > opening | 3 | 174 |

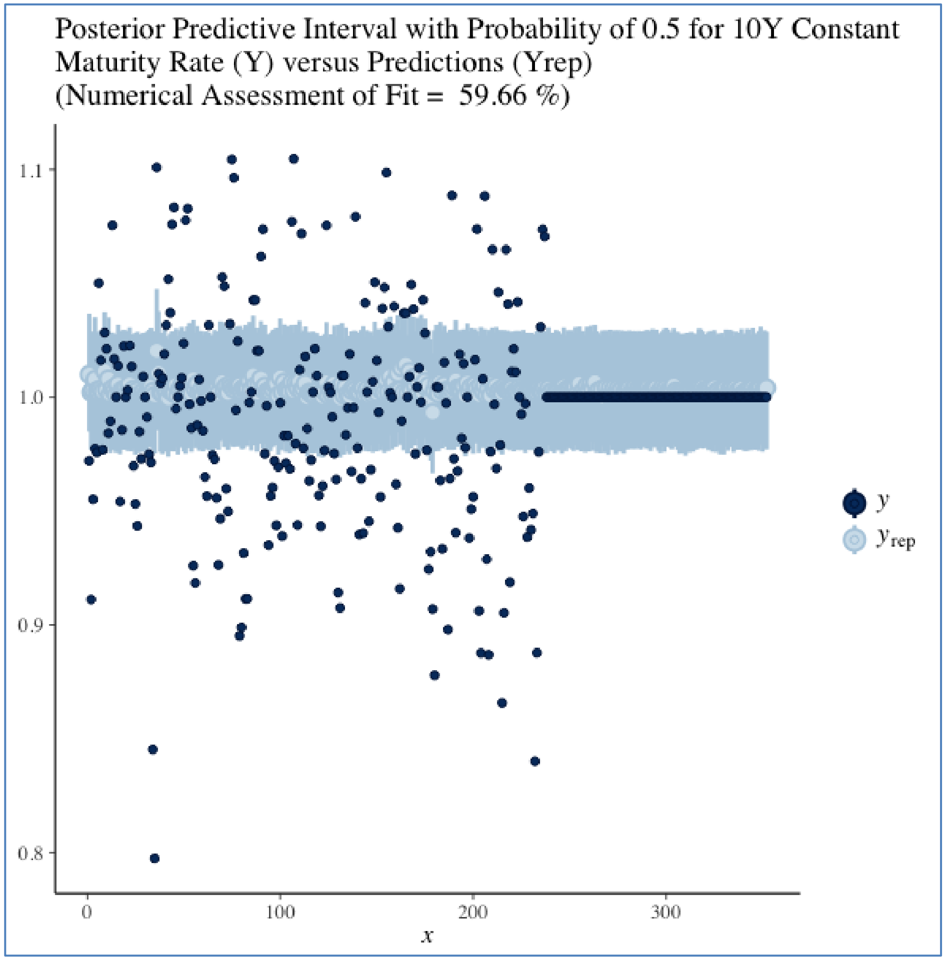
**Figure 9: VIX Index Model Results (Directional Prediction Accuracy)**

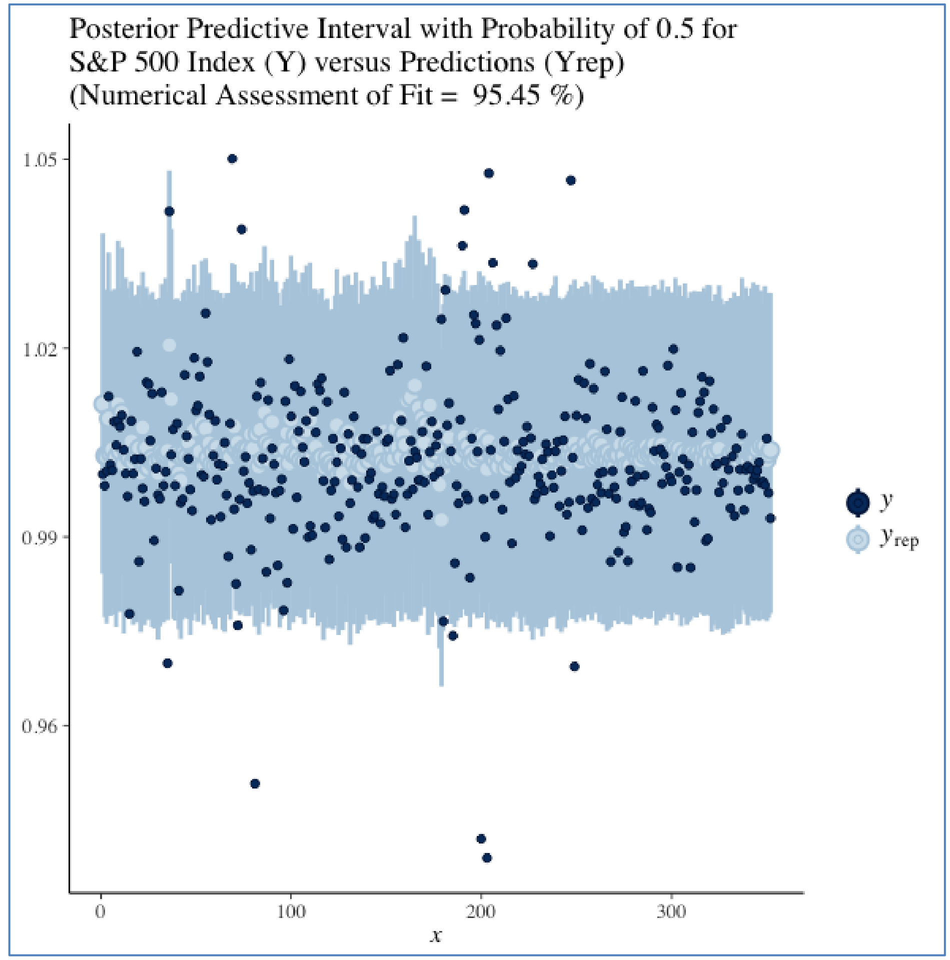
| | | Predicted | |
|---|---|---|---|
| | | closing < opening | closing > opening |
| **Observed** | closing < opening | 4 | 117 |
| | closing = opening | 0 | 2 |
| | closing > opening | 0 | 229 |

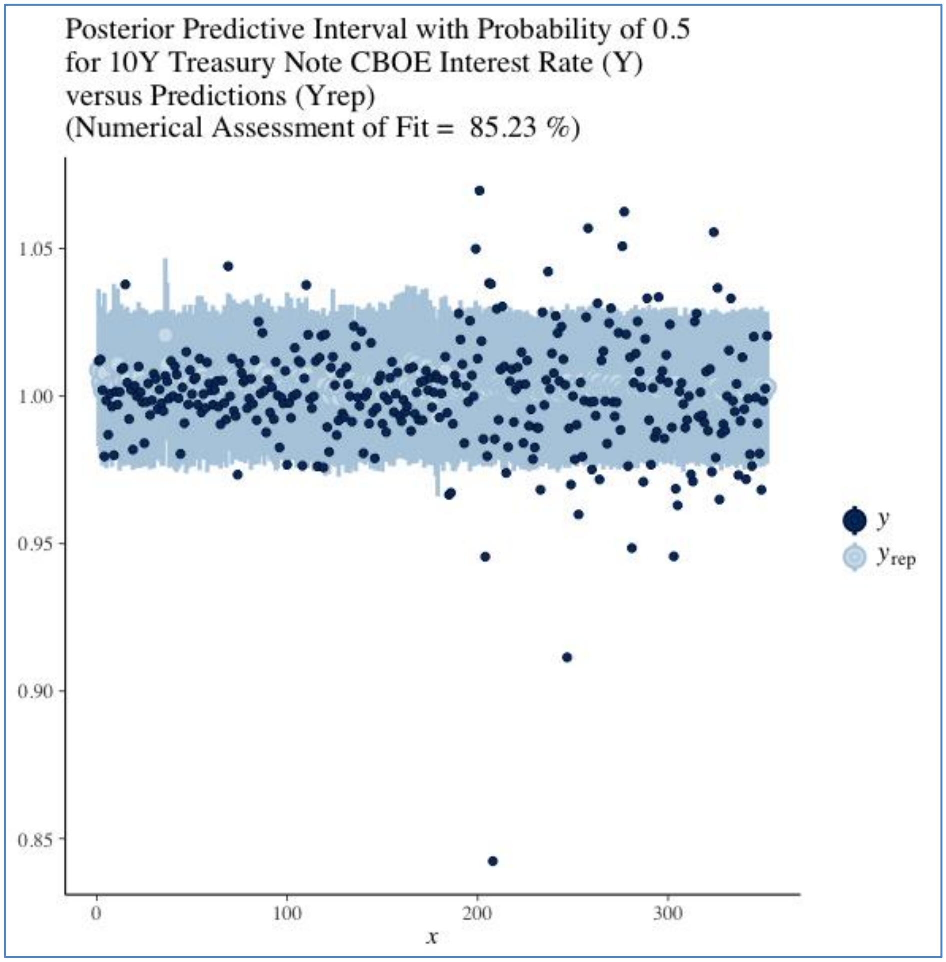
### Figure 10: Model Fit – Density Overlay Plots



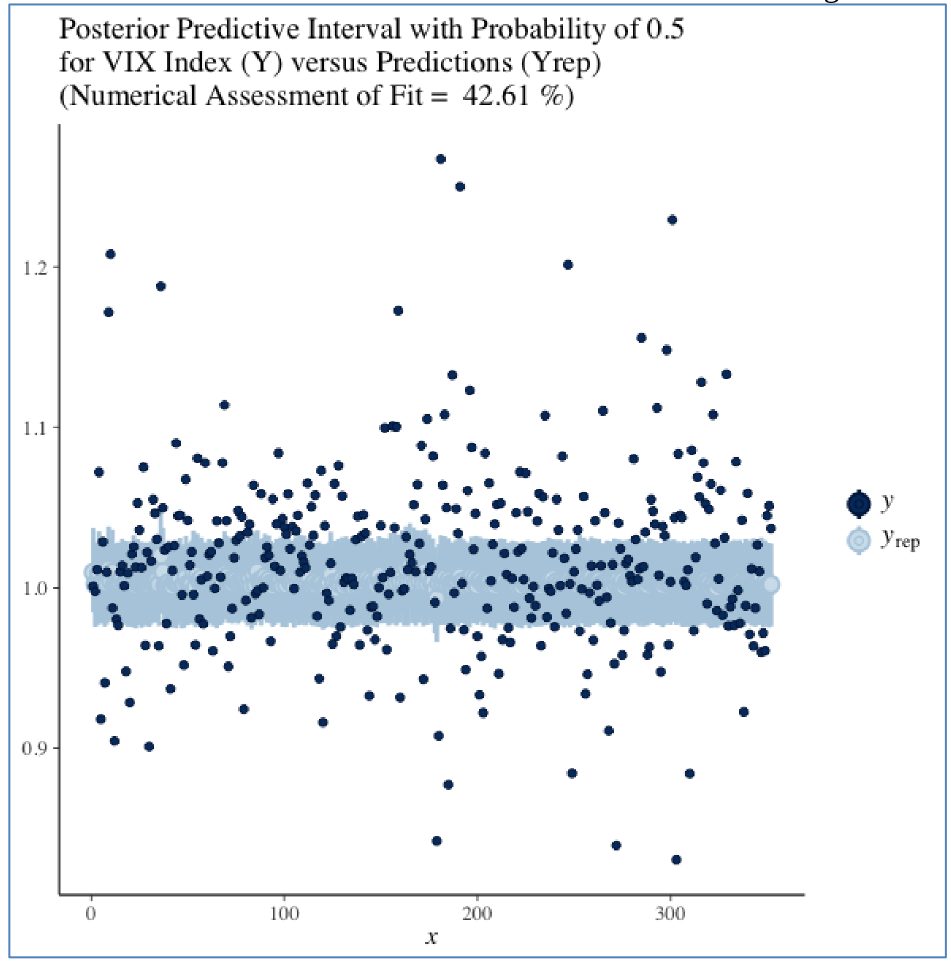
## Figure 11: Model Fit – Marginal Distribution Plots





Posterior Predictive Interval with Probability of 0.5
for 10Y Treasury Note CBOE Interest Rate (Y)
versus Predictions (Yrep)
(Numerical Assessment of Fit = 85.23 %)
1.05
1.00
0.95
0.90
0.85
0
100
200
300
x
y
$y_{rep}$

Posterior Predictive Interval with Probability of 0.5
for VIX Index (Y) versus Predictions (Yrep)
(Numerical Assessment of Fit = 42.61 %)
1.2
1.1
1.0
0.9
0
100
200
300
x
y
$y_{rep}$

**Figure 12: Trading Example**