UKRAINIAN CATHOLIC UNIVERSITY

MASTER THESIS

Stock market prediction utilizing central bank's policy statements

Author:
Roman MOISEIEV

Supervisor:
Andriy ZHOVTANETSKYY

A thesis submitted in fulfillment of the requirements for the degree of Master of Science

in the

Department of Computer Sciences
Faculty of Applied Sciences

Lviv 2020

Declaration of Authorship

I, Roman MOISEIEV, declare that this thesis titled, “Stock market prediction utilizing central bank’s policy statements” and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed:

Date:

"Nobody can predict interest rates, the future direction of the economy, or the stock market. Dismiss all such forecasts and concentrate on what's actually happening to the companies in which you've invested."

Peter Lynch

UKRAINIAN CATHOLIC UNIVERSITY

Faculty of Applied Sciences

Master of Science

Stock market prediction utilizing central bank's policy statements

by Roman MOISEIEV

Abstract

The stock market is quite unpredictable and affected by a vast number of factors. Moreover, many central banks, banks, hedge funds, and other financial institutions target their R&D departments to try to predict probabilities of market movements, possible black swans, and other risks. In this work, I target inefficiencies in the prediction of the market reaction on central bank policy statements. Such statements have two parts: action and information. Therefore in complicated cases, automatic trading systems react to actions and may not recognize vital insights from the informational component. To improve this, I collected historical data for monetary actions and press releases by Federal Reserve, stock price data, Fed Fund futures contract prices. Based on that, I build several classification models to predict the class of policy statements. Afterward, prepared pipeline and the econometric model that can incorporate a class of a policy statement for stock market reaction evaluation.

Acknowledgements

I want to thank Andriy Zhovtanetskyy for supervising my work. His advice in the understanding of the domain and context of the problem was invaluable. I am also grateful to Carlo Rosa for providing supplementary data used in his research. Besides that, I want to thank Ukrainian Catholic University, Faculty of Applied Sciences, and Oleksii Molchanovskyi personally for the excellent Master Program, which was an exceptional experience. Last but not least, I am immensely thankful to my friends and schoolmates that helped me through this endeavor and provided constant support.

Contents

List of Figures

List of Tables

List of Abbreviations

BERT	**B**idirectional **E**ncoder **R**epresentations from **T**ransformers
CME	**C**hicago **M**ercantile **E**xchange
ECB	**E**uropean **C**entral **B**ank
EFFR	**E**ffective **F**ederal **F**unds **R**ate
ETF	**E**xchange-**T**raded **F**und
FOMC	**F**ederal **O**pen **M**arket **C**ommittee
FRED	**F**ederal **R**eserve **E**conomic **D**ata
JSON	**J**avaScript **O**bject **N**otation
LIBOR	**L**ondon **I**nter**b**ank **O**ffered **R**ate
NBU	**N**ational **B**ank of **U**kraine
OHLC	**O**pen **H**igh **L**ow **C**lose
SVD	**S**ingular **V**alue **D**ecomposition
TF-IDF	**T**erm **F**requency–**I**nverse **D**ocument **F**requency

1 Introduction

1.1 Context

Central banks play a vital role in the development of the world economy. After the 2008 year financial crisis, they got even more influence. This influence is used to maintain unemployment low, prices stable, and to support the growth of an economy. As a result, central banks use different policy instruments which affect the whole economy and shift key indicators towards the target level. For example, The Federal Reserve System (central bank of the United States of America), among other things, determines the target range for the federal funds rate. This rate (see Figure 1.1) affects the money supply: for example, when the rate is high - all credit rates in banks are high, and business has expensive loans. Alternatively, when the rate is low - the money are cheap, governmental bonds yields are less profitable, companies make a capital investment in the property plant and equipment, and perform shares buyback. All mentioned, subsequently, creates new workplaces and keeps inflation stable.

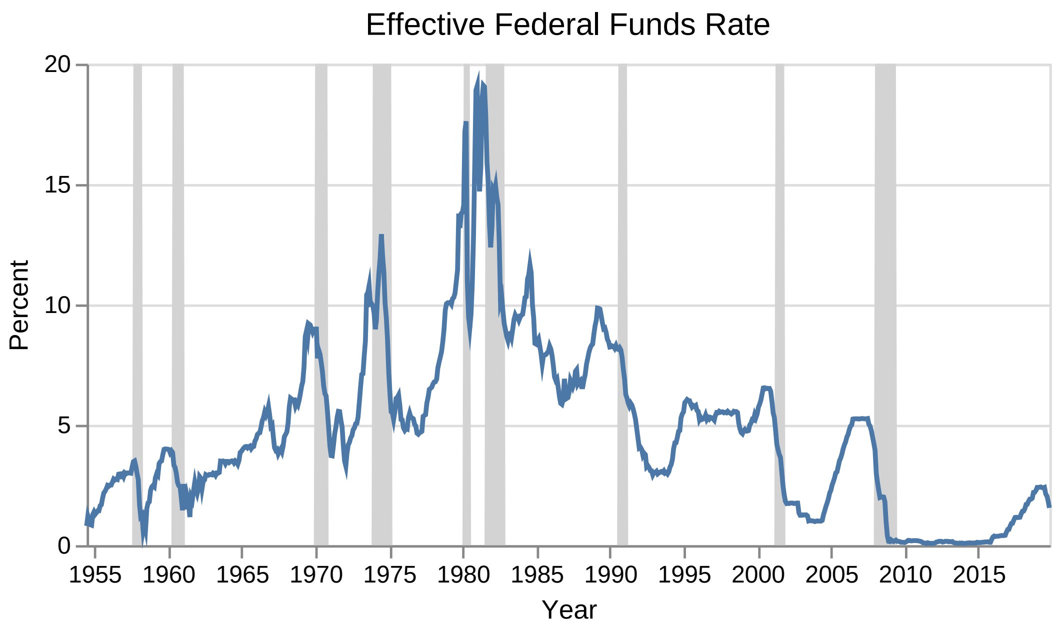

FIGURE 1.1: Effective Federal Funds Rate from 1955 to 2019. Shaded areas indicate U.S. recessions. Data source: Board of Governors of the Federal Reserve System (US), Effective Federal Funds Rate [FEDFUNDS][1].

Decisions about Monetary policy change consist of policy action and communication that accompanies one in the form of the Federal Open Market Committee

[1]retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/FEDFUNDS, December 29, 2019

Statement. The statement provides the motivation behind the decision and view of FOMC on future risks for the economy. Such press release may be classified as:

- "**hawkish**" - if statement implies tighter monetary policy stance to keep inflation low as priority
- **neutral**
- "**dovish**" - if statement implies weaker monetary policy position and prefer low unemployment.

1.2 Motivation

Although economists and market stakeholders know the importance of the communication component of the policymaking of a central bank, it may be underestimated in the algorithmic trading systems.

For example, the National Bank of Ukraine, in its April 2019 Inflation report, shows a case where changes to loosen monetary policy from European Central Bank in March 2019 lead to high market fluctuations (National Bank of Ukraine, 2019). When, firstly, algorithmic traders positively took the news about the policy change, and the price increased. However, on the press conference of the President of the ECB, underlying considerations with worse macroeconomic prognosis were provided. As a result, and the price of the Euro Stoxx 50 Index decreased below the pre-shock state (see Fig. 1.2).

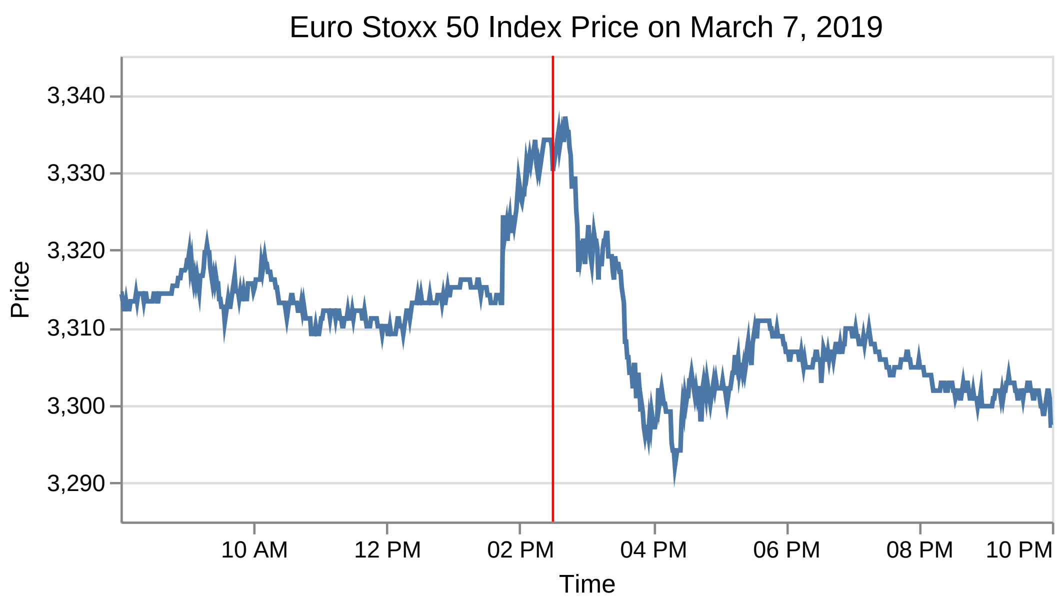

FIGURE 1.2: Price of the Euro Stoxx 50 index on March 7, 2019. Vertical line marks the time of President of the ECB press conference on the considerations underlying monetary policy decisions at a 2:30 PM CET. Data source: Dukascopy Bank SA[2].

Based on that, I believe that the ability to provide real-time evaluation of the informational component of policy decisions will help explain market behavior and

[2]retrieved from Dukascopy; https://www.dukascopy.com/swiss/english/marketwatch/historical/, January 3, 2020

improve existing automatical trading systems. Although achieving a human level of understanding is unlikely.

1.3 Goal and Research Question

The goals of this work are:

1. Collect a dataset with historical data for monetary actions, and press releases of FOMC, stock price data.

2. Construct meaningful features from the FOMC statement.

3. Conduct event-study for FOMC policy statements.

Research Question: Can the FOMC statements that accompany monetary policy decisions be utilized using natural language processing techniques to create statistically significant features that explain the stock market reaction?

1.4 Structure of the thesis

The remainder of this thesis is structured as follows. Section 2 gives an overview of related work and background. The description of collected datasets is given in Section 3. Section 4 presents the solution pipeline, followed by the evaluation of conducted experiments in Section 5. A conclusive discussion of obtained results is given in in Section 6.

2 Background and Related Work

2.1 Related Work

Related work is tied to economic event studies in the field of stock market reaction to macroeconomic news released by statistic agencies and central banks (Kearns and Manners, 2006; Andersson, 2007; Faust et al., 2007). In more narrow context we investigate the particular news - monetary policy action or statement of FOMC (Rigobon and Sack, 2004; Bernanke and Kuttner, 2005). Some studies investigate the impact of U.S. based news on global indexes (Andersen et al., 2007; Wongswan, 2009).

However, with the rise of importance of communication component - studies of Central Bank communications emerge. For example, (Rosa, 2011) use the classification of FOMC statements and show that 90% of explainable variability in stock after event attributed to news shock. (Hernández-Murillo and Shell, 2014) explore that complexity and length of FOMC statements rise over time (see Figure 2.1).

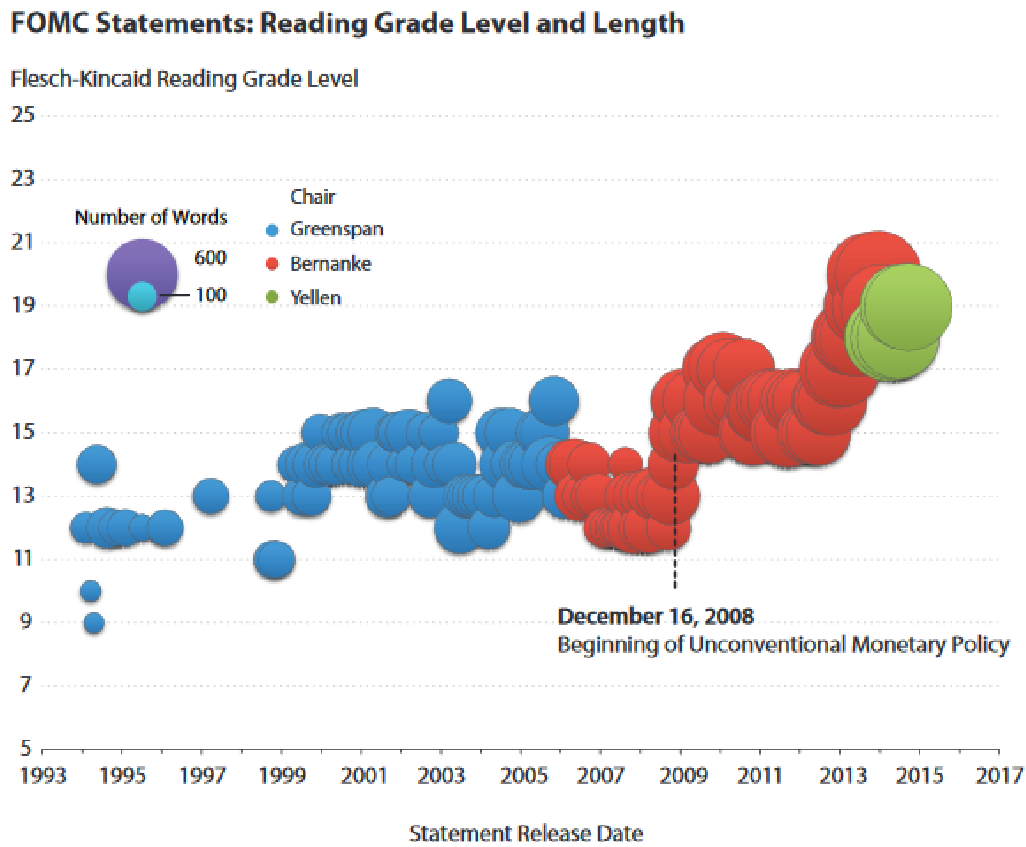

FIGURE 2.1: FOMC Statements: Reading Grade Level and Length. Source: (Hernández-Murillo and Shell, 2014)

Semantic analysis of FOMC statements shows the growth of semantic similarity and strong evidence that since the last financial crisis, FOMC statements have more similar content from meeting to meeting (Meade and Acosta, 2015). More recent research describes the fact that many central banks usually use previous press-releases as a baseline and add only marginal changes (Ehrmann and Talmi, 2019).

Other studies try to leverage advancements in deep learning to make stock prediction (Ding et al., 2015). However, such studies work for a wide range of possible news and learn only from news titles and not designed for usage in a more narrow context. Other applied for particular stock (Hiew et al., 2019) and not for macroeconomic news and their effect on indexes. (Araci, 2019) applies BERT for sentence classification in the financial domain.

My study places itself to be consistent with the spirit of econometric research. Furthermore, perform an attempt to benefit from recent advancements in NLP to automatize feature extraction for a particular event - FOMC statement.

2.2 Central banks. Central bank mandate. Monetary policy

A central bank is an institution that oversees the banking system of a country or union. Over the last decades, central banks significantly transformed. Moreover, after the 2008 year financial crisis, they got even more influence. Central banks usually have the right to increase the monetary base, set interest rates, and regulate commercial banks and financial institutions.

The central bank's independence is vital to reduce the level and variability of inflation and prevent political influence (Alesina and Summers, 1993; Crowe and Meade, 2007).

More transparency is another important transformation of central banks (Crowe and Meade, 2007). For example, the Federal Open Market Committee started announcing its decisions on the federal funds rate target in 1994 and begun a shift towards transparency in central bank's policies for the last decades. Before that, such policies were quite closed and unexpected (Geraats, 2002; Blinder et al., 2008), which created additional uncertainties and risks.

Major goals for monetary policy are the low unemployment rate, the stability of prices, and the growth of the economy (Friedman, 1995). The interest rate is used to achieve these goals. One approach to quantify this is the Taylor principle that shows that if inflation grew by percent, the interest rate should be increased by the central bank by more than percent (Taylor, 1999). The higher the interest rate, the lower the money supply is. An increase of a rate is used to cool down the economy and hold inflation. On the other hand, a decrease of the interest rate bolsters economic activity. Usually, there is an inverse relationship between the interest rate and stock market and direct one with the bonds market.

The Federal Reserve System - is the central banking system of the United States of America. It was created by the Federal Reserve Act in 1913, and reformed by the Federal Reserve Reform Act of 1977. The dual mandate of Federal Reserve from Congress to "promote effectively the goals of maximum employment, stable prices, and moderate long term interest rates" (Steelman, 2011).

2.3 Informational component of monetary policy decision

With the rise of transparency of monetary policy decision, not only policy actions themselves are important, but the role of communication has been increasing to a highly significant level. For example, on 28 January 2004, market participants expected that FOMC would not change the monetary policy, which is what happened. However, in the press statement phrase "considerable period" was dropped in comparison with the previous statement. Such change led to a 2% indices decrease - number compared to policy change effect (Rosa, 2011).

Therefore not only policy decisions play a vital role in the asset reactions, but also the central bank announcements about future policy intentions are an essential driver of stock market prices (Rosa, 2011). Deconstructing an influence of informational component from monetary policy actions shows that the effect of a monetary policy announcement on the economy is affecting the stock prices in addition to the response to policy action (Jarocinski and Karadi, 2018).

2.4 Market expectation

To calculate the market expectation of monetary policy action, we need to use 30 Day Federal Funds Futures. Usually, they used to hedge effective federal funds rates volatility, but also may be used for speculations on monetary policy action decision. Futures embodies near-term expectations of the Fed Funds rate. One of the possible problems is that settlement of contracts determined by an arithmetic average of daily effective federal funds rates (EFFR) during the contract month. And that requires reverse averaging to get expected Federal Funds rate properly. Another possible problem - calculation of the expectation when a meeting is scheduled at the end of the month. Because at that time, a large portion of daily EFFR of that month is already known, and errors contribute more to calculated expectation (Kuttner, 2001; Bernanke and Kuttner, 2005). Besides that, worth to point out that getting to a narrow window for intraday change of price may contaminate results with risk premia, as discussed in (Piazzesi and Swanson, 2008).

An alternative proxy for a market expectation of monetary policy action is eurodollar futures. Eurodollar - is dollars deposited at a bank outside of the USA. The price of futures contracts reflects the market expectation of a 3-months dollar LIBOR interest rate at the moment of maturity of contract (Gürkaynak, Sack, and Swanson, 2007).

To calculate the particular expectation from 30 Day Federal Funds Futures, we should differentiate whether there was an FOMC meeting in the month immediately before "meeting" month. If there was a meeting month before, we would use the month after contract for calculation. Otherwise - from the month before (CME Group, 2017).

If there was meeting calculation formula for expectated action is following:

$$\Delta i = (100 - p_{m+1}^{(current)}) - (\frac{D}{d} * [(100 - p_m^{(current)}) - \frac{D-d}{D}(100 - p_{m+1}^{(current)})]) \quad (2.1)$$

where D - is number day in the month, d - meeting day - 1, $p_{m+1}^{(current)}$ - current price of next month contract, $p_m^{(current)}$ - current price of this month contract.

Otherwise, calculation formula for expectated action is following:

$$\Delta i = \frac{D}{D-d}[(100 - p_m^{(current)}) - \frac{d}{D}(100 - p_{m-1}^{(end)})] - (100 - p_{m-1}^{(end)}) \quad (2.2)$$

where D - is number day in the month, d - meeting day - 1, $p_m^{(current)}$ - current price of this month contract, $p_{m-1}^{(end)}$ - last price of previous month

2.5 Event study

The classical event study for the stock market should have defined events and periods over which prices will be examined. After that, abnormal and normal returns are measured. For stock i and time T abnormal return defined as:

$$AR_{iT} = R_{iT} - E(R_{iT|X_T})$$

where $AR_{iT}, R_{iT}, E(R_{iT|X_T})$ are abnormal, actual and expected returns for T. X_T is the conditioning information for the normal return model.

With the parameter estimates for the normal model, we can calculate abnormal returns. And perform statistical test whether differences are significant (MacKinlay, 1997).

2.6 TF-IDF

The Term Frequency - Inverse Document Frequency is an approach that measures word importance inside the document based on word usage in the entire corpus. The main idea of the algorithm is to find a relative frequency of a word in the specific document in comparison to the inverse proportion of that word in all documents. Which represents how relevant the word is for the document. For example, articles and other common support words usually have negligible TF-IDF scores.

$$w_d = f_{w,d} * log\frac{|D|}{f_{w,D}} \tag{2.3}$$

where $f_{w,d}$ - is the number of times where w exists in d, $|D|$ is the size of the corpus, and $f_{w,D}$ - number of documents that contains w.

Therefore, when we multiply term frequency by inverse document frequency, we penalize words that usual for our corpus and select ones that are frequent for a particular document (Ramos, 2003).

2.7 BERT

Bidirectional Encoder Representations from Transformers (Devlin et al., 2018) is language representations model. BERT model architecture is multi-layer bidirectional Transformer encoder based on the (Vaswani et al., 2017) - original transformer model architecture (see Figure 2.2). Pre-trained models that is one of contributions of original authors exists in two sizes:

- $BERT_{BASE}$ with 12 layers (transformer blocks), 768 hidden size, 12 self-attention heads, total 110M parameters
- $BERT_{LARGE}$ with 24 layers (transformer blocks), 1024 hidden size, 16 self-attention heads, total 340M parameters

There are two main stages in the framework: pre-training and fine-tuning (see Figure 2.3).

Pre-training is done using two unsupervised tasks: masked language model (MLM) and the next sentence prediction (NSP). To train bidirectional representation in MLM, authors randomly mask 15% of all WordPiece tokens in each input and predict masked tokens. The NSP task is used to pre-train a model for Question

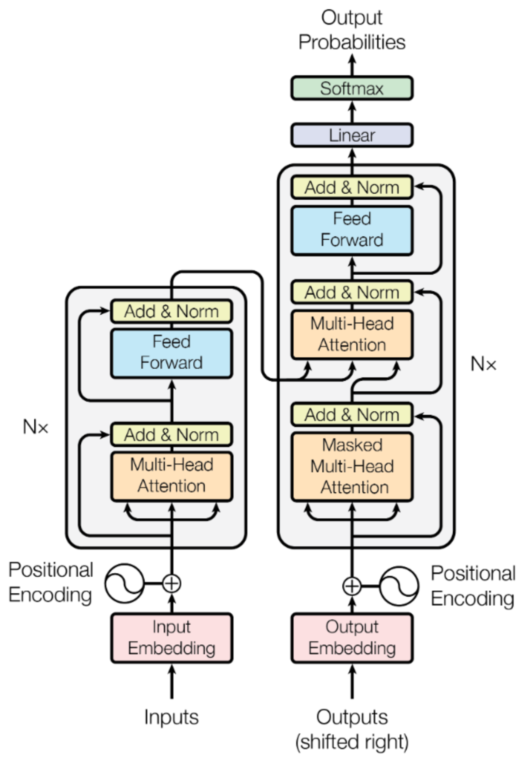

FIGURE 2.2: The Transformer - model architecture Source: (Vaswani et al., 2017)

Answering and Natural Language Inference tasks to teach a model understanding of a relationship between two sentences. For pre-training used data is BooksCorpus and English Wikipedia.

For fine-tuning, inputs and outputs added depending on the task. After that, we can fine-tune all parameters end-to-end. Therefore it is possible to download the pre-trained model and relatively quickly adapt it to a particular task that is required.

2.8 Multinomial Logistic Regression

In multiclass logistic regression the posterior probabilities are given by softmax transformation of linear functions of the feature variables:

$$p(C_k|\phi) = y_k(\phi) = \frac{exp(a_k)}{\sum_j exp(a_j)} \tag{2.4}$$

where 'activations' a_k are given by:

$$a_k = w_k^T \phi \tag{2.5}$$

Where ϕ is M-dimensional feature space, $p(C_k|\phi)$ - the posterior probability for class C_k, w_k - the parameters of model. To solve the equation, we can write down the likelihood function. Afterward, the optimization task with the cross-entropy loss is defined and minimized (Bishop, 2006).

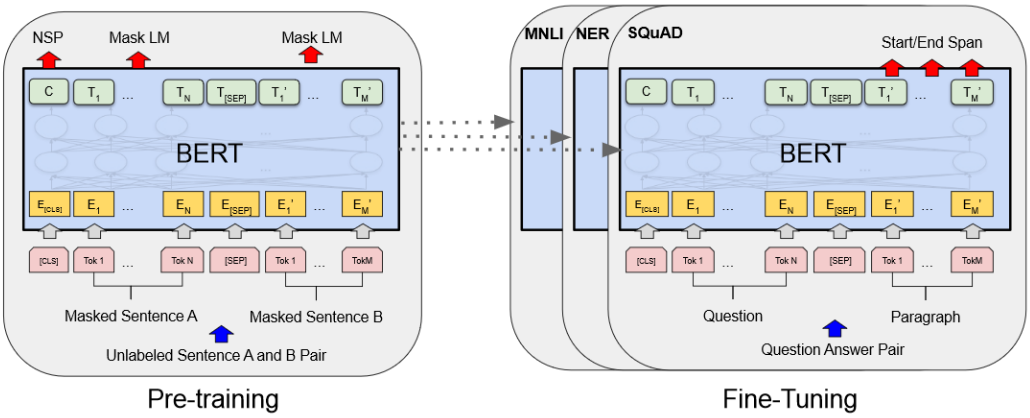

FIGURE 2.3: Pre-training and fine-tuning procedures for BERT Source: (Devlin et al., 2018)

2.9 Expected tone of the FOMC Statement

When monetary policy is determined, FOMC committee members look at the macroeconomic environment. Therefore, some expectation of the tone of the statement is built around key macroeconomic indicators. To calculate the expected tone of the FOMC statement, we can use the methodology developed by (Rosa, 2011):

$$Index^*_t = \gamma_1 Index^{OLD}_{t-5} + \gamma_2 PMI_{t-5m} + \gamma_3 \pi^e_{t-5m} + \gamma_4 Slope_{t-5m} + \epsilon_t \tag{2.6}$$

where $Index^*_t$ is the expected tone of FOMC statement, $Index^{OLD}_{t-5}$ is real tone of the previous statement, PMI_{t-5m} - purchasing manager index before statement, π^e_{t-5m} - median inflation expectation over next 12 months, $Slope_{t-5m}$ slope of the short-term yield curve 5 minutes before meeting, and proxied via difference of implied rate between 3 month ahead 30 Day Federal Future and current month contracts.

To convert results of this regression model to clasess, set of threshold values (δ_1, δ_2) is used

$$\begin{aligned} Index_t &= -1 \text{ if } Index^*_t \leq \delta_1 \\ Index_t &= 0 \text{ if } \delta_1 \leq Index^*_t \leq \delta_2 \\ Index_t &= 1 \text{ if } \delta_2 \leq Index^*_t \end{aligned} \tag{2.7}$$

And finally, the market participants expectation about central bank announcement calculated by:

$$E_{t-5m}[Index^{NEW}_t] = \sum_{i=-1}^{+1} Pr(Index^{NEW}_t = i) * i \tag{2.8}$$

where $E_{t-5m}[\cdot]$ is the expectation conditional on the information abailable 5 minutes before event and $Pr(Index^{NEW}_t = i)$ for $i = -1, 0, 1$ is computed by probit model.

3 Dataset Description

3.1 Federal Open Market Committee statements

The Federal Open Market Committee provides a press release usually issued at 12:30 p.m. or 2 p.m. to the public on the committee meeting day with a statement regarding its a policy decision. I have collected all statements for scheduled meetings from May 18, 1999, meeting to the December 10-11 2019 meeting. Overall - 174 items, varying from 527 to 6085 symbols.

The Source of FOMC statements is monetary policy materials from the official Federal Reserve web site[1]. Underlying data from monetary materials widget may be accessed in the JSON files for recent[2] and historical[3] events. Each press release located on the separate page, web-address for which may be located from mentioned JSON files. For example, on December 11, 2019, Federal Reserve issued the FOMC statement[4].

Although Federal Reserve web-site data have the date of each statement, they do not have time. To enhance data with time, for the period from May 18, 1999, to June 23, 2010, I have used data from the technical appendix table "Coding of FOMC statements" in (Rosa, 2011). For later period I used news data from bloomberg.com, reuters.com (Kulkarni, 2018) and FOMC meetings schedule directly[5] or through Wayback Machine[6].

Besides that, statements classification performed by (Rosa, 2011) for the period from May 18, 1999, to June 23, 2010, included as well.

3.2 Federal Funds Target Range

In the 2008 Federal Reserve switched monetary policy announcements from the target level to target range with 25 basis points spread. Both the upper and lower bound of this range are available as daily data starting from December 16, 2008. Changes

[1] Board of Governors of the Federal Reserve System (US), Federal Open Market Committee materials, retrieved from Federal Reserve System; https://www.federalreserve.gov/monetarypolicy/materials/, December 12, 2019

[2] Board of Governors of the Federal Reserve System (US), Federal Open Market Committee materials, retrieved from Federal Reserve System; https://www.federalreserve.gov/monetarypolicy/materials/assets/final-recent.json, December 12, 2019

[3] Board of Governors of the Federal Reserve System (US), Federal Open Market Committee materials, retrieved from Federal Reserve System; https://www.federalreserve.gov/monetarypolicy/materials/assets/final-hist.json, December 12, 2019

[4] Board of Governors of the Federal Reserve System (US), Federal Open Market Committee, retrieved from Federal Reserve System; https://www.federalreserve.gov/newsevents/pressreleases/monetary20191211a.htm, December 12, 2019

[5] Board of Governors of the Federal Reserve System (US), Federal Open Market Committee, retrieved from Federal Reserve System; https://www.federalreserve.gov/monetarypolicy/fomccalendars.htm, December 12, 2019

[6] Internet Archive, Wayback Machine, retrieved from Internet Archive; https://web.archive.org/web/https://www.federalreserve.gov/monetarypolicy/fomccalendars.htm, December 12, 2019

to target range usually made at scheduled FOMC meetings and are included in text form into FOMC statements.

The source is Federal Reserve Bank of St. Louis web site[7][8].

3.3 Stock data

Stock data were collected as intraday 1-minute OHLC data for ETFs SPY and AGG.

SPY (SPDR S&P 500 Trust ETF) is an exchange-traded fund listed on the NYSE and designed to track the SPX (Standard & Poor's 500 Index). It is the largest ETF by assets under management, which includes 500 largest US companies listed on stock exchanges and is widely accepted as an indicator for US stock market performance. Data for the period from January 3, 2011, 9:30 a.m. to November 29, 2019, 3:30 p.m. was scrapped from the historical download option on the Barchart web site[9].

AGG (iShares Core US Aggregate Bond ETF) is iShares ETF that tracks investment results of an index composed of the total US investment-grade bonds. Furthermore, it is the largest US bond ETF by assets under management. Data for the period from January 3, 2011, 9:30 a.m. to November 29, 2019, 3:30 p.m. was scrapped from the historical download option on the Barchart web site[10].

Following columns are included for both SPY and AGG:

- Time - EST (Eastern Standard Time) timestamp
- Open - the price at the start of a minute
- High - the highest price in a minute
- Low - the lowest price in a minute
- Last - close price in a minute
- Change - the difference between the last price of the previous minute and current minute
- Volume - the number of trades in a minute

3.4 30 Day Federal Funds Futures

30 Day Federal Funds Futures - is the futures contracts that are traded on Chicago Mercantile Exchange (CME), and each settled at the end of the month. Settlement of contracts determined by an arithmetic average of daily effective federal funds rates (EFFR) during the contract month. Typically the price is 100 minus the expectation of an arithmetic average of daily EFFR. Contracts traded for the first 36 months.

Contract ticker is constructed as following - ZQ prefix + month code letter + two last number of a year. For example, Federal Funds Futures for December 2019 will

[7] Board of Governors of the Federal Reserve System (US), Federal Funds Target Range - Lower Limit [DFEDTARL], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/DFEDTARL, January 4, 2020

[8] Board of Governors of the Federal Reserve System (US), Federal Funds Target Range - Upper Limit [DFEDTARU], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/DFEDTARU, January 4, 2020

[9] barchart, SPY intraday prices, retrieved from barchart; https://www.barchart.com/futures/quotes/SPY, November 30, 2019

[10] barchart, AGG intraday prices, retrieved from barchart; https://www.barchart.com/futures/quotes/AGG, November 30, 2019

have ticker ZQZ19. Months are encoded as follows: F - January, G - February, H - March, J - April, K - May, M - June, N - July, Q - August, U - September, V - October, X - November, Z - December.

I have collected the intraday data for ZQF11 to ZQZ19 (all future contacts from 2011 to 2019). Three years prior settlement for each ticker, 108 tickers overall.

All data is scrapped from the historical download option on the Barchart web site[11].

Following columns are included:

- Time - CST (Central Standard Time) timestamp
- Open - the price at the start of a minute
- High - the highest price in a minute
- Low - the lowest price in a minute
- Last - close price in a minute
- Change - the difference between the last price of the previous minute and current minute
- Volume - the number of trades in a minute

3.5 US Financial News Articles

As the financial news data source, I have used the dataset from Kaggle[12]

The dataset contains 306242 articles in JSON format. Articles were collected from the following sources: Bloomberg.com, CNBC.com, reuters.com, wsj.com, fortune.com. Publication time varies from January 2018 to May 2018.

[11] barchart, AGG intraday prices, retrieved from barchart; `https://www.barchart.com/futures/quotes/ZQF20/historical-prices?viewName=main&page=all`, November 30, 2019

[12] jeet2016, US Financial News Articles, retrieved from kaggle; `https://www.kaggle.com/jeet2016/us-financial-news-articles`, November 30, 2019

4 Solution pipeline

Although the research question of this work is related to extract the meaningful features from the FOMC statement, to work with that part, we need to build a full pipeline of the solution to a larger goal - prediction of SPY ETF reaction on monetary policy event as a whole. This includes the following steps:

1. Create a feature with the difference of market expectation of policy action and the action itself.

2. Create a feature with the difference between the market expectation of tone of FOMC statement and classified tone of the statement itself.

3. Conduct an event study to quantify the reaction of certain stocks on the monetary policy event.

4.1 Market Expectation of policy action

As discussed in Section 2.4 we can use 30 Day Federal Funds Futures or Eurodolar futures to calculate market expectation. Consistent with related works I decided to use 30 Day Federal Funds Futures for my calculation. And used collected futures prices data described in Section 3.4 and timestamps of events described in Section 3.1.

Although FOMC usually changes policy with a step of 25 basis points. We want to obtain the exact value of expectation. The general methodology follows (CME Group, 2017), with slight alterations, because that methodology goal is to calculate probability and not the expectation.

Two possible cases exist: if there was a meeting in the month before or not. The existence of a meeting implies possible changes to monetary policy and will contaminate the price of a futures contract, and it will not be a good proxy or estimation. Therefore, if the previous month is without a meeting, we can use the last price before settlement and Formula 2.2. In case if a meeting does exist, we can rely on the prices of next month's contract and Formula 2.1. From the nature of the schedule of FOMC meetings - there are no three consecutive months with meetings, therefore if the previous month had meeting - the following would not have one. With exception to unscheduled meetings, which for the recent period did not include FOMC statement and was not included in my analysis. Also, such approach offers some robustness of results in the second case from end-of-month behavior described in (Kuttner, 2001)

Intraday price may be contaminated results with risk premia, as discussed in (Piazzesi and Swanson, 2008). To choose the best approach, I have calculated using different level of aggregation:

- daily prices
- 30-minutes OLHC intraday prices

- 5-minutes intraday prices

For my setting, intraday data showed a somewhat smaller difference between action expectation and action. Therefore, I have used the closing price 30 minutes before the FOMC statement.

4.2 Policy statement classification

The classification of the policy statement is based on the view of FOMC on future policy tilt. This tilt is based on the expectation of committee members on economy development via macroeconomic indicators and perception of future movement in key indicators included in the central bank mandate that discussed in Section 2.2. For example, we can consider the following cases:

- On the August 9, 2011 meeting statement included the following: "are likely to warrant exceptionally low levels for the federal funds rate at least through mid-2013" when the previous one had "are likely to warrant exceptionally low levels for the federal funds rate for an extended period". That change showed a shift in views of FOMC regarding its future policy and was unanticipated by the market. 3 out of 10 present committee members did not support that change in the statement. This is an example of the so-called "hawkish" statement.

- April 29-30, 2014 meeting included the "The Committee sees the risks to the outlook for the economy and the labor market as nearly balanced." which may be strong evidence of neutral statement.

- On the March 17-18, 2015 meeting statement included the "the Committee judges that an increase in the target range for the federal funds rate remains unlikely at the April FOMC meeting. The Committee anticipates that it will be appropriate to raise the target range for the federal funds rate when it has seen further improvement in the labor market and is reasonably confident that inflation will move back to its 2 percent objective over the medium term" - which indicated lowered committee members expectation of near-future rate and shifted increase of rate somewhat further than expected by the market. This is an example of the so-called "dovish" statement.

Altogether, since I do not hold any economy degree to label statements, I have decided to use for training existing classified data from technical appendix table "Coding of FOMC statements" from (Rosa, 2011). The data contains classes of 96 statements for the period from May 18, 1999, to June 23, 2010. Data that is used for this work is more recent 71 statements for the period from January 26, 2011, to October 30, 2019. And the idea is to create a classifier based on (Rosa, 2011) data, and using it label the more recent statements.

4.2.1 TF-IDF

For TF-IDF, I follow the classical methodology described in Section 2.6 to build document representation. Although there are two different approaches. The first one is to build the TF-IDF on the corpus of the FOMC statements. The second one is to build the TF-IDF on the larger corpus of financial news articles.

In both options after execution, we have transformed FOMC statements that are ready for classification. However, since for each word in the corpus, we have a

FIGURE 4.1: Two word clouds from FOMC statements. Left - is for data for the period from May 18, 1999, to June 23, 2010. Right - for the period from January 26, 2011, to October 30, 2019.

feature (1272 for the first option and 118634 for the second) and only 96 data points, we can reduce dimensions by using SVD. We save the seven most important features that attribute 40-47% of explained variability in data (depending on the particular setting).

After that, we can classify documents using Multinomial Logistic Regression described in Section 2.8. The train and test split is done chronologically. First 50% go to train and last 50% to test because we need to have some representation from all three classes in both train and test and that the most convenient way, without introducing data leakage.

4.2.2 BERT

The second approach is to use for classification of FOMC statements is BERT described in Section 2.7. We use the pre-trained model $BERT_{BASE}$ and fine-tune it on our data. BERT allows up to 512 tokens on input, and that is sufficient for our FOMC statements. The main reasoning is to use a pre-trained model in this context because we have a small number of texts, and a pre-trained model may help.

For fine-tuning, we add one softmax layer on the output and train model on our data. The train and test split is done chronologically. First 50% go to train and last 50% to test because we need to have some representation from all three classes in both train and test and that the most convenient way, without introducing data leakage.

4.3 Event study methodology

After we have calculated the market expectation and classified the FOMC statement, we can introduce two important variables that are required for our model consistent with the framework of related work (Rosa, 2011). The first one is Monetary policy shock:

$$MPS_t = MPE_{t-30m} - MPA_t \tag{4.1}$$

where MPS_t is Monetary policy shock on time t, MPE_{t-30m} - Monetary policy expectation 30 minutes before event, calculation described in Section 4.1, MPA_t - monetary policy action anounced on time t - calculated from 3.2.

The second one is News shock:

$$NS_t = Index_t^{NEW} - E_{t-5m}[Index_t^{NEW}] \tag{4.2}$$

where NS_t is the news shock of event, $Index_t^{NEW}$ is the class of FOMC statement obtained from classification model explained in Section 4.2, $E_{t-5m}[Index_t^{NEW}]$ - expected tone of FOMC statement calculated as described in Section 2.9.

With both required features MPS_t and NS_t constructed, we can move forward to the econometric model, consistent with related work (Rosa, 2011) - Ordinary Least Squares with Heteroskedasticity-Consistent standard errors (White, 1980).

$$100 * log(\frac{Price_{t+25m}}{Price_{t-5m}}) = \beta_0 + \beta_1 MPS_t + \beta_2 NS_t + \epsilon_t \tag{4.3}$$

where $100 * log(\frac{Price_{t+25m}}{Price_{t-5m}})$ is cumulative return of particular stock for 30 minutes from 5 minutes before event to 25 minutes after.

5 Evaluation

The evaluation of our pipeline is divided into two parts:

- Evaluate Market Expectation of policy action via econometric model based on the MPS_t feature and compare the result with related work.
- Evaluate TF-IDF and BERT based classifications of FOMC statements using confusion matrix and F1 score.

5.1 Market Expectation of policy action

To evaluate calculated market expectation of policy action described in Section 4.1 I have used the econometric model defined in Section 4.3. For comparison, the proposed results from the appendix table "The response of ETF indices to Fed decisions and announcements" (Rosa, 2011). Both approaches using the same model, although using the data from different time periods. Table 5.1 displays the response of SPY ETF to the monetary policy shock.

	My SPY	**Other SPY**
Constant	$0.1573^{**}(0.064)$	$-0.066(0.056)$
MPS_t	$-5.7908^{*}(2.949)$	$-3.162^{***}(1.161)$
Adj R^2	0.031	0.013
Observations	71	65
F-test	3.856^{*}	7.69^{***}

TABLE 5.1: Comparison table for the response of SPY to FOMC policy action. *My SPY* reflects results of Ordinary Least Squared econometric model with Heteroskedasticity-Consistent standard errors in brackets. Observations on days of FOMC meetings, January 26, 2011 - October 30, 2019. Reference results (Rosa, 2011) *Other SPY* calculated using same approach, using observations on days of FOMC meetings May 1999 - June 2007. The superscripts ***, **, * indicate statistical significance at the 1%, 5% and 10% level respectively.

As an interpretation of our results, a 25 basis points surprise easing in the federal funds rate increase the SPY ETF price by 1.45% in the 25 minutes after FOMC statement, significant at the 10% level. For example, on December 19, 2018, meeting market participants expectation was 19 basis points tightening, while actual action was 25 basis points tightening. 6 basis points surprise "tightening" resulted in the -0.6% return on SPY in 25 minutes after the event (see Figure 5.1). Actual cumulative returns 25 minutes after FOMC statement usually less than 1% (see Figure 5.2).

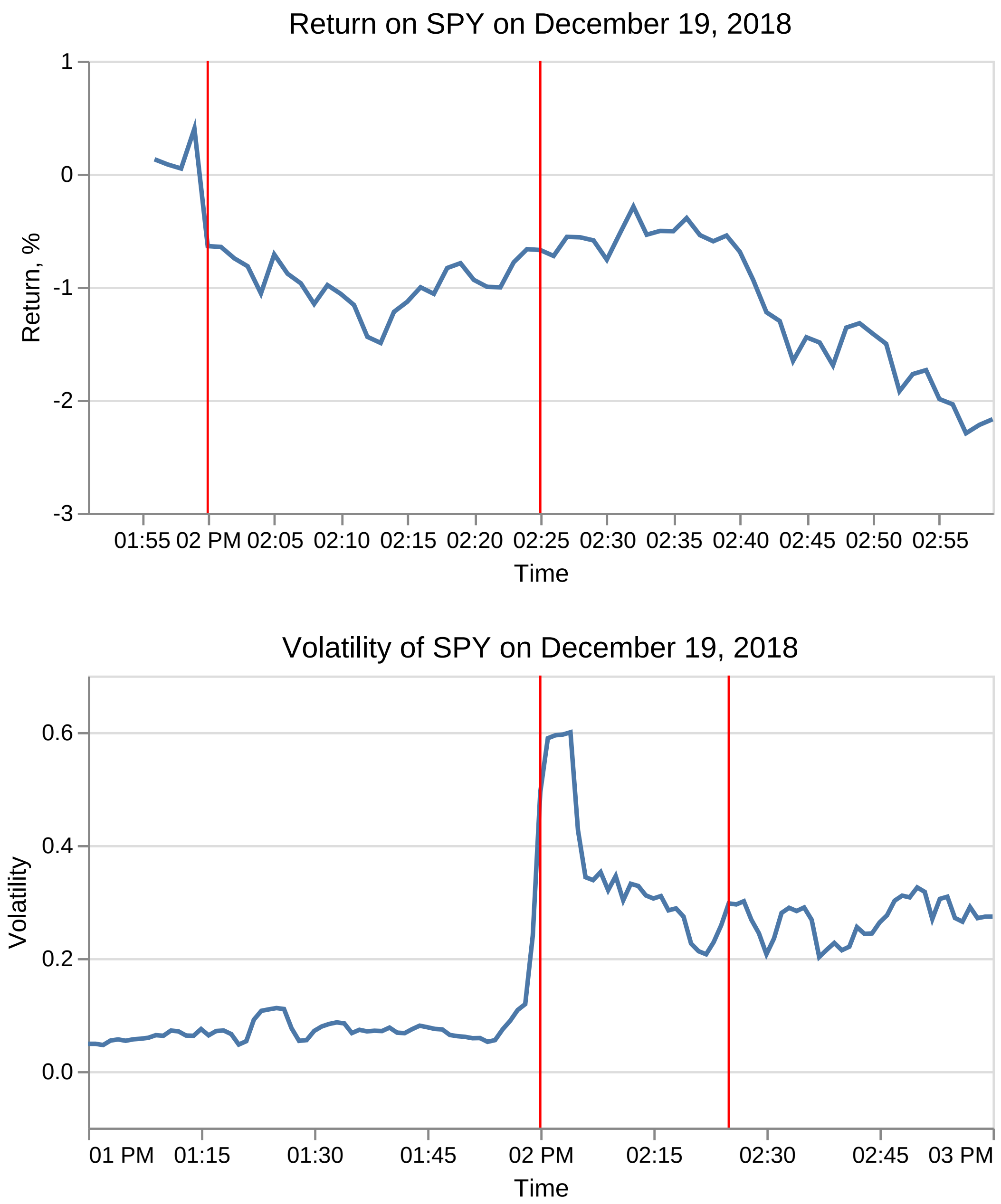

FIGURE 5.1: Return and Volatility on SPY on December 19, 2018. First vertical line marks the time of FOMC statement release, second - 25 minutes after. Volatility is estimated using Rogers-Satchell approach (Rogers and Satchell, 1991).

5.2 Policy statement classification

5.2.1 TF-IDF

We have two different TF-IDF approaches, as described in Section 4.2.1. The first one is to build TF-IDF on the corpus of only FOMC statements. Second - is to build TF-IDF on the corpus of financial news articles. After that, we train logistic regression on the first 50% of FOMC statements and evaluate on the others. Resulting confusion matrixes (see Figure 5.3) and metrics tables show 5.2 that option one is better than option two. This might be to the fact that financial news corpus has articles for

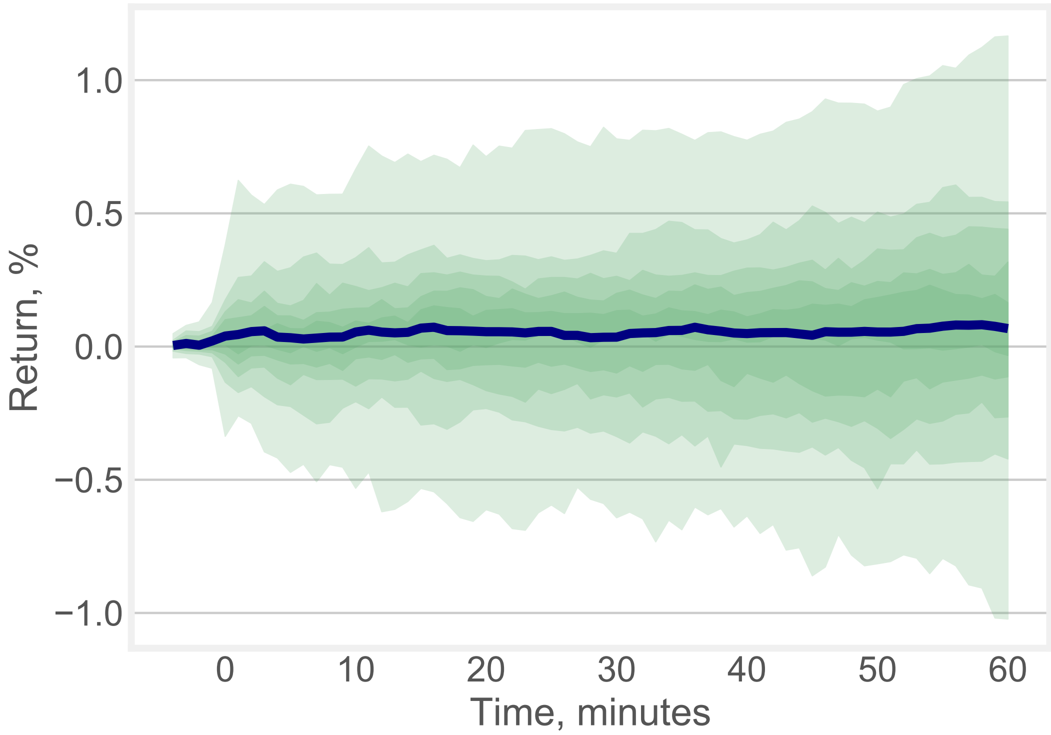

FIGURE 5.2: Fan chart of cumulative returns on SPY an hour after the FOMC statement release. The range of the fan chart is between 5 and 95 percentile. Blue line marks mean return. Observations on days of FOMC meetings, January 26, 2011 - October 30, 2019.

only five months of 2018, while FOMC statements are for more than a decade, and different wording is used. Another thing that should be mentioned is possible data leakage in the second option because we have a corpus of texts from the future and build TF-IDF on it.

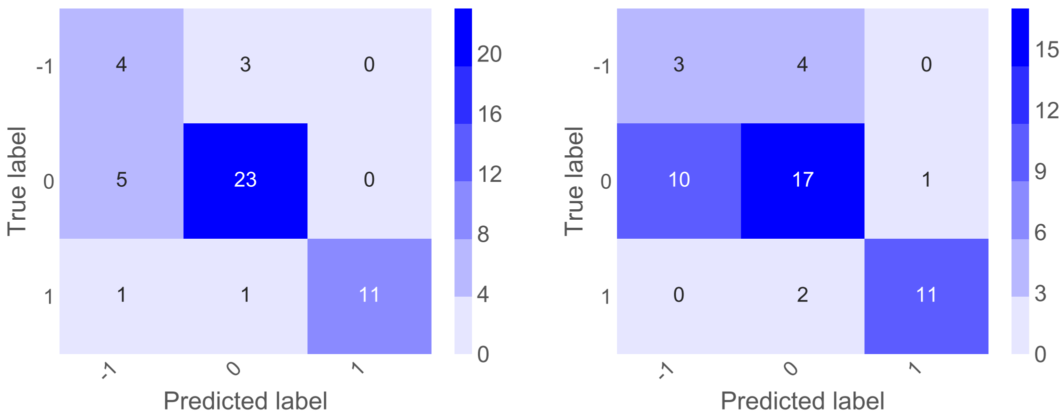

FIGURE 5.3: Confusion matrixes for classification by Logistic Regression based on two datasets transformed by different TF-IDF options. Left - is option one, build TF-IDF on FOMC Statements, and right is option two - build TF-IDF on financial news corpus.

Although we have decent classification accuracy on FOMC statements from the 1999-2010 period, and approach looks sustainable. When we train both options on

	TF-IDF 1	TF-IDF 2
Accuracy	0.79	0.65
Precision	0.83	0.71
Recall	0.79	0.65
F1 Score	0.80	0.67

TABLE 5.2: Comparison table with classification metrics of Logistic Regression build on two datasets transformed by different TF-IDF options. *TF-IDF 1* reflects results of TF-IDF build on corpus of FOMC statements. *TF-IDF 2* reflects results of TF-IDF build on corpus of financial news. FOMC statements used for train is May 1999 - November 2004. FOMC statements used for evaluation is December 2004 - June 2010. Target variable - classification done by (Rosa, 2011).

the whole labeled dataset from 1999 to 2010 and predict classes for FOMC Statements for 2011 to 2019, all of them are classified as neutral - 0. Which is simply impossible. I attribute that to the changes in the nature of FOMC statements related to increasing complexity and length of described by (Hernández-Murillo and Shell, 2014) (see Figure 2.1) and more similar content from meeting to meeting (Meade and Acosta, 2015). Some differences may be noticed on word clouds build from FOMC statements of both periods (see Figure 4.1)

For example, one of the FOMC statements that was classified by the model as neutral is from August 9, 2011. Although, the expectation of the market was 2.6 basis point tightening, and no action happened. The market reacted to FOMC Statement with the highest volatility on the meeting day in the period of the 2011-2019 year and resulted in the -2.2% return on SPY in 25 minutes after the event (see Figure 5.4). That reaction was caused by news shock, but our model was not able to capture that.

5.2.2 BERT

My approach for BERT classification was desctibed in Section 4.2.2. For implementation I have used the simpletransformers[1] library which is build on top of Transformers (Wolf et al., 2019) library. After grid search of possible hyperparameters and other tricks, I was not able to force a model to train on 48 FOMC statements. In the current state, every FOMC Statement in the evaluation set classified as "dovish", which is the most represented class in the train set. Furthermore, when I trained model on all labeled statements - all recent FOMC statements were classified as neutral.

[1] https://github.com/ThilinaRajapakse/simpletransformers

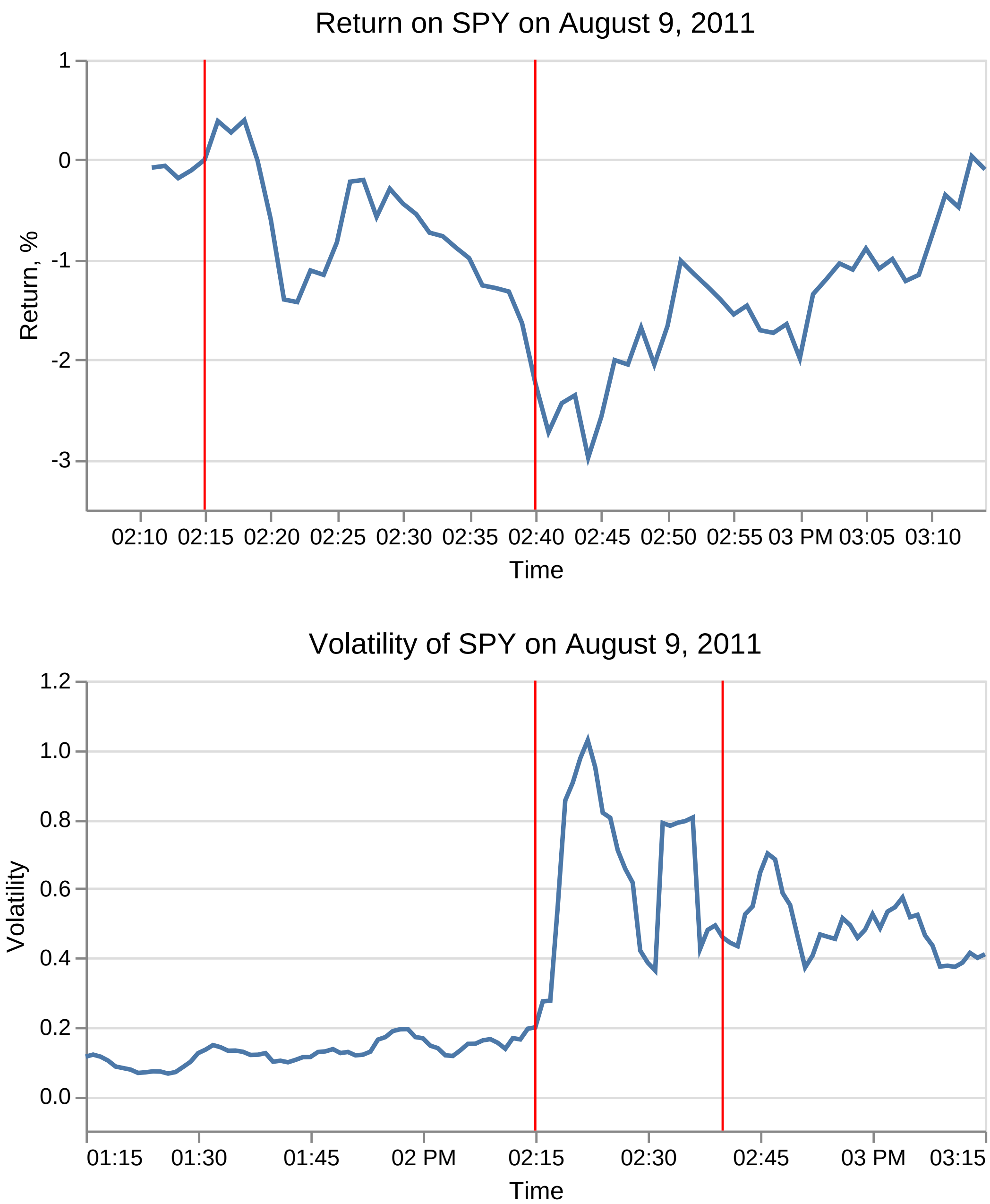

FIGURE 5.4: Return and Volatility on SPY on August 9, 2011. First vertical line marks the time of FOMC statement release, second - 25 minutes after. Volatility is estimated using Rogers-Satchell approach (Rogers and Satchell, 1991).

6 Conclusion

6.1 Contribution

The following is the list of goals achieved, previously defined in Section 1.3:

1. I collected dataset with FOMC statements with release timestamps, policy action, and market expectation of policy action, SPY cumulative return for 30-minute event window. Knowing the difficulties of obtaining financial data for research, discussed in Chapter 3 approaches, may be useful for a wide range of activities in the financial domain.

2. I have built a set of models that perform the classification of the FOMC statements with the best accuracy equal to 79%. Due to a small dataset of FOMC statements and changes in their nature over time discussed in Section 5.2.1 model does not capture differences in recent FOMC statements.

3. I have built an econometric model on the monetary policy shock for the recent period and compared it to related work. Since I do not have labeled classification of FOMC statements as mentioned previously, they were not included in the econometric model.

As an answer to Research Question defined in Section 1.3: **Can the FOMC press release in a monetary policy decision be utilized using natural language processing techniques to create statistically significant features that explain the stock market reaction?** I can state that after experiments conducted in Section 5.2.1 I see that model trained on 48 FOMC statements able to classify correctly 79% of the next 48 statements that were labeled by a human. This allows me to conclude that it is possible to create a significant feature out of FOMC Statements. Although it does not contribute sufficiently to the more broad goal of prediction of stock returns due to the changes in FOMC statements nature. And more research on the topic is required.

6.2 Future work

Recent deep learning studies on financial data shows that a hot topic of interest is the usage of news from regular sources such as Bloomberg or Reuters to data from twitter. The main idea behind is to train on a large set of articles or even titles, and gather sentiment that might be used for stock prediction.

Contrary, our approach is to take a small amount of highly important articles - FOMC statements and evaluate them. I think that it is possible to achieve better results by combining approaches, by firstly train the neural network on some larger corpus of financial sentences, and only afterward fine-tune further by labeled sentences taken out of the FOMC statements. Basically - we need to relabel data sentence by sentence to achieve better results.

Besides that, a model that trains on n recent FOMC statements and evaluates only the next one may perform significantly better. Due to the fact that recent statements tend to be more similar than distant ones.

A Supplementary data

Supplementary data to this thesis that is eligible for redistribution is available upon request from the author.

Bibliography

Alesina, Alberto and Lawrence H Summers (1993). "Central bank independence and macroeconomic performance: some comparative evidence". In: *Journal of Money, credit and Banking* 25.2, pp. 151–162.
Andersen, Torben G et al. (2007). "Real-time price discovery in global stock, bond and foreign exchange markets". In: *Journal of international Economics* 73.2, pp. 251–277.
Andersson, Magnus (2007). "Using intraday data to gauge financial market responses to Fed and ECB monetary policy decisions". In:
Araci, Dogu (2019). "FinBERT: Financial Sentiment Analysis with Pre-trained Language Models". In: *arXiv preprint arXiv:1908.10063*.
Bernanke, Ben S and Kenneth N Kuttner (2005). "What explains the stock market's reaction to Federal Reserve policy?" In: *The Journal of finance* 60.3, pp. 1221–1257.
Bishop, Christopher M (2006). *Pattern recognition and machine learning*. springer.
Blinder, Alan S et al. (2008). "Central bank communication and monetary policy: A survey of theory and evidence". In: *Journal of Economic Literature* 46.4, pp. 910–45.
CME Group (2017). *Understanding the CME Group FedWatch Tool and Fed Funds Futures Probability Tree Calculator*. https://www.cmegroup.com/education/demos-and-tutorials/fed-funds-futures-probability-tree-calculator.html.
Crowe, Christopher and Ellen E Meade (2007). "The evolution of central bank governance around the world". In: *Journal of Economic Perspectives* 21.4, pp. 69–90.
Devlin, Jacob et al. (2018). "Bert: Pre-training of deep bidirectional transformers for language understanding". In: *arXiv preprint arXiv:1810.04805*.
Ding, Xiao et al. (2015). "Deep learning for event-driven stock prediction". In: *Twenty-fourth international joint conference on artificial intelligence*.
Ehrmann, Michael and Jonathan Talmi (2019). "Starting from a blank page? Semantic similarity in central bank communication and market volatility". In: *Journal of Monetary Economics*.
Faust, Jon et al. (2007). "The high-frequency response of exchange rates and interest rates to macroeconomic announcements". In: *Journal of Monetary Economics* 54.4, pp. 1051–1068.
Friedman, Milton (1995). "The role of monetary policy". In: *Essential Readings in Economics*. Springer, pp. 215–231.
Geraats, Petra M (2002). "Central bank transparency". In: *The economic journal* 112.483, F532–F565.
Gürkaynak, Refet S, Brian P Sack, and Eric T Swanson (2007). "Market-based measures of monetary policy expectations". In: *Journal of Business & Economic Statistics* 25.2, pp. 201–212.
Hernández-Murillo, Rubén, Hannah Shell, et al. (2014). "The rising complexity of the FOMC statement". In: *Economic Synopses* 23.
Hiew, Joshua Zoen Git et al. (2019). "BERT-based Financial Sentiment Index and LSTM-based Stock Return Predictability". In: *arXiv preprint arXiv:1906.09024*.
Jarocinski, Marek and Peter Karadi (2018). "Deconstructing monetary policy surprises: the role of information shocks". In:

Kearns, Jonathan, Phil Manners, et al. (2006). "The impact of monetary policy on the exchange rate: A study using intraday data". In: *International Journal of Central Banking* 2.4, pp. 157–183.

Kulkarni, Rohit (2018). *The Historical Reuters News-Wire*. DEACCESSIONED VERSION. DOI: https://doi.org/10.7910/DVN/XDB74W.

Kuttner, Kenneth N (2001). "Monetary policy surprises and interest rates: Evidence from the Fed funds futures market". In: *Journal of monetary economics* 47.3, pp. 523–544.

MacKinlay, A Craig (1997). "Event studies in economics and finance". In: *Journal of economic literature* 35.1, pp. 13–39.

Meade, Ellen E, Miguel Acosta, et al. (2015). *Hanging on every word: Semantic analysis of the FOMC's postmeeting statement*. Tech. rep. Board of Governors of the Federal Reserve System (US).

National Bank of Ukraine (2019). "Inflation Report: April 2019". In: pp. 13–14.

Piazzesi, Monika and Eric T Swanson (2008). "Futures prices as risk-adjusted forecasts of monetary policy". In: *Journal of Monetary Economics* 55.4, pp. 677–691.

Ramos, Juan et al. (2003). "Using tf-idf to determine word relevance in document queries". In: *Proceedings of the first instructional conference on machine learning*. Vol. 242. Piscataway, NJ, pp. 133–142.

Rigobon, Roberto and Brian Sack (2004). "The impact of monetary policy on asset prices". In: *Journal of Monetary Economics* 51.8, pp. 1553–1575.

Rogers, L Christopher G and Stephen E Satchell (1991). "Estimating variance from high, low and closing prices". In: *The Annals of Applied Probability*, pp. 504–512.

Rosa, Carlo (2011). "Words that shake traders: The stock market's reaction to central bank communication in real time". In: *Journal of Empirical Finance* 18.5, pp. 915–934.

Steelman, Aaron et al. (2011). "The Federal Reserve's" dual mandate": the evolution of an idea". In: *Richmond Fed Economic Brief* Dec.

Taylor, John B (1999). "A historical analysis of monetary policy rules". In: *Monetary policy rules*. University of Chicago Press, pp. 319–348.

Vaswani, Ashish et al. (2017). "Attention is all you need". In: *Advances in neural information processing systems*, pp. 5998–6008.

White, Halbert et al. (1980). "A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity". In: *econometrica* 48.4, pp. 817–838.

Wolf, Thomas et al. (2019). "HuggingFace's Transformers: State-of-the-art Natural Language Processing". In: *ArXiv* abs/1910.03771.

Wongswan, Jon (2009). "The response of global equity indexes to US monetary policy announcements". In: *Journal of International Money and Finance* 28.2, pp. 344–365.