Stockholm School of Economics
Department of Economics
5350 Master's thesis in economics
Academic Year 2017-2018

Classification of Monetary Policy Decisions through Text Mining Techniques

Manuel von Krosigk (40888)

Abstract

Ellingsen, Söderström and Masseng (2003) describe an empirical strategy to test their model about how central bank interventions affect asset prices. They manually analyse a Wall Street Journal column in order to determine whether bond traders interpreted a target rate change as reaction to new information about the economy or change of central bank preferences. Instead this thesis tries to replicate their results during another time period by applying Machine Learning and Natural Language Processing techniques. Even though deterministic as well as trained algorithms are applied in order to achieve a consistent classification, not all of their hypotheses can be verified. This can likely be traced back to the extraordinary time period underlying the empirical test in this sample.

Keywords: Sentiment analysis, text mining, monetary policy, term structure of interest rates
JEL: C80, C88, E43, E52, E58

Supervisor: Jesper Lindé
Date submitted: 10 December 2017
Date examined: 19 December 2017
Discussant: Kalle Kantanen
Examiner: Mark Sanctuary

Acknowledgements:

I would like to thank my supervisor Jesper Lindé for taking the time and effort to share his expertise in monetary policy, economic research and the process of structuring academic work with me. His comments have been very helpful and guided me through the process of finishing this thesis.

Furthermore, I would like to express my gratitude towards Paul Söderlind who supported me from HSG side and contributed greatly to me staying on the right track from the beginning as well as Ferdinand Graf from the company d-fine who provided access to some of the newspaper articles and supported me with helpful comments about my empirical strategy.

Finally, my great thanks go out to my fellow classmates and lecturers at SSE who provided such an inspiring learning environment that really got me excited about all the topics we covered.

Contents

List of Abbreviations

BOJ Bank of Japan

BOE Bank of England

CB Count-based

ECB European Central Bank

Fed Federal Reserve Bank

FOMC Federal Open Market Committee

GSE Government-sponsored enterprise

KNN k-nearest neighbours

LDA Latent Dirichlet allocation

LSAP Large-scale asset purchases

MBS Mortgage-backed securities

ME Maximum entropy

ML Machine learning

NB Naïve Bayes

NLP Natural language processing

OMO Open market operation

SVM Support vector machines

tdm Term-document matrix

ZIRP Zero interest rate policy

ZLB Zero lower bound

List of Figures

List of Tables

1. Introduction

Monetary policy actions have very influential effects on capital markets. Especially target rate changes impact the bond market directly and hence affect interest rates of all maturities. Ellingsen and Söderström (2001), however, call attention to the limitations of explanatory models and dissent among scholars about the ramifications of target rate adjustments by central banks, in particular when it comes to the response of the yield curve. More precisely, empirical evidence shows short and long term market rates moving in the same direction after most monetary policy changes while occasionally the yield curve tilts which to this point, no coherent theory can explain.

In order to eradicate this shortcoming, Ellingsen and Söderström (ibid.) introduced a model based on the presumption that a change in monetary policy can be traced back to two intentions. Either, new information about the state of the economy has been discovered and thus monetary policy authorities respond to it or their objective functions, i.e. their preferences, changed. The former case is referred to as *endogenous* response to new information while the latter is characterised as *exogenous* shift in preferences. While short rates co-move closely with the target rate in both cases, the model of Ellingsen and Söderström (ibid.) predicts the long end of the yield curve to move in the same direction as the target rate whenever monetary authorities endogenously respond to new information about the state of the economy while long rates move in the opposite direction of the target rate, i.e. a tilt in the yield curve when a policy action can be traced back to an exogenous shift in preferences of the respective monetary authority. The yield curve response happens through actions of bond market participants who update their expectations about future interest rate targets upon observing and interpreting the policy action and thus, price the information into the yield curve.

Implementing this distinction into a model resolves a mismatch between macroeconomic intuition and empirical observations. The former would imply that short and long rates are linked together through arbitrage consider-

ations as well as a change in inflation expectations by market participants. Hence, long interest rates should fall when the target rate is increased as Ellingsen, Söderström and Masseng (2003) report, mainly due to a change in inflation expectations. Empirical research shows, however, that the yield curve sometimes tilts while in most cases long and short maturity rates move in the same direction. As asset prices change once a target rate adjustment is announced, their model succeeds in disentangling how bond market participants change their perception about the state of the economy and the central bank's objectives as a response to the monetary authority's actions. Further evidence for this theory is Gürkaynak, Sack and Swanson (2004) who find that monetary policy actions as well as their accompanying statements influence asset prices.

On top of setting up a model to describe the influence of monetary policy on the term structure of interest rates, Ellingsen and Söderström (2001) perform empirical tests in order to support their theory. While some authors such as Peersman (2002) and Evans and Marshall (1998) run some variations of Vector Autoregression analysis to determine changes in policy preferences, Ellingsen and Söderström (2001) utilize the interpretation of bond traders and analysts as described in a column of the Wall Street Journal (WSJ) surrounding the day of a target rate adjustment by the Federal Reserve Bank as described in Ellingsen, Söderström and Masseng (2003) and use this classification as input to a straightforward regression analysis.

The authors point out that their classification approach is limited to the extent that the daily frequency of the newspaper column might not reflect the immediate opinion of the bond traders whose move seconds after the decision would be the cleanest measure for the classification; the trader's opinion would furthermore not be biased through interpretations by others at that point. Apart from that, the source of input for the classification is very limited to one newspaper and a few journalists. Even though it can be argued that the resulting continuity and consistency of such an approach supports a well founded classification, it might be biased through limited sample size and the twofold human interaction while interviewing traders

and interpreting the finished article. The latter implies that the journalist could have misunderstood the trader's initial interpretation of the policy change and, on top of that, the published article might not unambiguously convey the trader's interpretation.

Both of the shortcomings listed above can be overcome by applying techniques developed to deal with big data, namely, Natural Language Processing (NLP) and Machine Learning (ML), assuming that opinions of traders and journalists converge to the prevalent belief with amount of information. Firstly, the limited sample that can be analysed by hand can be extended by training an algorithm. Secondly, setting exact rules ex-ante ensures a consistent interpretation of articles and thus, removes one human interaction that is exposed to misinterpretation. Furthermore, outliers in the articles have a smaller impact on the final classification due to the increased sample size and inclusion of different sources. Finally, even though the articles analysed still not necessarily reflect the bond trader's immediate opinion, it can be argued that the prevailing interpretation of a policy event by bond traders will win through in the mass of articles covering the event and thus, serve as a good approximation of the immediate interpretation.

Even though methods built upon these have been applied in economics and finance, the full extent of possibilities has not yet reached these fields of academic research. Most applications only consider specific documents such as the employment report (Hautsch and Hess 2002; Hess 2004) while others merely look at the existence of such reports and effects around their release (Bomfim 2003; Hautsch, Hess and Veredas 2011; Lucca and Moench 2015). Tetlock (2007), like Ellingsen and Söderström (2001), analyses daily content from only one WSJ column where sometimes even different topics are discussed and the wanted information is not present. On the other hand, Manela and Moreira (2017) look at newspaper articles over time but only consider those from the front page in order to make the sample size feasible for analysis. This way, the selection can be interpreted to cover the most important news as decided by the publishing agency but since they have completely different goals than a researcher looking at impact of news,

this selection procedure is arbitrary and important information might be lost. Conversely, analysing all articles is computationally not feasible and furthermore will include too much noise through statements irrelevant for the respective topic at hand.

This thesis aims to replicate the results of the theoretical model of Ellingsen and Söderström (2001) using NLP and ML techniques. Since the availability of newspaper articles was very limited during their sample period between 1988 and 2003, I analysed newspaper articles between 2001 and 2017 for which data quality was high throughout the sample. Alas, the latter period is heavily affected by the recession following the financial crisis of 2007 as well as quantitative easing (QE) efforts undertaken by monetary authorities. Hence, rather than finding evidence for the economic model about the connection of monetary policy and market interest rates, this thesis evaluates the model's validity during unprecedented economic conditions and central bank interventions. On top of that, it establishes state of the art methods to deal with today's increased amount of data and introduces a strategy to identify and classify relevant newspaper articles to perform text analytics on a sample of feasible size.

The remainder of the paper is organized as follows. Section 2 provides an overview over monetary policy and its goal during the past decades as well as introducing methodology based on text analysis. The utilized data is discussed in section 3 while section 4 goes into detail about the empirical strategy applied to answer the research question; section 5 presents the results of the empirical analysis. A critical evaluation of the chosen procedure can be found in section 6; section 7 concludes.

2. Monetary policy and text mining background

Rather than suggesting alternative econometric methods to find empirical evidence to the model of market interest rates and monetary policy introduced in Ellingsen and Söderström (ibid.), this thesis attempts to evaluate

their classification strategy as stated in Ellingsen, Söderström and Masseng (2003) by removing human influence and automating the task to the largest extent possible to the best of the author's knowledge. For that reason, the methodology suggested in section 4 bases on techniques used in the area of text analytics and applies it to newspaper articles about monetary policy. The basic concepts of both areas are laid out in the following.

2.1. Monetary policy

The primary objectives of monetary policy as well as actions in order to achieve their goals have varied over time, development status and political system of an economy. Mishkin (2007) makes a strong case about how the past decades after the Great Depression and their experiences have contributed to the current monetary system in the developed world, in particular in the U.S.A. as described below.

2.1.1. The beginnings of inflation targeting

After the events of the Great Depression, the prevalent approach among monetary authorities was Keynesian and later influenced by Samuelson and Solow (1960). In their interpretation of the Phillip's curve, a trade-off between unemployment and inflation has to be resolved in the long-run. Hence, monetary and fiscal policy had the objective to achieve full employment at the cost of a slight rise in inflation. Alas, inflation exceeded the ten percent mark and employment even decreased compared to its level in the 1950s.

Milton Friedman and his stream of monetarists argued, however, that there was no long-run trade-off but rather a natural rate of unemployment would be the equilibrium, irrespective of inflation. For this reason, monetary policy should target inflation instead of output, the determinant for employment, by ensuring a steady growth in the money supply. This argument was later confirmed by Robert Lucas' rational expectations theory. With the oil price shock of 1973, awareness of the importance of a nominal anchor rose as the high costs that inflation accommodates became more apparent. The mech-

anism of such an anchor would facilitate low and stable inflation expectations that lead to stable price and wage setting behaviour of firms, decreasing level and volatility of inflation.

With the realisation that expansive monetary policy does not lead to higher output in the long run, inflation is costly and a nominal anchor is beneficial, many industrialised nations adapted monetary targeting in the mid-1970s. Keeping inflation under control using this strategy hinges upon one critical assumption; there has to be a strong relationship between the goal variable, i.e. inflation or nominal income, and the target aggregate. During the 1980s, it became apparent that this was not valid any more and should be abandoned. Observing how successful German and Swiss central banking had become through the adoption of very transparent policy moves using target ranges, Mishkin (2007) writes, a numerical and clearly communicated long-run goal would support in creating less volatile inflation expectations and still left the central bank enough leeway to deal with short-run fluctuations.

Nevertheless, the prevalent monetary targeting strategy faced difficulties due to its weak relationship between money supply and nominal income, making it impossible to reach the desired inflation outcome. Furthermore, monetary aggregates were no longer a useful signal about the attitude of monetary authorities. As a result, monetary targeting could not properly be used as a nominal anchor and support steering inflation expectations in the vast majority of cases.

In order to make use of advantages of monetary targeting compared with the German and Swiss communication strategy as well as providing a strong nominal anchor, inflation became the new target during the 1990s in many developed economies. Research by Barro and Gordon (1983), Calvo (1978) and Kydland and Prescott (1977) showed that a strong nominal anchor such as inflation could even solve the time-inconsistency problem. The defined long-run commitment to price stability, expressed through a numerical value, holds central banks accountable and makes their performance easy to assess and thus less vulnerable to influence for politicians who are incentivised to

use central bank tools for short-run expansive policy. Output, and thus employment, are still apparent in the monetary authorities' objective function as Svensson (1997) showed. It does, however, consider the former's long-run perspective rather than cyclical behaviour as in the 1960s and 1970s and is thus referred to as *flexible inflation targeting* by Mishkin (2007).

Naturally, it is debatable whether central banks should adapt their strategy subject to current developments. The real estate bubble in the U.S.A. stirred questions about how to react to asset prices, including exchange rates, while the following period of interest rates at the Zero Lower Bound (ZLB) challenged central banks worldwide as their traditional instruments such as target rates could no longer be applied. Thus, central banks such as the Fed and the ECB starting purchasing government bonds in order to raise inflation while smaller economies such as Sweden and Switzerland introduced slightly negative interest rates and the Czech Central Bank pegged the Czech Crown to the Euro. Furthermore, as Mishkin (ibid.) points out, the prevalent opinion diverges about what extend of central bank transparency is still beneficial to an economy. He argues that the relative weights of inflation and output in the goal function should stay occult as the public would not be able to identify when a central bank reacts to economic events and when it changes weights. As shown by Ellingsen and Söderström (2001), reactions on the bond market indicated, however, that markets make inferences from Fed statements which rationale is underlying a target rate change.

2.1.2. Recent history of Fed actions and target rates

Once the Fed started targeting the interest rate in 1988, transparency of policy moves increased compared to the period of money stock measures as pointed out in Ellingsen, Söderström and Masseng (2003). The aftermath of the financial crisis erupting after the collapse of the real estate bubble in the U.S.A. in 2007 had the most influential central banks cut interest rates to their natural lower bound, hindering the traditional monetary transition mechanism, and introduce alternative measures to conduct monetary policy.

Table 2 on page 17 lists the Federal Open Market Committee's (FOMC) target federal funds rate or range, change (basis points) and level as published on Federal Reserve System (2017). While the target rate reached their peak in June 2006, two and a half years later, after the Lehman Brothers collapse, interest rates reached their all-time low and a new era of Zero Interest Rate Policy (ZIRP) commenced.

As Fawley, Neely et al. (2013) state, the Fed executed measures to stimulate economic growth at a time of short rates approaching zero accompanied by the ECB, BOJ and BOE. Since the importance of banks and the bond market varied across their respective economies, the Fed and BOJ focussed their efforts on bond purchases while the other two lent to banks directly. Table 1 on the next page summarises the Fed's engagements to counter the repercussions of the financial crisis while their target rate remained at the ZLB.

Over the course of time, the Fed executed four major large-scale asset purchase (LSAP) programs, usually referred to as QE1, QE2, Operation Twist (Maturity Extension Program and Reinvestment Policy) and QE3. As Fawley, Neely et al. (ibid.) point out, the Fed kept its balance sheet size by reinvesting maturing assets into treasuries at first, and later MBS and GSE debt into MBS. All four programs combined led to a threefold increase of the monetary base compared to pre-crisis levels while, due to the comprehensive augmentation of excess reserves through banks, broader aggregates increased to a more reasonable extent.

As stated in Blinder et al. (2010), the first Fed intervention, QE1, was meant to change the composition of the Fed's portfolio in order to increase liquidity at the capital markets, in particular housing credit markets, by divesting of Treasuries and purchase of less liquid assets such as GSE debt and MBS. By contrast, QE2 aimed at raising inflation and long-term real interest rates and was conducted through the liabilities side of the Fed's balance sheet as stated in Blinder et al. (ibid.) who emphasizes the Treasury's borrowing activities, indicating a combination of monetary and fiscal policy to achieve the goal of price stability. Further increasing it's balance

Table 1: Federal Reserve open market operations during and after the financial crisis based on Fawley, Neely et al. (2013, p. 61).

Date	Program	Description	Target rate
25.11.2008	QE1 (12/2008-03/2010)	Fed announces to purchase $100bn in government-sponsored enterprise (GSE) debt and $500bn in mortgage-backed securities (MBS).	1.00%
18.03.2009	QE1 (12/2008-03/2010)	Fed announces to purchase $300bn in long-term Treasuries and additional $750bn in MBS and $100bn in GSE.	0.00-0.25%
03.11.2010	QE2 (11/2010-06/2011)	Fed announces to purchase $600bn in Treasuries	0.00-0.25%
21.09.2011	Operation Twist (09/2011-12/2012)	Fed announces to purchase $400bn of Treasuries with remaining maturities of 6 to 30 years and divest $400bn with maturities at most 3 years.	0.00-0.25%
20.06.2012	Operation Twist (09/2011-12/2012)	Program to purchase and divest ~$45bn/month extended to the end of 2012.	0.00-0.25%
13.09.2012	QE3 (09/2012-10/2014)	Fed announces to purchase $40bn of MBS per month until the outlook for the labour market improves in the context of price stability.	0.00-0.25%
16.12.2015	End of ZIRP (12/2008-12-2015)	Economic activity has been expanding at a moderate pace, leading to the Fed opting for an increase in the Federal Funds Rate.	0.25-0.5%

sheet through lending operations, total Fed reserves more than doubled from $907bn in September 2008 to $2.214bn in November 2008 (Blinder et al. 2010, p. 468). Blinder et al. (ibid.) also interprets these quantitative easing efforts as effective, at least in parts, as short-term as well as long-term interest rate spreads smoothed gradually. While the Maturity Extension Program and Reinvestment Policy did not expand the monetary base but rather twisted the yield curve, QE3 was introduced while the Operation Twist was still running in order to improve labour market conditions.

2.2. Text mining

Empirical methods across disciplines have utilized quantitative, structured data for decades in order to find evidence for theoretical models, assess the effect of a change in policy and many others. While econometricians have applied knowledge about statistical distributions to empirical observations in order to make inferences about how trustworthy a result might be, the structure of the underlying data was always of major importance which is why many standardised software packages work in terms of two- and three dimensional arrays.

With the rise of computer science, capabilities to analyse as well as availability of data skyrocketed and with it one of the oldest preservable communication means of mankind, written text. Even though grammatically correct sentences come natural to human beings, in order for them to make sense to a machine, a structure needs to be introduced which is the basis of text analytics. The standard way to do so for the classical applications, according to Meyer, Hornik and Feinerer (2008), is based on term frequencies and distance measures. Even though there are standardized procedures when it comes to text analysis, a large variety of features require special attention to obtain meaningful results. For instance, whether the goal text is taken from a newspaper agency, blog, chat history or product feedback makes a huge difference. Different languages might have been used, terms might be misspelled, acronyms used, spam included and many others. Meyer, Hornik

and Feinerer (ibid.) point out, however, that most software packages include the following five features and are thus applicable as basic framework.

Preprocessing (1) encompasses importing the data and cleaning it in such a way that it can be structured without noise through different use in grammar and language, generally. Association analysis (2) refers to counting co-occurrence of terms while clustering (3) assigns similar documents to groups. A general summary (4) returns the major concepts within a text while categorisation (5) classifies texts into predefined categories.

In most cases of applications, after the data has been imported and cleaned, the structuring takes place in the form of a term-document matrix[1] (tdm). The latter can easily be created from a corpus, a collection of text documents, and takes the form of a matrix where the rows indicate documents and columns terms; the matrix elements are consequently the frequency of appearance for each term in every document. Once the texts are in this format, it is possible to write functions to deterministically identify documents and train models.

While clustering documents into different buckets according to specified or unspecified criteria, usually referred to as topic modelling, is one major text mining research area, this discussion does not contribute to the understanding of the methodology applied in this discussion and is thus excluded from the paper. The rationale for not applying topic modelling in this case is that most algorithms in this area work on an unsupervised level and hence do not entail a predetermined deterministic causality which I try to achieve. For this reason, I chose to make a clear distinction between clustering and classification. A general overview and code examples for the software package *R* about the former topic, however, can be found in Silge and Robinson (2017, ch. 6).

[1] In some cases it might make more sense to use a document-term matrix which is the inverse of the tdm.

2.2.1. Machine learning

As indicated in chapter 1, ML techniques for data mining in economics do not receive the same attention as in other fields when it comes to testing theories. One major reason for this is probably a shortcoming in establishing causality as economic models do. Athey and Imbens (2017) present a recent overview paper on applied econometrics that cover precisely this issue and point out why the underlying data structure can solve this problem. Since the mere magnitude of data opens up possibilities to keep huge training and test sets, it is possible to identify causal elements in contrast to classical regression models where a high level of identification is necessary to achieve significant and meaningful results. Thus, some ML techniques may offer an attractive way to serve as additional robustness check for the validity of economic models as well as empirical results from conventional econometrics. In this, ML differs from classical statistics as pointed out in Breiman et al. (2001). While the latter assumes that the underlying data is generated by a stochastic model, ML utilizes algorithmic approaches and treats underlying mechanisms as unknown. Hence, research in ML often focusses on predictive performance and empirical results rather than extensive formal deductive proofs, neglecting distributional properties and derivation of tests basing on asymptotic assumptions. As a result, ML develops at a high speed and most publications are found in conferences and their accompanying proceedings while statisticians mainly publish in journals.

Like most textbooks, Friedman, Hastie and Tibshirani (2001) split data mining techniques into supervised and unsupervised learning. The former is related to classical statistical literature in the sense that we have predictor/independent variables that influence certain response/dependent variables. Solely the way how this influence works differs among various methods. Formally, if random variables (X, Y) are represented by some joint probability density $Pr(X, Y)$, supervised learning tries to determine the properties

of the conditional density $Pr(Y|X)$. More specifically, finding

$$\mu(x) = \underset{\theta}{\operatorname{argmin}}\ E_{Y|X} L(Y, \theta)$$

that minimize the expected error at x where $L(y, \hat{y})$ is some loss function. In unsupervised learning, on the other hand, one has a number of observations from a probability vector X with joint density $Pr(X)$ from which the properties of the probability density are to be inferred directly without the help of a supervisor (Friedman, Hastie and Tibshirani 2001, pp. 485-486). In order to make the results of this analysis comparable to those of Ellingsen, Söderström and Masseng (2003) who use clear conditions on their classification theory, this discussion will only utilize supervised learning techniques as commonly used in the field of sentiment analysis according to Wiebe, Bruce and O'Hara (1999).

2.2.2. Sentiment analysis

The computational study of how opinions, sentiments, subjectivity, evaluations, attitudes appraisal, affects, views, emotions, etc. are expressed in text is referred to as opinion mining or sentiment analysis (Liu 2012). Even though this exercise aims at classifying newspaper articles with respect to predefined criteria, the applied methodology is to a large extent derived from the literature on sentiment analysis and thus described below. According to the review of Feldman (2013), sentiment analysis can be on document, sentence or aspect level or of a comparative nature. While it might make sense to analyse the sentiment of different aspects or sentences in a product review, the classification of newspaper articles with respect to endogeneity versus exogeneity is most useful on a document level, simply because most articles are structured such that objective facts are presented and in some cases accompanied by the author's or some interviewed person's subjective interpretation on the underlying causes of an event.

On this level, Feldman (ibid.) points out the feasibility of supervised learn-

ing approaches as it can be assumed that there is a finite set of classes to which each document belongs, for instance endogenous response to new information and exogenous shift in preferences. When provided with training data, commonly used classification algorithms encompass naïve Bayes (NB), k-nearest neighbours (KNN), support vector machines (SVM) and maximum entropy (ME). Given the classification algorithm, new documents can be assigned the respective sentiment. As shown in Pang, Lee and Vaithyanathan (2002), documents merely represented by a collection of words without structure still yield good accuracy.

In general, an opinion lexicon has to be created in order to extract the sentiment of a statement. The former entails a list of words and expressions that are used to express people's subjective feelings and opinions. Apart from manually creating this list or accessing available dictionaries such as WordNet® by Esuli and Sebastiani (2006) and Fellbaum (1998), it is possible to rely on syntactic patterns in large corpora as has been suggested in Ding, Liu and Yu (2008), Hatzivassiloglou and McKeown (1997), Kanayama and Nasukawa (2006), Turney (2002) and Yu and Hatzivassiloglou (2003); an elegant algorithm to find WordNet® synonyms and antonyms is suggested in Kamps et al. (2004). Even though Feldman (2013) claims that sentiment lexicons are crucial for most sentiment analysis algorithms, it makes sense to use them with care in cases where sentiment does not refer to terms with clear and universe meaning, such as indicators for positive and negative features of a product or service. While the latter can usually be unambiguously derived from reviews, the endogeneity/exogeneity task builds on terms in a specific setting that might be used in different contexts where its sentiments diverge which is why the manual approach should be preferred to standardised lexica.

Other approaches to solve the sentiment classification task in a different context include preselecting articles by sorting out objective statements that do not contribute to the classification (Wiebe, Bruce and O'Hara 1999) or using an ontology tree based on aspects in every document. Wang, Lu and Zhai (2010) perform such an aspect-based analysis and assume an underlying linear combination of the problem as Bayesian regression to determine an

overall classification. Apart from the direct rating through so called regular opinions as laid out above, Ding, Liu and Zhang (2009), Ganapathibhotla and Liu (2008) and Jindal and Liu (2006), among others, analyse comparative opinions which are particularly important in product reviews. Finally, unsupervised learning can be applied; for instance, Turney (2002) introduces an unsupervised three-step learning algorithm to mark a review as positive or negative[1] based on the log-likelihood ratio test which Yu and Hatzivassiloglou (2003) extend further. Another common tool in this field is Latent Dirichlet Allocation (Blei, Ng and Jordan 2003) which is particularly helpful in topic modelling and can serve as a basis for sentiment analysis as in Guo et al. (2009) who develop a latent semantic association model.

3. Data

While all analyses in this discussion are performed by the author, if not indicated otherwise, the raw data is collected from a variety of sources. Since the project can be divided into two distinctive parts, the classification of open market operations and yield curve movements, the origin of the data is disclosed below in a similar fashion.

3.1. Financial time series

Federal Funds target rates as well as major announcements delivered by the board are taken from the website of the Federal Reserve (Federal Reserve System 2017) as well as Fawley, Neely et al. (2013) and newspaper articles as collected by the Factiva database. The daily yield curve utilized is provided by Thompson Reuters Datastream. Figure 1 on the following page compares the Federal Funds target rate (range) to the daily constant maturity US treasury 1M middle rate. Vertical lines indicate extraordinary announcements by the Fed board as described in section 2 that had a

[1] Neutral sentiment is ignored in many cases as a clear distinction from the non-neutral sentiments is not possible.

significant impact on financial markets.

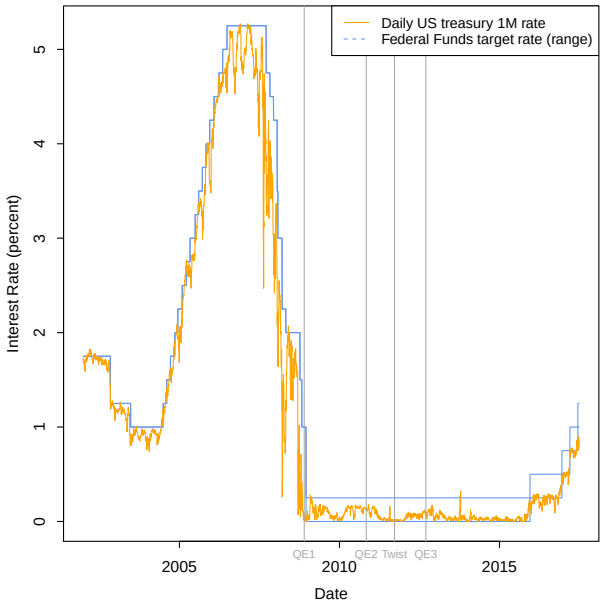


Figure 1: Federal funds target range alongside the one month US treasury rate; *source:* own depiction.

Table 2 on the next page lists all absolute Fed target rate changes during the sample period, one-day relative changes in market interest rates as well as the final classification as described in section 4. The sample ranges from January, 1st 2002 until June, 27th 2017, implying 4,040 changes in the yield curve of which 33 took place during a target rate adjustment of the Fed.

3.2. Articles and text

The final dataset contains 834 newspaper articles of different length in English language from Bloomberg, Financial Times, Reuters, Wall Street Journal, Dow Jones Newswires, AFX, Market News International, the Associated Press and others as collected from the Factiva database from which 9,017 terms are extracted for the analysis. Figures 5 and 6 on page XI in the appendix serve as examples of how the articles look like. For demonstration

Table 2: Federal funds target rate adjustments and corresponding yield curve movements.

Date	Sched	Tgt_{low}	Tgt_{high}	Δ_{low}	Δ_{high}	Δ1M	Δ3M	Δ6M	Δ1Y	Δ2Y	Δ3Y	Δ5Y	Δ7Y	Δ10Y	Δ20Y	Δ30Y	Class
2002-11-06	1	0.0125	0.0125	-0.0500	-0.0500	-0.14	-0.13	-0.11	-0.01	0.03	0.02	0.00	0.00	-0.00	-0.00	0.00	Exog
2003-06-25	1	0.0100	0.0100	-0.0025	-0.0025	0.12	0.11	0.12	0.15	0.14	0.09	0.05	0.04	0.03	0.02	0.02	Exog
2004-06-30	1	0.0125	0.0125	0.0025	0.0025	-0.02	-0.04	-0.04	-0.05	-0.05	-0.04	-0.03	-0.02	-0.02	-0.01	-0.01	R
2004-08-10	1	0.0150	0.0150	0.0025	0.0025	0.01	-0.01	0.01	0.02	0.04	0.03	0.02	0.02	0.01	0.00	0.00	Endog
2004-09-21	1	0.0175	0.0175	0.0025	0.0025	0.01	-0.01	0.01	0.01	0.01	0.01	0.00	-0.01	-0.00	-0.01	0.00	Exog
2004-11-10	1	0.0200	0.0200	0.0025	0.0025	-0.01	0.00	0.00	0.00	0.01	0.01	0.01	0.00	0.01	0.00	0.01	Endog
2004-12-14	1	0.0225	0.0225	0.0025	0.0025	-0.01	-0.01	-0.01	-0.00	0.00	-0.00	-0.00	-0.01	-0.00	-0.00	-0.00	Exog
2005-02-02	1	0.0250	0.0250	0.0025	0.0025	-0.00	0.00	-0.00	0.00	0.01	0.01	0.01	0.00	0.00	-0.00	-0.00	R
2005-03-22	1	0.0275	0.0275	0.0025	0.0025	0.02	0.01	0.01	0.02	0.04	0.03	0.03	0.02	0.02	0.01	0.01	Exog
2005-05-03	1	0.0300	0.0300	0.0025	0.0025	-0.01	-0.01	0.00	0.00	0.01	0.01	0.01	0.00	0.00	-0.00	-0.01	Endog
2005-06-30	1	0.0325	0.0325	0.0025	0.0025	0.01	0.00	0.00	0.00	0.00	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01	Endog
2005-08-09	1	0.0350	0.0350	0.0025	0.0025	-0.00	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01	-0.00	-0.00	-0.00	Endog
2005-09-20	1	0.0375	0.0375	0.0025	0.0025	0.02	0.01	0.01	0.01	0.02	0.02	0.01	0.01	0.00	-0.00	-0.01	Endog
2005-11-01	1	0.0400	0.0400	0.0025	0.0025	0.00	-0.01	-0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	Endog
2005-12-13	1	0.0425	0.0425	0.0025	0.0025	0.01	0.00	0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	Exog
2006-01-31	1	0.0450	0.0450	0.0025	0.0025	0.05	-0.00	-0.01	-0.00	0.00	0.00	0.00	0.00	-0.00	-0.01	0.00	Endog
2006-03-28	1	0.0475	0.0475	0.0025	0.0025	0.01	0.00	0.01	0.01	0.02	0.02	0.02	0.02	0.02	0.01	0.01	Exog
2006-05-10	1	0.0500	0.0500	0.0025	0.0025	-0.01	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	-0.00	-0.00	Exog
2006-06-29	1	0.0525	0.0525	0.0025	0.0025	-0.02	0.00	-0.01	-0.01	-0.02	-0.01	-0.01	-0.01	-0.01	-0.01	-0.00	Exog
2007-09-18	1	0.0475	0.0475	-0.0050	-0.0050	0.01	-0.03	-0.04	-0.04	-0.04	-0.02	-0.02	-0.00	0.00	0.01	0.01	Endog
2007-10-31	1	0.0450	0.0450	-0.0025	-0.0025	-0.00	-0.01	0.00	0.02	0.03	0.02	0.02	0.02	0.02	0.01	0.01	Endog
2007-12-11	1	0.0425	0.0425	-0.0025	-0.0025	-0.05	-0.04	-0.04	-0.05	-0.07	-0.06	-0.06	-0.06	-0.04	-0.03	-0.03	Endog
2008-01-22	0	0.0350	0.0350	-0.0075	-0.0075	-0.18	-0.18	-0.16	-0.15	-0.12	-0.11	-0.08	-0.06	-0.04	-0.02	-0.01	Endog
2008-01-30	1	0.0300	0.0300	-0.0050	-0.0050	-0.09	-0.03	-0.03	-0.01	0.00	0.05	0.03	0.03	0.02	0.02	0.02	Endog
2008-03-18	1	0.0225	0.0225	-0.0075	-0.0075	-0.39	-0.17	0.01	0.06	0.17	0.14	0.09	0.06	0.04	0.01	0.01	Endog
2008-04-30	1	0.0200	0.0200	-0.0025	-0.0025	-0.09	-0.02	-0.06	-0.05	-0.03	-0.02	-0.03	-0.02	-0.02	-0.02	-0.01	Endog
2008-10-08	0	0.0150	0.0150	-0.0050	-0.0050	-0.41	-0.17	-0.06	0.01	0.12	0.11	0.10	0.08	0.06	0.04	0.02	Endog
2008-10-29	1	0.0100	0.0100	-0.0050	-0.0050	-0.02	-0.19	-0.12	-0.08	-0.04	-0.03	0.01	0.01	0.01	0.02	0.02	R
2008-12-16	1	0.0000	0.0025	-0.0100	-0.0075	NA	0.33	-0.18	-0.10	-0.13	-0.14	-0.11	-0.08	-0.06	-0.04	-0.04	Endog
2015-12-16	1	0.0025	0.0050	0.0025	0.0025	-0.05	0.08	-0.06	0.01	0.04	0.05	0.02	0.02	0.01	0.01	0.01	Exog
2016-12-14	1	0.0050	0.0075	0.0025	0.0025	0.04	0.02	0.00	0.05	0.09	0.08	0.05	0.04	0.02	0.00	0.00	Endog
2017-03-15	1	0.0075	0.0100	0.0025	0.0025	-0.08	-0.06	-0.04	-0.04	-0.05	-0.05	-0.05	-0.05	-0.03	-0.02	-0.02	Endog
2017-06-14	1	0.0100	0.0125	0.0025	0.0025	0.01	0.01	0.00	-0.02	-0.02	-0.02	-0.03	-0.03	-0.03	-0.03	-0.03	Endog

purposes, comparatively short snippets have been chosen here. In order to read them into the software, all articles have been converted to simple text files for which all descriptive content in the beginning has been discarded. The sample spans across 33 Fed target rate adjustments for which between 11 and 52 articles of varying length have been analysed, respectively.

4. Empirical strategy

Liu (2012) summarizes the most influential papers in the area of sentiment analysis or opinion mining, a procedure to determine whether a snippet of text, be it in the form of a commentary on a product, post on social media or article in a newspaper, communicates a positive, negative or neutral message about the topic at stake using natural language processing (NLP). Application possibilities of these tools are ample, ranging from businesses improving their products through online reviews to automated fraud and insider trading detection through monitoring analysts' messaging behaviour and are thus of major importance for businesses, individuals and policy makers alike. Building on Feldman (2013), Friedman, Hastie and Tibshirani (2001), Liu (2012) and Silge and Robinson (2017), I derive a procedure to extract an opinion on published articles by major international newspapers and agencies in order to determine the prevalent public opinion on the nature of a target rate move by the Fed.

4.1. Terminology

As texts are a typical example of unstructured data, it needs to be converted to structured data in order to perform meaningful analyses. Therefore, I

follow Liu (2010) in defining an opinion as the quintuple[1]

$$(e_j, a_{jk}, so_{ijkl}, h_i, t_l) \tag{1}$$

where e_j is the target entity which forms the opinion target together with a_{jk}, an aspect of the former; in a product review the target entity might be a laptop while battery life would be one of its aspects. Naturally, a sentiment analysis would take both parts into account and hence their combination is referred to as opinion target. so_{ijkl} refers to the sentiment value of the opinion source, h_i, on a_{jk} of e_j at time t_l. Since the first step of the analysis, the identification of relevant articles as described in section 4.2 determines the opinion target while time and source are given exogenously, I focus on discovering the sentiment of each text snippet.

4.2. Article identification

Articles have been accessed from the Factiva database which offers a user-friendly interface through which content can be screened with respect to different criteria. In order to obtain a sample of useful text snippets for the sentiment determination, articles plus/minus one day around each FOMC meeting have been taken into account; of those, I considered the ones in English language dealing with interest rates and central bank interventions in the U.S.A. that covered the terms *Fed* or *Federal Reserve* as well as *Interest Rate*. Since some agency reports appeared repeatedly, I manually selected a few across most conventional agencies and newspaper publishers, including statements of economists. This procedure ensures that the absolute majority of text sources is highly relevant for the classification exercise which is important for the further analysis. Hence, the quintuple above is defined

[1]Even though the literature on opinion mining takes subjectivity and emotion into account (Riloff, Patwardhan and Wiebe 2006; Wiebe 2000; Wiebe et al. 2004), I refrain from doing so as the source of the data comes from professional and reviewed media sources exclusively and the aim of the analysis is not to determine how individuals feel about a certain product or situation but rather extract the summary of public opinion from a newspaper article statement.

for every open market operation of interest apart from so_{ijkl}. In order to achieve this, the text data have to be cleaned and organised by imposing a quantitative structure as described below.

4.3. Pre-processing

After reading in the data by collecting all text snippets allocated to one FOMC meeting, a few manipulations have to be undertaken in order to prepare the text such that it is feasible for quantitative analysis. The rationale for this is simply that a large part of the text does not contain useful information and would hence distort the results by adding too much noise and taking up computation power. This procedure is commonly referred to as *pre-processing* and entails the removal of stop words and stemming as well as manipulations depending on the respective task.

Apart from a few technical manipulations to make the input into R easier, the pre-processing in this study begins by discarding all articles in a language other than English. The removal of stop words, i.e. terms that inherit no intrinsic meaning, is performed in two consecutive steps[1]. First, an algorithm is applied to tokenize every word in each text. These Part-of-Speech (POS) tags as presented in table 7 on page XII in the appendix are applied to the articles through a pre-trained model by Hornik (2016) that assigns POS tags based on the probability of what the correct POS tag is for newspaper language and selects the one with highest probability; an example of a tagged article snippet can be found in figure 7 on page XII in the appendix. Once every word is tagged, those identified to have only subordinate or auxiliary purpose are discarded. Secondly, all remaining elements of the text are scanned for punctuation, numbers, unnecessary white space and de-capitalised while an extended built-in function filled with common stop words subtracts the remaining ones. Special care has been taken with

[1]Another approach is based on term-weighting as described in Silge and Robinson (2017) accompanied by their published R package Silge and Robinson (2016). Since the articles are pre-selected in an earlier step, this procedure is deemed unnecessary, however.

valence shifters, presuppositional items and modal auxiliary verbs that have been combined to unigrams, i.e. terms without spaces. The pre-processing is concluded by stemming the terms left such that words from the same family in different conjugations are detected as equivalent[1]. Once this manipulation is fulfilled, a corpus is formed from all pre-processed text files allocated to a target rate change.

4.4. Sentiment determination

In order to identify the sentiment of each text piece as endogenous or exogenous, I apply a deterministic, count-based (CB) approach as well as state-of-the-art ML algorithms. While the latter have shown stable performance in text mining studies in different contexts, this exercise demands high precision and thus I chose not to rely on learning algorithms alone but created a deterministic algorithm that classifies target rate changes based on predefined terms as well as significance tests. Rather than combining all texts around one target rate adjustment and determining its sentiment, I chose to conduct the classification on an article level and based on the number of endogenous and exogenous results determine the overall sentiment for two reasons. First, this procedure is more transparent as it shows how many articles of the one kind oppose the other. Secondly, since article length varies a lot, it is easier to control for outliers as every article counts equally towards the final classification.

4.4.1. Count-based evaluation

One classical way to analyse texts through predefined methods is assuming that terms with higher occurrence frequencies are more important than others. Because of the simplicity of this approach, it is widely used throughout

[1]When it comes to customer reviews, sarcasm and opinion spam are two more important aspects to look for; since this exercise mainly contains facts about monetary policy and interpretation by market participants, I deem this problem to be of subordinate nature. In particular, since the vast magnitude of words per article relativises this issue.

the field as pointed out in Meyer, Hornik and Feinerer (2008). In order to design a promising approach, a list of words and expressions has to be set up which can be done in three different ways as discussed in section 2. The most straight forward one, which has been chosen here for its comprehensible and deterministic nature, being manually setting it up in a one-time effort. The sentiment defining words are chosen such that they are representative for an endogenous or exogenous event as defined in Ellingsen, Söderström and Masseng (2003) together with a synonym finder. A list of the most relevant terms[1] used, excluding synonyms, can be found in table 8 on page XIII in the appendix. The polarity is reversed whenever a term is preceded by a negation. In contrast to Ellingsen, Söderström and Masseng (ibid.), endogenous events are not defined residually. In this discussion, it is sufficient to look for simple quantitative occurrence of expressions since, intuitively, crucial events as well as comments accompanying FOMC statements appear across many articles as common in count-based evaluation. Furthermore, the binary nature of the classification problem leads to analysing increments as common in ratings, for instance, being negligible whereas dictionary-based methods usually find the total sentiment of a piece of text by adding up the individual sentiment scores for each word in the text (Silge and Robinson 2017).

As input, my function takes a pre-processed corpus consisting of articles surrounding one FOMC meeting in which a target rate change has been decided and announced as well as a list of terms that are indicative of endogenous or exogenous sentiment, respectively; finally, a confidence level has to be decided ex ante. Naturally, across dates, the input parameters apart from the corpi are equivalent in order to make results comparable and consistent. The function then compares occurrences of the respective terms with predefined sentiment to each document in a corpus and prints the amount

[1]This list is by no means comprehensive and some of the expressions could appear in different contexts or even indicate the respective other sentiment. Nevertheless, given the previous pre-selection of articles and the mere magnitude of available articles, I am confident that the listed terms succeed in extracting the overall tendency across articles around one date.

of words with endogenous and exogenous sentiment, respectively, as well as their difference. Thus, the output of the function is a list of documents from every FOMC date which states by how much the number of endogenous words exceeds or deceeds that of exogenous words.

Even though Meyer, Hornik and Feinerer (2008) describe this procedure as sufficient, varying article length by construction has a huge impact on the absolute number of sentiment word appearances across documents surrounding one FOMC meeting. For that reason, I included a one-sided t-test for comparison of means among the different sentiment words across documents per FOMC meeting. Given that both vectors are created from the same articles ranging from a few hundred to a few thousand words, each, and consist of positive integer values, applying a standard student t-test seems to be justified. Depending on whether the test yields a significant difference in means, the function returns a final classification stemming from the deterministic, count-based function or flags it as *ambiguous*.

Since across the data mining literature as pointed out in section 2, machine learning techniques receive more attention and trust than simple count-based approaches, I selected 7 articles with obvious endogenous and 12 with exogenous sentiment as training set which I extend with the text snippets from Ellingsen, Söderström and Masseng (2003); 36 for the endogenous and 26 for the exogenous training set. This way, both classes have a similar length and can be used to train the most common supervised learning techniques in sentiment analysis according to Liu (2010) and Feldman (2013). Generally, train and test sets have to contain records that are representative of the entire dataset in order to yield internally valid parameter estimates which is why I took all articles for the training set from the initial 834 ones and selected none to three across all dates on top of the snippets from the Ellingsen, Söderström and Masseng (2003) sample from the 80s and 90s.

4.4.2. Naïve Bayes

Friedman, Hastie and Tibshirani (2001, p. 211) list the Naïve Bayes (NB) classifier among the linear methods for classification as a variant of linear discriminant analysis. As such, it adapts the loss function to the fact that the output variable G is categorical such that prediction errors are penalised appropriately. Utilising a zero-one loss function in which every misclassification is charged one single unit, Friedman, Hastie and Tibshirani (ibid., pp. 20-21) simplify the expected prediction error $EPE = E[L(G, \hat{G}(X))]$ such that the

$$\hat{G} = \max_{g \in \mathcal{G}} Pr(g|X = x) \tag{2}$$

conveys the intuition that classification is done to the most probable case in which the conditional discrete distribution $Pr(G, X)$ is used. NB enhances this approach by assuming that the inputs are conditionally independent in each class, i.e. that every class density is a product of marginal densities. Even though this assumption is generally not fulfilled as certain word combinations appear consistently, Friedman, Hastie and Tibshirani (ibid.) as well as Rish, Hellerstein and Thathachar (2001) emphasise that it outperforms more sophisticated alternatives in many cases. One reason for this is that the prediction depends only on the maximum probability, not its actual value. Furthermore, dependencies cancel out in many cases when working with a large set of features.

4.4.3. Maximum entropy

Maximum entropy (ME), on the other hand, as described in Berger, Della Pietra and Della Pietra (1996) does not impose the restrictive conditional independence assumption and can be applied when underlying distributions are not known ex ante. This is achieved in the machine learning context through a training set that produces output values y from a finite set $\mathcal{Y}$

while the inputs x are from $\mathcal{X}$. Their empirical probability distribution is

$$\tilde{p}(x, y) = \frac{1}{N} \times number\, of\, times\, that\, (x, y)\, occurs\, in\, the\, sample$$

where N is the size of the training set. Furthermore, introducing an indicator function, often referred to as *feature*,

$$f_j(x, y) = \begin{cases} 1 & \text{if } y = c_i \text{ and } x \text{ contains } w_k, \\ 0 & \text{otherwise} \end{cases}$$

in which c_i denotes a member of the different classes $\mathcal{C}$ possible while w_k is a word. That means the indicator function returns the value one if a document belongs to class c_i and contains the word w_k. Based on their features, Berger, Della Pietra and Della Pietra (1996) show that the model p^* should be selected to be as close as possible to uniform according to the ME principle

$$p^* = \underset{p \in \mathcal{C}}{arg\,max} \Big(- \sum_{x,y} \tilde{p}(x) p(y|x) log\, p(y|x) \Big) \tag{3}$$

which can be solved through a Lagrangian approach where the multipliers can be estimated through an iterative scaling algorithm.

4.4.4. Knn

As stated in Friedman, Hastie and Tibshirani (2001, p. 465), k-nearest-neighbours is best applied in settings where every class has a lot of different prototypes and the decision boundary is irregular. Above that, this family of classifiers does not need a model to fit as it is memory-based. It works with a query point x_0 for which the k training points $x_{(r)}$, $r = 1, \ldots, k$ are found which are the closest to x_0, according to some distance metric. In the article classification task, for every row of the test set corpus, the k closest training set vectors, as determined through Euclidean distance, are found and the classification is achieved through majority vote where ties are broken at

random. Should there be ties for the kth nearest vector, all candidates are included in the vote as stated in the R package documentation.

4.4.5. Support vector machines

Finally, support vector machines (svm) perform well, according to Williams (2011, p. 293), on assignments that are non-linear, sparse and high-dimensional. The underlying rationale is to transform the data such that the classes in the training set become linearly separable by a hyperplane. The classification is then conducted by transforming new data in the same way as in the training set and determine on which side of the hyperplane the points of interest lie. A mathematical representation of this concept is not included in this discussion since too many technicalities would have to be introduced; the basics are explained, however, in Friedman, Hastie and Tibshirani (2001, p. 417).

4.4.6. Classification of target rate adjustment dates

All ML algorithms utilised in order to determine the sentiment distinction are performed through the standard functions in the R software package; their classifications are listed in table 3 on the following page. Due to the low dimensionality of terms in the dtm, linear kernels have been chosen for ME and SVM. Table 3 on the next page lists the number of articles that have been classified as endogenous and exogenous per target rate adjustment date and decision algorithm as well as overall date classification by quantitative comparison of classified articles. The count-based approach yields an overall number of 13 endogenous and five exogenous events while 16 cannot be determined statistically significant and are thus marked as ambiguous. Using an NB classifier to predict articles from the action set shows that the class distribution for endogenous/exogenous training articles is 39/37[1]. As out-

[1]The difference to the initial training set stems from the fact that the algorithm omits table entries for attributes with missing values. Since some of the short snippets from the training set can thus not be used, there are four endogenous and one exogenous text snippet missing from the initial 81.

Table 3: Target adjustment dates with corresponding number of as endogenous and exogenous classified articles through different text mining procedures.

Date	CB	CB_{end}	CB_{ex}	NB	NB_{end}	NB_{ex}	ME	ME_{end}	ME_{ex}	KNN	KNN_{end}	KNN_{ex}	SVM	SVM_{end}	SVM_{ex}
2001-12-11	Ambig	6	8	Endog	21	0	Exog	10	11	Exog	5	16	Endog	15	6
2002-11-06	Ambig	18	19	Endog	52	0	Exog	15	37	Exog	13	39	Endog	42	10
2003-06-25	Ambig	16	16	Endog	39	0	Exog	13	26	Exog	11	28	Endog	27	12
2004-06-30	Endog	18	11	Endog	35	1	Endog	25	11	Exog	10	26	Exog	17	19
2004-08-10	Endog	17	9	Endog	29	0	Endog	16	13	Exog	11	18	Exog	14	15
2004-09-21	Exog	7	14	Endog	27	0	Endog	15	12	Exog	13	14	Endog	14	13
2004-11-10	Ambig	13	10	Endog	24	0	Exog	11	13	Exog	9	15	Endog	16	8
2004-12-14	Ambig	13	13	Endog	28	0	Exog	10	18	Exog	9	19	Endog	18	10
2005-02-02	Ambig	8	10	Endog	23	1	Exog	8	16	Exog	7	17	Endog	22	2
2005-03-22	Exog	9	15	Endog	28	0	Exog	10	18	Exog	9	19	Endog	21	7
2005-05-03	Ambig	9	14	Endog	25	2	Exog	8	19	Exog	9	18	Endog	17	10
2005-06-30	Ambig	7	4	Endog	13	0	Exog	5	8	Exog	4	9	Endog	10	3
2005-08-09	Ambig	11	11	Endog	25	0	Exog	12	13	Exog	8	17	Endog	17	8
2005-09-20	Ambig	13	10	Endog	26	0	Ambig	13	13	Exog	12	14	Endog	16	10
2005-11-01	Endog	14	7	Endog	23	0	Exog	8	15	Exog	5	18	Endog	17	6
2005-12-13	Exog	6	14	Endog	24	0	Exog	6	18	Exog	5	19	Endog	20	4
2006-01-31	Ambig	5	6	Endog	13	0	Exog	4	9	Endog	7	6	Endog	8	5
2006-03-28	Ambig	6	10	Endog	16	1	Exog	3	14	Exog	5	12	Endog	11	6
2006-05-10	Exog	2	13	Endog	17	0	Exog	6	11	Exog	7	10	Endog	11	6
2006-06-29	Exog	6	11	Endog	19	0	Exog	8	11	Exog	6	13	Endog	12	7
2007-09-18	Endog	24	0	Endog	25	0	Exog	6	19	Exog	7	18	Endog	23	2
2007-10-31	Endog	12	2	Endog	15	0	Exog	7	8	Exog	3	12	Endog	9	6
2007-12-11	Endog	9	1	Endog	11	1	Endog	7	5	Exog	2	10	Exog	5	7
2008-01-22	Endog	8	2	Endog	12	0	Exog	3	9	Exog	0	12	Endog	8	4
2008-01-30	Endog	13	2	Endog	16	0	Exog	4	12	Exog	2	14	Endog	14	2
2008-03-18	Endog	10	4	Endog	18	0	Exog	0	18	Exog	3	15	Endog	18	0
2008-04-30	Endog	21	4	Endog	25	0	Exog	4	21	Exog	2	23	Endog	22	3
2008-10-08	Endog	8	2	Endog	11	0	Exog	3	8	Exog	2	9	Endog	8	3
2008-10-29	Endog	12	4	Endog	17	0	Exog	4	13	Exog	3	14	Endog	15	2
2008-12-16	Endog	13	4	Endog	21	0	Exog	6	15	Exog	8	13	Endog	19	2
2015-12-16	Ambig	10	7	Endog	18	1	Endog	11	8	Exog	5	14	Endog	11	8
2016-12-14	Ambig	7	8	Endog	19	0	Exog	8	11	Exog	4	15	Endog	10	9
2017-03-15	Ambig	8	7	Endog	20	1	Endog	11	10	Exog	5	16	Endog	14	7
2017-06-14	Ambig	6	8	Endog	15	0	Exog	7	8	Exog	6	9	Endog	9	6

put, the classifier displays a table for each of the 9,016 predictor terms which states the conditional probability given the target class for each attribute level. In a consecutive step, predictions are made based on these conditional probabilities through the terms in every article to classify which yields only none to two exogenous articles and thus, NB classifies all events as endogenous. The ME classifier, on the other hand, directly predicts the sentiment along with its corresponding probability based on the probability distribution that best fits the data given their underlying constraints. Comparing the absolute number of classified articles and simply selecting the one with the higher amount yields six endogenous, 27 exogenous and one ambiguous target rate adjustment. Similarly, knn directly outputs the classification of each article based on the most likely affiliation due to terms in its immediate proximity. Taking a simple majority vote, knn classifies one date as endogenous and 33 as exogenous; compared to ME, there are less close calls. SVM, finally, yields direct classifications with corresponding probabilities based on a logistic distribution using maximum likelihood. In contrast to ME which estimates the data's empirical probability distribution, SVM probabilities range from .50 to .95 with a .75-quantile of .55 while ME probabilities lie between .51 and 1.00 with median close to one. Even though SVM appears to be more sensitive to the sentiment in the articles, 31 events are classified as endogenous while only three are determined to be exogenous.

It is obvious that the NB model strongly prefers the endogenous classification while knn determines most decisions to be exogenous. ME and SVM appear to be slightly more attentive towards the sentiment in the articles but come to different conclusions as well. Even though the ML learning methods come with a classification certainty[1], it is apparent that the endogeneity/exogeneity assignment is too much of a specialised task as it builds upon very detailed formulations in the texts for which the training sample is too small, in particular for such an extensive action set. As the count-based

[1]I chose not to include the probabilities for every article classification by method for purposes of readability. Furthermore, since results diverge this strong, they do not add significant value or credibility to the respective results.

algorithm includes a significance test across articles for every adjustment date, I base the selection for the final classification on the latter and refer to the most promising of the ML techniques whenever CB returns an ambiguous classification and at least one of the other algorithms shows a clear tendency towards one sentiment; the list of open market operations (OMO) with corresponding classification is outlined in table 2 on page 17. In accordance with Ellingsen, Söderström and Masseng (2003), all dates that inherit another important economic release by the Bureau of Labor Statistics such as the employment report, have been excluded from the sample and marked as R while those changes that have not been noticed by markets are indicated with a U. The information about relevant economic releases was taken from Labor Statistics (2017).

4.4.7. Performance evaluation

In order to get an idea about the accuracy of the described classification procedures, I performed a simulation across all ML algorithms. For this purpose, I took a test set consisting of 39 endogenous and 37 exogenous articles and randomly drew 90% of those from which I predicted the sentiment of the remaining 10%; I calculated the accuracy measure of correctly specified articles over total articles as the fraction of the sum of the diagonal elements of the confusion matrix over the total number of articles to be classified. Repeating this procedure 500 times and taking the average accuracy as well as a few indicators about the distribution yields the values denoted in table 4 on the next page. It is obvious that svm is significantly outperformed in terms of variation and mean by the other three techniques which vary across 60% of correctly specified articles on average while their standard deviation is around 20%. ME seems to perform best with knn being second while having the largest standard deviation among the three top ones. Its predictions have to be scrutinized with particular care since they range from no correct prediction to 100% correct predictions. The performance of NB might be traced back to it almost entirely predicting endogenous sentiment while

Table 4: Simulated accuracy of different ML algorithms.

	Naive Bayes	Maximum entropy	knn	SVM
Mean	0.5134	0.6374	0.5700	0.1983
Median	0.5714	0.7143	0.5714	0.1429
Std	0.1747	0.1789	0.1905	0.2497
Min	0.1429	0.1429	0.0000	0.0000
Max	1.0000	1.0000	1.0000	1.0000

in general, endogenous events seem to appear more often than exogenous ones, in this sample as well as during the 80s and 90s period of Ellingsen, Söderström and Masseng (2003).

Even though the results do indicate that most algorithms are a better choice than randomly allocating sentiment, it has to be emphasized that 90% of the data predict the remaining 10% and in the final task, the action set is significantly larger than the training set. Since the algorithm is trained, however, to detect the same sentiment in this evaluation and does classify many articles correctly, in some compositions even with more than 85% for maximum entropy, I conclude that some of the learning techniques do find the right indicators for both sentiment types. The high variation in the final results as well as some clearly misspecified combinations might be due to outliers in the articles. Whenever the latter become very long and cover a variety of topics, the ML algorithms tend to misclassify them. One possibility to mitigate this shortcoming is to restrict the analysis for sentiment determining terms to those in immediate proximity of key phrases. For this analysis, this does not seem to add much value since most articles are really short once stopwords are discarded and the longer ones are too complex in nature, i.e. the interpreted rationale for a target rate change is not necessarily written immediately after the objective description of the event itself.

When it comes to the count-based approach, a comparable straight forward evaluation procedure is not at hand. Instead, I applied it to some of the

articles used in Ellingsen, Söderström and Masseng (2003). Even though I could only access ten of their articles and most of those were not feasible for testing as they have been classified as unnoticed or coincided with employment report releases, the count-based approach detects the correct sentiment in about 70-80% of the cases as far as obvious sentiment can intuitively be interpreted from the data. For instance, the target rate adjustment on May 1st, 1991 could correctly be specified as exogenous. Even though March 11th, 1991 and May 5th, 1991 coincided with employment report releases and were thus not included in their assessment, looking at the content of the WSJ column[1] reveals a tendency towards endogeneity that has successfully been recognized by my deterministic function. Even though this supports the use of this method, the WSJ column used by the authors covers a variety of topics relevant for financial markets and is thus broader than the coverage in the articles selected for this analysis. Furthermore, the idea of applying automated text mining in this exercise is to apply the function to a variety of newspaper articles in order to detect the general tendency towards the interpretation of Fed motives underlying a target change rather than focussing thoroughly on one source of information.

4.5. Monetary policy and the bond market

As described in section 1, it is well established in empirical research that monetary policy, on average, has a positive effect on market interest rates of all maturities. Looking at past developments more carefully, however, reveals that interest rates with longer maturities in some cases move in the opposite direction of the target rate. Ellingsen and Söderström (2001) capture this observation in a theoretical model developed by Svensson (1997) in which the

[1] 'Among other things, they note the recent growth in the money supply, which now is closely watched by the bond market because Federal Reserve Board Chairman Alan Greenspan has said he views it as an important indicator of economic growth' on March 8th and 'instead, the job outlook is worsening, after-tax disposable income is declining, the effective cost of instalment credit is rising and consumer access to credit is being cut back by creditors. The Fed will be forced to ease significantly further' on March 11th.

central bank minimizes the quadratic deviation of its inflation and output targets. The central bank influences the economy through setting the one period interest rate which affects output and inflation with a lag. Ellingsen and Söderström (2001) extend the model with an equation describing the term structure of interest rates and endogenously derive the central bank's reaction function.

The model predicts that in case the central bank's objective function is observed by all market participants, they fully anticipate possible target rate adjustments and thus market rates respond immediately and move in the same direct as the target rate. If the central bank rate changes unanticipated, the model predicts two different outcomes depending on the underlying rationale of the change. First, market participants might recognize that the central bank has private information and thus, interpret an increase in the target rate as indicator for increasing inflation. Their actions in this scenario lead to interest rates of all maturities increasing as well. Secondly, an increase in the target rate could be interpreted as reaction to a high weight towards price stability in the central bank's objective function. Since the perception of the state of the economy is not changed for the bond traders, they react to the fact that inflation, on average, will be fought more vigorously by the central bank and hence will be lower on average. This leads to interest rates with sufficiently high maturity decreasing while shorter maturity rates increase along with the target rate.

In order to perform an empirical test of the model and formulate hypotheses, days on which the central bank adjusts the target rate (policy days) and days on which it does not act (non-policy days) have to be separated since in the model the target rate is adjusted every period, conversely to the actions of the Fed. Private information of the central bank can only be revealed on policy days since leaving the rate unchanged, in the vast majority of cases, is due to the state of the economy which is observed by all market participants alike as the central bank does not reveal any private information. If the target rate is adjusted due to private information about the state of the economy, the model predicts that market rates move in the

same direction as the target rate since bond traders adjust their inflation expectations and hence the new information is priced into bond rates; the correlation with long rates is lower than for short rates. If the adjustment is due to a change in weights of the central bank's objective function, long rates are even negatively correlated, i.e. the yield curve tilts. The stated model predictions are formulated in four hypotheses by Ellingsen, Söderström and Masseng (2003) and tested empirically by analysing the one-day change of the 3-month treasury bill rate after a policy adjustment date which serves as a proxy for the policy innovation as the 3-month rate is mainly affected by current developments and not as noisy as shorter maturity rates.

4.6. Term structure of interest rates

Using the classification established in section 4.4.6, I follow Ellingsen, Söderström and Masseng (ibid.) in testing the hypotheses about how market interest rates respond to Fed target rate adjustments by estimating two regressions in which the one day change of interest rates of different maturities is explained by the change in the 3-month rate which represents the policy innovation of the Fed, depending on the classification of the OMO. First, Ellingsen, Söderström and Masseng (ibid.) test their theory whether the relationship between long and short rates differs on policy days and non-policy days. For that purpose, their regression model is set up as

$$\Delta i_t^n = \alpha_n + (\beta_n^{NP} d_t^{NP} + \beta_n^P d_t^P)\Delta i_t^{3m} + \nu_t^n \tag{4}$$

where Δi_t describes the one day change in the interest rate on day t and its respective superscript denotes its maturity. Consequently, Δi_t^{3m} is the change in the 3-month rate. d_t are dummy variables for policy (P) and non-policy (NP) days and ν_t^n is the standard idiosyncratic error; α_n denotes the intercept. The regression thus measures the reaction of interest rates of maturities 6 months, 1, 2, 3, 5, 7, 10, 20 and 30 years to the policy innovation of the Fed on non-policy (β_n^{NP}) and policy (β_n^P) days, respectively.

As formulated above, the authors hypothesise whether the information on policy days differs from the one on non-policy days where the effect on long rates should be less clear than for short rates due to the different reasons why a policy innovation took place which has opposite effects on the long end of the yield curve according to the model prediction. This can be formulated in their first hypothesis

Hypothesis 1 *For large n,* $\beta_n^P < \beta_n^{NP}$.

Secondly, Ellingsen, Söderström and Masseng (2003) dive deeper and investigate directly whether the long and short rates behave differently on policy days classified as endogenous or exogenous as well as whether non-policy days have a similar impact as endogenous policy days. Therefore,

$$\Delta i_t^n = \alpha_n + (\beta_n^{NP} d_t^{NP} + \beta_n^{End} d_t^{End} + \beta_n^{Ex} d_t^{Ex}) \Delta i_t^{3m} + \nu_t^n \tag{5}$$

is estimated similar to model 4 where d_t^m, $m \in \{End; Ex\}$ are dummy variables taking the value of one if the corresponding day has been classified as endogenous or exogenous, respectively. From their theoretical model, hypotheses 2 through 4 are formulated such that

Hypothesis 2 *For large n,* $\beta_n^{Ex} < 0 < \beta_n^{End}$,

Hypothesis 3 $\beta_n^{NP} = \beta_n^{End} = 0 \; \forall \; n$,

Hypothesis 4 β_n^j *is decreasing in n for j = NP, End.*

Hypothesis 2 goes a step further than hypothesis 1 and tries to disentangle positive and negative effects after endogenous and exogenous policy innovations and is thus the key to the model formulated by Ellingsen, Söderström and Masseng (ibid.) as it looks at reactions at the long end of the yield curve. Hypothesis 3 is a test about the revelation of private information after an endogenous OMO since if the central bank does not reveal new information to markets by adjusting its target rate, when it does not happen because of a change in preferences, market rates should not behave differently than on

non-policy days as all information about the state of the economy is already priced in. Nevertheless, the amount of influence of the target rate is expected to decrease with maturity since short-term fluctuation become less important for the medium and long run expectations about inflation which is the basis for hypothesis 4.

Since this discussion covers a time period with interest rates at the ZLB and central banks on the verge of inability to use their traditional means of interference, other announcements have been delivered such as the various QE programs described in section 2. As comparable announcements of unconventional monetary policy are not present during the sample period of Ellingsen, Söderström and Masseng (2003) but impact financial markets as pointed out in Hattori, Schrimpf and Sushko (2016), I ran the regressions in equation 4 and 5 with an additional dummy interaction for a QE coefficient as control where the date was chosen to be the official QE announcement date as stated in table 1 on page 9. Even though this should increase the credibility of the model, the sample period is heavily affected by the events of the financial crisis and thus many of the Fed chairman announcements will have triggered market responses as stated in Blinder et al. (2010) while not all of them can be controlled for as they are of ambiguous nature or concur with other statements and report releases and hence, their effects cannot be disentangled clearly in such a limited sample size.

5. Results

As presented in Ellingsen, Söderström and Masseng (2003), short term market rates have a close relationship with the target rate while the yield on long-term bonds behaviour tends to be ambiguous. While, as commonly explained through arbitrage arguments, long rates should be linked to short-term rates, a tilt of the yield curve has been observed in numerous occasions, possibly due to a change in inflation expectations. The model proposed in Ellingsen and Söderström (2001) implements this feature through the per-

ceived underlying reason for a monetary authority's decision to adjust the target rate. This implies that it is either because of new information about the economy (endogenous response to new information) or a change in the preferences towards the trade-off between inflation and unemployment (exogenous shift in preferences).

5.1. Policy versus non-policy days

The first test of Ellingsen, Söderström and Masseng (2003) addresses hypothesis 1 and thus, whether long and short rates have a different relationship on days when the target rate was adjusted by the Fed. I follow them by defining a policy day to be any event when the monetary authority changed the target rate and count all other days as non-policy days; this includes those days on which the FOMC met and did not adjust the target rate. Even though no action can be considered as policy innovation, the FOMC meetings take place regularly and between 2009 and 2015 there was not a single adjustment, as announced, while meetings took place. For the purpose of coherence, hence, all FOMC meetings without target rate adjustment are treated as standard non-policy days.

Using regression (4) on the dataset separated into 4,007 non-policy days and 30 policy days supports the findings of Ellingsen, Söderström and Masseng (ibid.) as shown in table 5 on the next page; the remaining three are treated as non-policy days since major economic reports have been published on the same day the target rate was adjusted. Since all changes in policy were discussed in the press, I refrain from excluding policy days due to them being unnoticed by markets. Since a Breusch-Pagan test on the residuals revealed heteroskedasticity for maturities of up to 10 years, robust standard errors have been reported in these cases. While for the shorter maturities the relationship with the three month rate is even bigger for policy days, maturities of two years and more is considerably weaker and even approaches zero for above ten years; mainly due to insignificant coefficients. This is in line with the model prediction as policy adjustments can take place out

Table 5: Regression results for yield curve response to short rate movements on policy days and non-policy days.

	6m	1y	2y	3y	5y	7y	10y	20y	30y
α_n	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	(0.95)	(0.86)	(0.63)	(0.57)	(0.50)	(0.47)	(0.42)	(0.34)	(0.40)
β_n^{NP}	0.59	0.46	0.36	0.35	0.32	0.28	0.23	0.17	0.16
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
β_n^{P}	0.83	0.64	0.33	0.28	0.2	0.16	0.1	0.03	0.01
	(0.00)	(0.01)	(0.39)	(0.49)	(0.62)	(0.64)	(0.75)	(0.76)	(0.93)
$\bar{R}^2$	0.58	0.36	0.11	0.09	0.06	0.04	0.03	0.02	0.02
$\beta_n^{NP} = \beta_n^{P}$	0.00	0.00	0.69	0.39	0.18	0.22	0.13	0.08	0.07

of two different reasons which have opposite effects on the long end of the yield curve which is why they even out and result in a zero policy coefficient. For non-policy days the coefficient is still significantly positive. Controlling for the QE efforts of the Fed does not change the results as can be seen in table 9 on page XIV in the appendix. This is an indication that policy days reveal information about the monetary authority's preferences as β_n^P decays much faster with n than β_n^{NP} and both coefficients statistically differ and thus inherit different information for higher maturities.

5.2. Endogenous versus exogenous days

In order to find evidence for hypotheses 2 through 4, the sample of policy days is further divided into 20 ones that are interpreted in the press as endogenous and ten which are seen to be exogenous. A first graphical representation of possible difference in their nature can be observed by figures 2-4 which show the change in the 10-year rate against the change in the 3-month rate at classified policy days. Compared to the analysis of Ellingsen, Söderström and Masseng (2003), figure 2 on the next page, which includes all policy events, does not show a clear positive relationship although a slight tendency is noticeable. Since the 3-month rate is used as proxy for the policy innovation, it can thus be concluded that in general, the long rate responds with a change

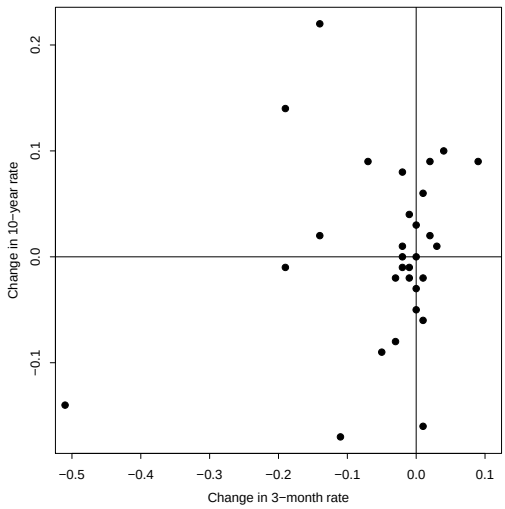


Figure 2: Response of the 10-year interest rate to a change in the 3-month rate for all classified policy events; *source:* own depiction based on Ellingsen, Söderström and Masseng (2003).

in the same direction which is in line with what most authors find on this topic, according to Ellingsen, Söderström and Masseng (2003). For policy days classified as endogenous the picture is clearer, as can be seen from figure 3 on the following page. Here, the relationship is obviously positive while for figure 4 on the next page the variation of the policy innovation is too small in order to make valid conclusions from simply eye-balling the data. Even though Ellingsen, Söderström and Masseng (ibid.) have more observations and variation in both rates, they do not find a clear sign of their correlation for exogenous policy days which supports their theoretical predictions as well as the findings in this sample.

For a statistically more sound procedure, regression (5) with robust standard errors for maturities up to 10 years has been estimated in order to test the authors' hypotheses. Table 6 on page 40 lists the estimated parameters alongside their corresponding p-values in parentheses underneath. While the coefficient for non-policy days is highly significant and decreasing with n, β_n^{End} is not significantly different from zero for higher maturities; this is

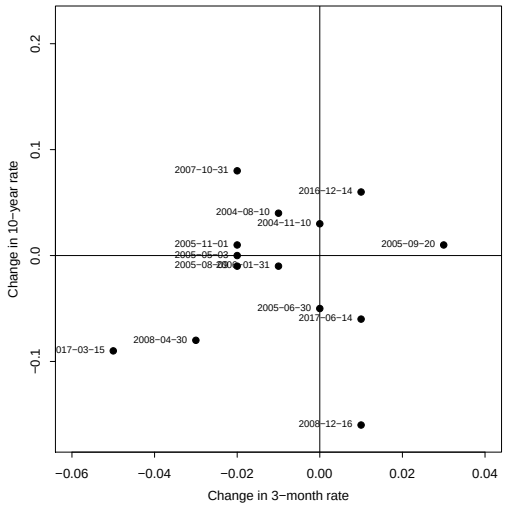


Figure 3: Response of the 10-year interest rate to a change in the 3-month rate for endogenous policy events; *source:* own depiction based on Ellingsen, Söderström and Masseng (2003).

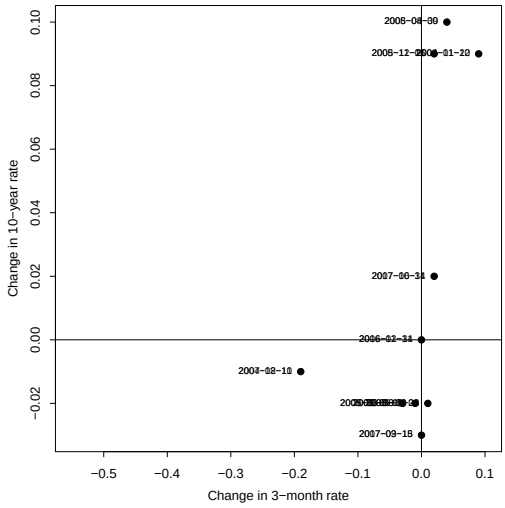


Figure 4: Response of the 10-year interest rate to a change in the 3-month rate for exogenous policy events; *source:* own depiction based on Ellingsen, Söderström and Masseng (2003).

Table 6: Regression results for yield curve response to short rate movements on classified policy days.

	6m	1y	2y	3y	5y	7y	10y	20y	30y
α_n	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	(0.92)	(0.91)	(0.65)	(0.59)	(0.51)	(0.47)	(0.43)	(0.34)	(0.40)
β_n^{NP}	0.59	0.46	0.36	0.35	0.32	0.28	0.23	0.17	0.16
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
β_n^{End}	0.81	0.65	0.31	0.27	0.18	0.15	0.07	0.01	0.01
	(0.00)	(0.08)	(0.63)	(0.70)	(0.79)	(0.82)	(0.89)	(0.91)	(0.93)
β_n^{Ex}	0.83	0.43	0.26	0.22	0.34	0.39	0.35	0.33	0.2
	(0.00)	(0.68)	(0.88)	(0.89)	(0.79)	(0.75)	(0.71)	(0.19)	(0.40)
$\bar{R}^2$	0.58	0.35	0.11	0.09	0.06	0.04	0.03	0.02	0.02
$\beta_n^{NP} = \beta_n^{End}$	0.00	0.00	0.58	0.35	0.17	0.19	0.11	0.08	0.09
$\beta_n^{End} = \beta_n^{Ex}$	0.87	0.12	0.81	0.83	0.59	0.41	0.32	0.24	0.45

even more pronounced for β_n^{Ex}. Given the high number of non-policy days and very limited sample of classified policy days, this is not surprising. Nevertheless, hypothesis 2, which states that long-term interest rates positively correlate to endogenous policy moves but respond negatively to exogenous ones, cannot be answered with certainty since neither coefficient is statistically significant from zero. Even though it appears hard to justify the tilt of the yield curve through exogenous policy moves due to the corresponding positive coefficients, this might be traced back to the lack of sufficient variation in yield curve changes after exogenous and endogenous policy moves on top of the limited amount of them in the sample. Furthermore, since not even the sign of higher maturity coefficients is according to theory, a different type of conduct of monetary policy than in earlier periods might be a reason that the findings of Ellingsen, Söderström and Masseng (2003) could not be replicated. As discussed in section 2, the aftermath of the financial crisis has been accompanied by very transparent monetary policy with the Fed engaging in clear forward guidance in order to calm markets and re-establish trust among financial institutes. Hence, the QE announcements as well as major interest rate adjustments were well anticipated which might

have affected significance of the respective coefficients.

For hypothesis 3, which states that all interest rates behave similarly and positive to information on non-policy days as well as endogenous policy days, I come to the same conclusion as Ellingsen, Söderström and Masseng (2003); a statistical test for equality of parameters clearly rejects it for maturities larger than one year. Equality is also rejected for the comparison of coefficients for endogenous and exogenous policy days. The final prediction of the model in Ellingsen and Söderström (2001) is that for all maturities, the magnitude of the effect after a non-policy day or endogenous policy day falls with maturity as expressed in hypothesis 4. This result can be replicated in this sample even though significance of parameters is low for higher maturities. Including announcements of unconventional monetary policy actions into the regression does not change the results as reported in table 10 on page XIV in the appendix. One reason for the insignificant QE dummies might be that most of these decisions were well anticipated as they have been addressed in previous FOMC meetings and were hence already priced in before the actual decision. Running the regressions for the non-crisis period only, i.e. before January 1st 2008, does not change the tendency of the results as can be seen in tables 11 and 12 on page XV in the appendix which might have to do with the considerably smaller sample than in the initial paper by Ellingsen, Söderström and Masseng (2003) since even the differentiation between policy and non-policy days does not fully correspond to their findings even though the classification is not taken into account on this level.

I refrain from including adjusted estimates as well as a sensitivity analysis as conducted by Ellingsen, Söderström and Masseng (ibid.) since the aim of this thesis is to suggest an alternative classification procedure rather than an in depth assessment of the economic model introduced in Ellingsen and Söderström (2001).

6. Discussion

As discussed above, the main prediction of the model in Ellingsen and Söderström (2001) could not be verified in this thesis. Since previous studies have shown empirical evidence for yield curve tilts after exogenous target rate adjustments and shifts after endogenous target rate adjustments, I believe the reason for this mismatch is due to two main differences when it comes to this analysis. First, the difference in sample periods is most likely the major factor to bias the regression outcomes. Secondly, the classification task has been conducted in an automated way rather than human inspection.

When it comes to the sample period, section 2 describes how the conduct of monetary policy differed throughout the 20th century. Ellingsen, Söderström and Masseng (2003) even noted a few adjustment dates that were completely unnoticed by markets and had to refer to previous studies in order to identify such. In the 21st century, however, transparency of monetary authorities has already become state of the art practice in the Western world and seen as an indispensable element of successful implementation of policy. Even more, due to the financial crisis of 2007 and its aftermath, the Fed took particular care to clearly laying out the future steps it intended to take in order to calm markets and aid resuming trust into the financial system. Hence, most steps were well communicated and expected before, even though not with clear dates they would come into action, and the moment of surprise was consequently much lower than during the 80s and 90s. As Ellingsen, Söderström and Masseng (ibid.) point out, unanticipated Fed actions have a more pronounced impact on financial markets and thus this strategy of increased forward guidance might be one of the key reasons for why exogenous events did not have the influence on the yield curve that was expected. Naturally, the ZLB as well as QE engagements together with comparatively few target rate adjustments might have supported this fact as well, even though it is hard to explain the wrong sign on the coefficient for exogenous events for long maturities. Nevertheless, I tried to control for unusual monetary policy actions such as QE in the regressions in chapter 5

in order to make results comparable to those in the 80s and 90s where such monetary authority manoeuvres had not been in place. This way, the true coefficients of endogenous and exogenous policy moves could have been estimated but since there were rumours on top of well communicated future actions, it is practically impossible to pin down one exact date at which QE became expected by financial markets; this fact might be the most plausible explanation for the QE coefficients being insignificant throughout the analysis.

The classification task, on the other hand, has been delegated to be performed by computer algorithms rather than the personal intervention of a trained economist. Even though this procedure clearly benefits due to its increase in objectivity and replicability, breaking texts down to a few terms is a strong simplification, especially for the highly specified task at hand. Nevertheless, since many results were comparable to Ellingsen, Söderström and Masseng (2003) and running the program on their sample showed promising results, I am confident that the classification is of subordinate cause for the mismatch in empirical findings. Especially the ML section is in line, however, with Pang, Lee and Vaithyanathan (2002) who compare NB, ME and svm to classify reviews and find that these algorithms perform not as good for sentiment classification as on classical topic-based categorization. Finally, as pointed out in section 5, for a classification task with a specification level this high, a much larger training set compared to the action set is needed in order to achieve meaningful results.

7. Conclusion

Ellingsen and Söderström (2001) introduce an economic model which relates monetary policy actions to yield curve movements and succeeds in implementing the empirical feature of yield curve parallel shifts after a target rate adjustment has been performed which was interpreted by market participants as endogenous response to new information about the economy. If the change

is instead interpreted as exogenous shift in preference by the Fed, the model mimics observations that show a tilt in the yield curve. While Ellingsen, Söderström and Masseng (2003) find empirical evidence for this theoretical model by first classifying target rate adjustments by analysing newspaper articles by hand, this thesis automates their procedure, using state of the art text mining and machine learning procedures.

Even though a variety of intuitive and simulation checks have been performed on my alternative classification task, I fail to replicate their findings which is mainly due to the limited amount of variation in the target rate adjustments as well as unprecedented interventions by the monetary authorities during the sample period at hand. Conversely, this implies that the model does not account for alternative policy actions that go beyond adjusting the target rate and falls short of covering the complete range of possible monetary policy actions and their impact on market interest rates.

Furthermore, the techniques applied are standard methods which are well discussed in the computer science literature and applied on a variety of topics. This implies that their utility has been proven in many fields but conversely, their level of specialisation is quite low. Since causality cannot be established as transparently as with common econometric means, they have to be applied with great care in such a specialised setting and a lot of manual labour has to be put into the selection of informative training sets. An additional issue with respect to the feasibility of the technique might be that the classification algorithm could not be applied on the sample used by Ellingsen, Söderström and Masseng (ibid.) for which only few articles were available which would have supported the accuracy of the automated technique.

One extension that might mitigate this issue is applying bootstrapping to the training sample in order to get more meaningful data to train the algorithms. Since this is beyond the scope of this thesis, it is left for further discussions alongside the application of more sophisticated supervised learning algorithms as well as clustering and unsupervised methods. On top of that, the elastic database (Gormley and Tong 2015) offers a very convenient way to extend a dataset consisting of articles through its implemented

feature to scan existing data with respect to similarities. By utilising this feature, re-running the analysis of this thesis on a sample period with normal interest rates will be a great asset in evaluating the accuracy of the economic model at hand as well as the applied methodology.

References

Athey, Susan and Guido W. Imbens (2017). "The state of applied econometrics: Causality and policy evaluation". In: *The Journal of Economic Perspectives* 31.2, pp. 3–32.

Barro, Robert J. and David B. Gordon (1983). "Rules, discretion and reputation in a model of monetary policy". In: *Journal of monetary economics* 12.1, pp. 101–121.

Berger, Adam L., Vincent J. Della Pietra and Stephen A. Della Pietra (1996). "A maximum entropy approach to natural language processing". In: *Computational linguistics* 22.1, pp. 39–71.

Blei, David M., Andrew Y. Ng and Michael I. Jordan (2003). "Latent dirichlet allocation". In: *Journal of machine Learning research* 3.Jan, pp. 993–1022.

Blinder, Alan S. et al. (2010). "Quantitative easing: entrance and exit strategies". In: *Federal Reserve Bank of St. Louis Review* 92.6, pp. 465–479.

Bomfim, Antulio N. (2003). "Pre-announcement effects, news effects, and volatility: Monetary policy and the stock market". In: *Journal of Banking & Finance* 27.1, pp. 133–151.

Breiman, Leo et al. (2001). "Statistical modeling The two cultures (with comments and a rejoinder by the author)". In: *Statistical science* 16.3, pp. 199–231.

Calvo, Guillermo A. (1978). "On the time consistency of optimal policy in a monetary economy". In: *Econometrica: Journal of the Econometric Society*, pp. 1411–1428.

Ding, Xiaowen, Bing Liu and Philip S. Yu (2008). "A holistic lexicon-based approach to opinion mining". In: *Proceedings of the 2008 international conference on web search and data mining*, pp. 231–240.

Ding, Xiaowen, Bing Liu and Lei Zhang (2009). "Entity discovery and assignment for opinion mining applications". In: *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 1125–1134.

Ellingsen, Tore and Ulf Söderström (2001). "Monetary policy and market interest rates". In: *The American Economic Review* 91.5, pp. 1594–1607.

Ellingsen, Tore, Ulf Söderström and Leonardo Masseng (2003). "Monetary policy and the bond market". In: *Unpublished manuscript, Stockholm School of Economics.*

Esuli, Andrea and Fabrizio Sebastiani (2006). "Determining Term Subjectivity and Term Orientation for Opinion Mining". In: *EACL*. Vol. 6, p. 2006.

Evans, Charles L. and David A. Marshall (1998). "Monetary policy and the term structure of nominal interest rates: evidence and theory". In: *Carnegie-Rochester Conference Series on Public Policy*. Vol. 49, pp. 53–111.

Fawley, Brett W., Christopher J. Neely et al. (2013). "Four stories of quantitative easing". In: *Federal Reserve Bank of St. Louis Review* 95.1, pp. 51–88.

Federal Reserve System, Board of Governors of the (2017). *Open Market Operations*. URL: `https://www.federalreserve.gov/monetarypolicy/openmarket.htm` (visited on 20/08/2017).

Feldman, Ronen (2013). "Techniques and applications for sentiment analysis". In: *Communications of the ACM* 56.4, pp. 82–89.

Fellbaum, Christiane (1998). *WordNet*. Wiley Online Library.

Friedman, Jerome, Trevor Hastie and Robert Tibshirani (2001). *The elements of statistical learning*. Vol. 1. Springer series in statistics New York.

Ganapathibhotla, Murthy and Bing Liu (2008). "Mining opinions in comparative sentences". In: *Proceedings of the 22nd International Conference on Computational Linguistics*. Vol. 1, pp. 241–248.

Gormley, Clinton and Zachary Tong (2015). *Elasticsearch: The Definitive Guide: A Distributed Real-Time Search and Analytics Engine*. O'Reilly Media, Inc.

Guo, Honglei et al. (2009). "Product feature categorization with multilevel latent semantic association". In: *Proceedings of the 18th ACM conference on Information and knowledge management*, pp. 1087–1096.

Gürkaynak, Refet S., Brian P. Sack and Eric T. Swanson (2004). "Do actions speak louder than words? The response of asset prices to monetary policy actions and statements". In: *FEDS Working paper* 66.

Hattori, Masazumi, Andreas Schrimpf and Vladyslav Sushko (2016). "The response of tail risk perceptions to unconventional monetary policy". In: *American Economic Journal: Macroeconomics* 8.2, pp. 111–136.

Hatzivassiloglou, Vasileios and Kathleen R. McKeown (1997). "Predicting the semantic orientation of adjectives". In: *Proceedings of the eighth conference on European chapter of the Association for Computational Linguistics*, pp. 174–181.

Hautsch, Nikolaus and Dieter Hess (2002). "The processing of non-anticipated information in financial markets: Analyzing the impact of surprises in the employment report". In: *European Finance Review* 6.2, pp. 133–161.

Hautsch, Nikolaus, Dieter Hess and David Veredas (2011). "The impact of macroeconomic news on quote adjustments, noise, and informational volatility". In: *Journal of Banking & Finance* 35.10, pp. 2733–2746.

Hess, Dieter (2004). "Determinants of the relative price impact of unanticipated information in US macroeconomic releases". In: *Journal of Futures Markets* 24.7, pp. 609–629.

Hornik, Kurt (2016). *openNLP: Apache OpenNLP Tools Interface*. URL: `https://CRAN.R-project.org/package=openNLP`.

Jindal, Nitin and Bing Liu (2006). "Identifying comparative sentences in text documents". In: *Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval*, pp. 244–251.

Kamps, Jaap et al. (2004). "Using WordNet to Measure Semantic Orientations of Adjectives". In: *LREC*. Vol. 4, pp. 1115–1118.

Kanayama, Hiroshi and Tetsuya Nasukawa (2006). "Fully automatic lexicon expansion for domain-oriented sentiment analysis". In: *Proceedings of the 2006 conference on empirical methods in natural language processing*, pp. 355–363.

Kydland, Finn E. and Edward C. Prescott (1977). "Rules rather than discretion: The inconsistency of optimal plans". In: *Journal of political economy* 85.3, pp. 473–491.

Labor Statistics, Bureau of (2017). *Economic Releases*. URL: `https://www.bls.gov/bls/archived_sched.htm` (visited on 15/09/2017).

Liu, Bing (2010). "Sentiment Analysis and Subjectivity". In: *Handbook of natural language processing* 2, pp. 627–666.

Liu, Bing (2012). "Sentiment analysis and opinion mining". In: *Synthesis lectures on human language technologies* 5.1, pp. 1–167.

Lucca, David O. and Emanuel Moench (2015). "The Pre-FOMC Announcement Drift". In: *The Journal of Finance* 70.1, pp. 329–371.

Manela, Asaf and Alan Moreira (2017). "News implied volatility and disaster concerns". In: *Journal of Financial Economics* 123.1, pp. 137–162.

Marcus, Mitchell P., Mary Ann Marcinkiewicz and Beatrice Santorini (1993). "Building a large annotated corpus of English: The Penn Treebank". In: *Computational linguistics* 19.2, pp. 313–330.

Meyer, David, Kurt Hornik and Ingo Feinerer (2008). "Text mining infrastructure in R". In: *Journal of statistical software* 25.5, pp. 1–54.

Mishkin, Frederic S. (2007). *Monetary policy strategy*. MIT Press.

Pang, Bo, Lillian Lee and Shivakumar Vaithyanathan (2002). "Thumbs up? Sentiment classification using machine learning techniques". In: *Proceedings of the ACL-02 conference on Empirical methods in natural language processing*. Vol. 10, pp. 79–86.

Peersman, Gert (2002). "Monetary policy and long term interest rates in Germany". In: *Economics Letters* 77.2, pp. 271–277.

Riloff, Ellen, Siddharth Patwardhan and Janyce Wiebe (2006). "Feature subsumption for opinion analysis". In: *Proceedings of the 2006 conference on empirical methods in natural language processing*, pp. 440–448.

Rish, Irina, Joseph Hellerstein and Jayram Thathachar (2001). "An analysis of data characteristics that affect naive Bayes performance". In: *IBM TJ Watson Research Center* 30.

Samuelson, Paul A. and Robert M. Solow (1960). "Analytical aspects of anti-inflation policy". In: *The American Economic Review* 50.2, pp. 177–194.

Silge, Julia and David Robinson (2016). "tidytext: Text Mining and Analysis Using Tidy Data Principles in R". In: *JOSS* 1.3. DOI: 10.21105/joss.00037. URL: http://dx.doi.org/10.21105/joss.00037.

Silge, Julia and David Robinson (2017). *Text Mining with R: A Tidy Approach*. O'Reilly Media, Inc.

Svensson, Lars E. O. (1997). "Inflation forecast targeting: Implementing and monitoring inflation targets". In: *European economic review* 41.6, pp. 1111–1146.

Tetlock, Paul C. (2007). "Giving content to investor sentiment: The role of media in the stock market". In: *The Journal of Finance* 62.3, pp. 1139–1168.

Turney, Peter D. (2002). "Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews". In: *Proceedings of the 40th annual meeting on association for computational linguistics*, pp. 417–424.

Wang, Hongning, Yue Lu and Chengxiang Zhai (2010). "Latent aspect rating analysis on review text data: a rating regression approach". In: *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 783–792.

Wiebe, Janyce (2000). "Learning subjective adjectives from corpora". In: *AAAI/IAAI* 20.0.

Wiebe, Janyce et al. (2004). "Learning subjective language". In: *Computational linguistics* 30.3, pp. 277–308.

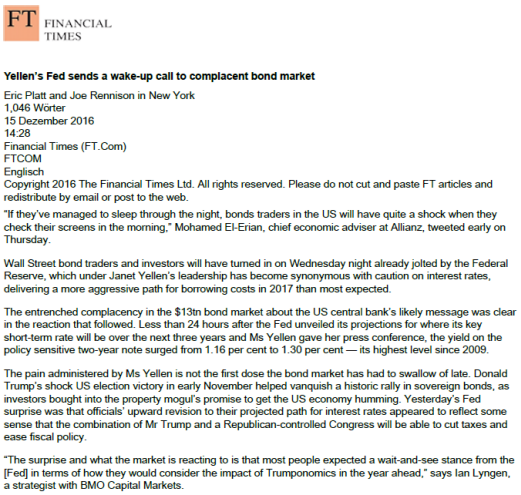
Wiebe, Janyce M., Rebecca F. Bruce and Thomas P. O'Hara (1999). "Development and use of a gold-standard data set for subjectivity classifications". In: *Proceedings of the 37th annual meeting of the Association for Computational Linguistics on Computational Linguistics*, pp. 246–253.

Williams, Graham (2011). *Data mining with Rattle and R: The art of excavating data for knowledge discovery*. Springer Science & Business Media.

Yu, Hong and Vasileios Hatzivassiloglou (2003). "Towards answering opinion questions: Separating facts from opinions and identifying the polarity of opinion sentences". In: *Proceedings of the 2003 conference on Empirical methods in natural language processing*, pp. 129–136.

A. Appendix

A.1. Exemplary newspaper articles



FT FINANCIAL TIMES

Yellen's Fed sends a wake-up call to complacent bond market

Eric Platt and Joe Rennison in New York
1,046 Wörter
15 Dezember 2016
14:28
Financial Times (FT.Com)
FTCOM
Englisch

"If they've managed to sleep through the night, bonds traders in the US will have quite a shock when they check their screens in the morning," Mohamed El-Erian, chief economic adviser at Allianz, tweeted early on Thursday.

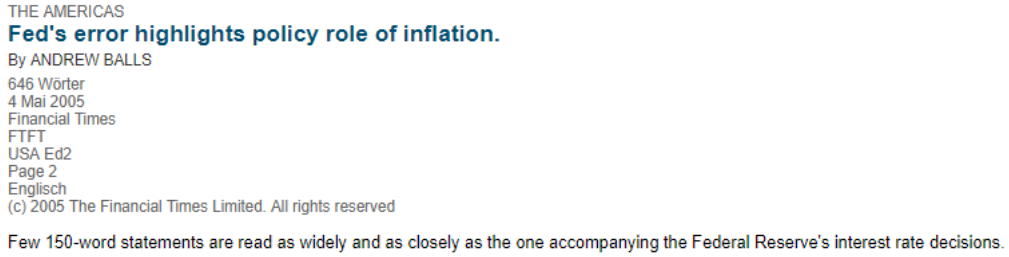
Wall Street bond traders and investors will have turned in on Wednesday night already jolted by the Federal Reserve, which under Janet Yellen's leadership has become synonymous with caution on interest rates, delivering a more aggressive path for borrowing costs in 2017 than most expected.

The entrenched complacency in the $13tn bond market about the US central bank's likely message was clear in the reaction that followed. Less than 24 hours after the Fed unveiled its projections for where its key short-term rate will be over the next three years and Ms Yellen gave her press conference, the yield on the policy sensitive two-year note surged from 1.16 per cent to 1.30 per cent — its highest level since 2009.

The pain administered by Ms Yellen is not the first dose the bond market has had to swallow of late. Donald Trump's shock US election victory in early November helped vanquish a historic rally in sovereign bonds, as investors bought into the property mogul's promise to get the US economy humming. Yesterday's Fed surprise was that officials' upward revision to their projected path for interest rates appeared to reflect some sense that the combination of Mr Trump and a Republican-controlled Congress will be able to cut taxes and ease fiscal policy.

"The surprise and what the market is reacting to is that most people expected a wait-and-see stance from the [Fed] in terms of how they would consider the impact of Trumponomics in the year ahead," says Ian Lyngen, a strategist with BMO Capital Markets.

Figure 5: Example of an article that has been used for the sentiment determination.



THE AMERICAS

Fed's error highlights policy role of inflation.

By ANDREW BALLS
646 Wörter
4 Mai 2005
Financial Times
FTFT
USA Ed2
Page 2
Englisch

Few 150-word statements are read as widely and as closely as the one accompanying the Federal Reserve's interest rate decisions.

Figure 6: Example of an article that has been used for the sentiment determination.

A.2. Part-of-Speech tags.

Table 7: The Penn English Treebank POS tagset (Marcus, Marcinkiewicz and Santorini 1993).

CC	Coordinating conjunction
CD	Cardinal number
DT	Determiner
EX	Existential *there*
FW	Foreign word
IN	Preposition/subordinating conjunction
JJ	Adjective
JJR	Adjective, comparative
JJS	Adjective, superlative
LS	List item marker
MD	Modal
NN	Noun, singular or mass
NNS	Noun, plural
NNP	Proper noun, singular
NNPS	Proper noun, plural
PDT	Predeterminer
POS	Possessive ending
PRP	Personal pronoun
PRP$	Possessive pronoun
RB	Adverb
RBR	Adverb, comparative
RBS	Adverb, superlative
RP	Particle
SYM	Symbol (mathematical or scientific)
TO	*to*
UH	Interjection
VB	Verb, base form
VBD	Verb, past tense
VBG	Verb, gerund/present participle
VBN	Verb, past participle
VBP	Verb, non-3rd person singular present
VBZ	Verb, 3rd person singular present
WDT	*wh*-determiner
WP	*wh*-pronoun
WP$	Possessive *wh*-pronoun
WRB	*wh*-adverb

```
[1] "c(/JJ \"/`` With/IN almost/RB a/DT year/NN having/VBG passed/VBN since/IN it/PRP completed/VBD a/DT steady/JJ barrage/N
N of/IN interest/NN rate/NN cuts/NNS meant/VBD to/TO boost/VB the/DT economy/NN in/IN 2002/CD ,/, the/DT Federal/NNP Reserve
/NNP ran/VBD out/IN of/IN patience/NN Wednesday/NNP and/CC fired/VBD one/CD more/JJR double-barreled/JJ blast/NN in/IN the/D
T hope/NN of/IN changing/VBG the/DT outlook/NN for/IN 2003.\"/CD ,/, \"/`` \"/`` ,/, \"/'' In/IN making/VBG a/DT larger-than
-expected/JJ half-percentage-point/JJ cut/NN in/IN short-term/JJ rates/NNS ,/, the/DT Fed/NNP said/VBD that/IN consumer/NN a
nd/CC business/NN worries/NNS about/IN the/DT slow/JJ economy/NN and/CC war/NN with/IN Iraq/NNP have/VBP been/VBN causing/VB
G business/NN activity/NN to/TO falter/VB ,/, despite/IN a/DT nearly/RB two-year-old/JJ policy/NN to/TO make/VB money/NN eas
ier/JJR to/TO borrow/VB ./. In/IN making/VBG its/PRP$ decision/NN ,/, the/DT Fed/NNP also/RB indicated/VBD that/IN it/PRP pr
obably/R... <truncated>
```

Figure 7: Example of an article snippet with POS tags.

A.3. Sentiment indicators for count-based evaluation

Table 8: Terms that have been defined as indicators for endogenous or exogenous sentiment.

Endogenous	Exogenous
unchanged	switch
remain	unexpected
steady	larger
continue	sharp
sustain	less
prevent	policy change
as a result	new
economic recovery	surprise
recovery	longer period
slowdown	for a while
downturn	despite
recession	gradual
crisis	signal
slow growth	more/less aggressive
high growth	tightening
unemployment	loosening
labor	cautious
productivity	CPI
real estate	pressure
housing	productivity
financial market	ahead
natural catastrophe	pause
hurricane	
terrorist attack	

A.4. Additional regression results

Table 9: Regression results for yield curve response to short rate movements on policy days and non-policy days after controlling for QE announcements.

	6m	1y	2y	3y	5y	7y	10y	20y	30y
α_n	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	(0.96)	(0.85)	(0.66)	(0.58)	(0.54)	(0.52)	(0.47)	(0.36)	(0.42)
β_n^{NP}	0.59	0.46	0.36	0.35	0.32	0.28	0.23	0.17	0.16
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
β_n^{P}	0.83	0.64	0.33	0.28	0.2	0.16	0.1	0.03	0.01
	(0.00)	(0.01)	(0.39)	(0.49)	(0.62)	(0.64)	(0.75)	(0.76)	(0.93)
β_n^{QE}	0	0.01	-0.03	-0.02	-0.06	-0.08	-0.07	-0.04	-0.03
	(0.15)	(0.44)	(0.49)	(0.69)	(0.24)	(0.14)	(0.30)	(0.20)	(0.25)
$\bar{R}^2$	0.58	0.36	0.11	0.09	0.06	0.05	0.03	0.02	0.02
$\beta_n^{NP} = \beta_n^{P}$	0.00	0.00	0.69	0.39	0.18	0.22	0.13	0.08	0.07

Table 10: Regression results for yield curve response to short rate movements on classified policy days after controlling for QE announcements.

	6m	1y	2y	3y	5y	7y	10y	20y	30y
α_n	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	(0.93)	(0.90)	(0.68)	(0.60)	(0.55)	(0.52)	(0.47)	(0.36)	(0.42)
β_n^{NP}	0.59	0.46	0.36	0.35	0.32	0.28	0.23	0.17	0.16
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
β_n^{End}	0.81	0.65	0.31	0.27	0.18	0.15	0.07	0.01	0.01
	(0.00)	(0.08)	(0.63)	(0.70)	(0.79)	(0.82)	(0.89)	(0.91)	(0.93)
β_n^{Ex}	0.83	0.43	0.26	0.22	0.34	0.39	0.35	0.33	0.2
	(0.00)	(0.68)	(0.88)	(0.89)	(0.79)	(0.75)	(0.71)	(0.19)	(0.40)
β_n^{QE}	0	0.01	-0.03	-0.02	-0.06	-0.08	-0.07	-0.04	-0.03
	(0.15)	(0.44)	(0.49)	(0.69)	(0.24)	(0.14)	(0.30)	(0.20)	(0.25)
$\bar{R}^2$	0.58	0.35	0.11	0.09	0.06	0.05	0.04	0.02	0.02
$\beta_n^{NP} = \beta_n^{End}$	0.00	0.00	0.58	0.35	0.18	0.19	0.11	0.08	0.09
$\beta_n^{End} = \beta_n^{Ex}$	0.87	0.12	0.81	0.83	0.59	0.41	0.32	0.24	0.45

Table 11: Regression results for yield curve response to short rate movements on policy days and non-policy days before the period of ZIRP.

	6m	1y	2y	3y	5y	7y	10y	20y	30y
α_n	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	(0.32)	(0.68)	(0.84)	(0.69)	(0.58)	(0.52)	(0.54)	(0.47)	(0.51)
β_n^{NP}	0.52	0.46	0.34	0.32	0.3	0.26	0.22	0.15	0.13
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
β_n^{P}	1.05	0.78	0.62	0.54	0.54	0.51	0.41	0.29	0.18
	(0.00)	(0.15)	(0.39)	(0.41)	(0.31)	(0.30)	(0.28)	(0.36)	(0.56)
$\bar{R}^2$	0.54	0.31	0.09	0.07	0.06	0.05	0.04	0.02	0.02
$\beta_n^{NP} = \beta_n^{P}$	0.00	0.01	0.15	0.29	0.24	0.21	0.31	0.40	0.76

Table 12: Regression results for yield curve response to short rate movements on classified policy days before the period of ZIRP.

	6m	1y	2y	3y	5y	7y	10y	20y	30y
α_n	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	(0.28)	(0.60)	(0.90)	(0.73)	(0.61)	(0.55)	(0.56)	(0.48)	(0.52)
β_n^{NP}	0.52	0.46	0.34	0.32	0.3	0.26	0.22	0.15	0.13
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
β_n^{End}	1.29	1.17	1	0.88	0.73	0.62	0.41	0.17	0.14
	(0.00)	(0.00)	(0.23)	(0.22)	(0.49)	(0.57)	(0.67)	(0.85)	(0.88)
β_n^{Ex}	0.85	0.43	0.24	0.19	0.32	0.37	0.35	0.32	0.2
	(0.00)	(0.69)	(0.88)	(0.90)	(0.80)	(0.76)	(0.72)	(0.72)	(0.81)
$\bar{R}^2$	0.54	0.32	0.09	0.07	0.06	0.05	0.04	0.02	0.02
$\beta_n^{NP} = \beta_n^{End}$	0.00	0.00	0.03	0.08	0.18	0.25	0.51	0.94	1.00
$\beta_n^{End} = \beta_n^{Ex}$	0.02	0.00	0.06	0.10	0.34	0.53	0.87	0.68	0.86