Understanding "Fedspeak": Identifying the Sources of Market Sentiment in Central Bank Communications

A thesis presented by

Kevin J. Wu

to the department of
Applied Mathematics
in partial fulfillment of the honors requirements
for the degree of
Bachelor of Arts

Harvard College
Cambridge, MA
April 1, 2014

Acknowledgments

The writing of this thesis has been a truly humbling experience, and I am grateful for the guidance and support of those who have helped bring this project to its conclusion. I am incredibly fortunate to have had as my advisor Professor Benjamin Friedman, who patiently guided me through every stage of the thesis-writing process and kept me grounded with the big-picture questions. I am also enormously indebted to Professor Luke Miratrix, whose contributions to feature selection for text classification form the backbone of this paper; without his direction and generosity, this paper would not have been what it is. I also want to thank Professors John Campbell and Greg Lewis for their guidance when this thesis was still an idea in its infancy; and Professors David Parkes and Alice Oh, who introduced me to the tools and techniques of the field of Natural Language Processing. Next, I want to thank Andrew Kellerman and Evan Schnidman for their unique perspective on the relationship between Fed communications and market sentiment. Finally, I am grateful for the advice and support of Siddarth Viswanathan, Rohit Ramani, and Yi-an Chang, whose honest appraisals of my work have been crucial in its final stages.

But were it not for the support and encouragement of my family, I do not know how I would have begun to embark on such a project. To my parents, whose passion for learning I have always admired, I want to extend my warmest gratitude and appreciation.

Abstract

Financial market sentiment analysis is a rapidly expanding topic of research both in academic and industry circles, in no small part due to the availability and diversity of new textual sources such as social media. In this thesis, I apply methods from the statistical machine learning and Natural Language Processing (NLP) fields to examine the stock market effect of one of the most highly anticipated economic releases in the financial community: the Federal Open Market Committee's post-meeting announcement. Although financial market sentiment analysis is commonly used to predict returns from textual data, I use returns as a way to categorize statements as either positive or negative. The novelty of this approach is the use of sparse regression to identify of a small number of phrases from central bank announcements that characterize either positive or negative market sentiment. I find that sparse regression is able to detect the presence of key phrases and sentences from Fed communications, and that potential sources of positive market sentiment are more prevalent in the text of FOMC statements than sources of negative market sentiment. However, sparse regression falls short of generating a statistically significant subset of textual features.

Contents

List of Figures

List of Tables

1 Introduction and Background

"There is excessive emphasis on reading the entrails of the Federal Reserve. We get put on a table and sliced open."

— Richard Fisher, *New York Times* (2008)

Investors have learned to heed central bankers' every word for insight into the future path of asset purchases and interest rates. As President of the Federal Reserve Bank of Dallas Richard Fisher pointed out in 2008, public scrutiny of the Fed is unrelenting. Between the minutes and transcripts of the Federal Open Market Committee, the speeches delivered by its voting members, and the press conferences delivered by central bankers, no part of the Fed's communication strategy is left unexamined by market participants and pundits in the financial press. It is not surprising, therefore, that one of the most heavily anticipated press releases in the financial community is the public statement released by the Federal Open Market Committee (FOMC) after each of its regularly scheduled meetings[1]. As the primary means by which changes in monetary policy are announced, these press releases form the crux of the central bank's communication strategy. Moreover, the text of each statement contains the Fed's assessment of the current state of the economy and its outlook for inflation and economic growth—key determinants of the path of future monetary policy.

The deliberation involved in the drafting of each statement and the scrutiny over the Fed's every word warrant an examination into how particular language used by the Fed has the potential to move markets. Although the effect of the Fed's announcements on financial markets has been the focus of extensive research by economists and policymakers, there are relatively few papers which treat the text of central bank communications itself as data. This thesis aims to bridge that gap, incorporating methods from the statistical machine learning and natural language processing (NLP) fields to examine the relationship between the semantic content of Fed statements and financial markets. Using the direction of returns on S&P index prices over a narrow time window surrounding each statement's release, I use sparse regression techniques to establish a relationship between particular phrases used by the Federal Reserve and market sentiment.

The structure of this thesis is as follows: in Chapter 2, I summarize the existing academic literature on the relationship between central bank communications and financial markets and discuss the need for a way to directly map language onto a financial market response variable. In Chapter 3, I examine the use of high-frequency intraday S&P 500 ETF prices as a gauge of market sentiment and explore the application of statistical NLP techniques to the corpus of Federal Reserve statements. In Chapter 4, I introduce the statistical literature on textual feature selection and present the sparse regression model and algorithm I use to extract phrases from Fed statements. Finally, in Chapter 5, I present

[1] Referred to informally as "FOMC statements," "Federal Reserve statements," or "Fed statements"—terms I will use interchangeably in this paper.

the results of that analysis. I find that although feature selection is able to flag key sentences from Fed statements as indicators of market sentiment, the results are not strong enough to be significant—in part, I argue, due to the relatively small number of observations and the inherent difficulty of using a statistical model to capture more complex linguistic structures.

2 A Review of the Economic Literature on Fed Communications

Former Fed chairman Ben Bernanke and Kenneth Kuttner (2005) are one of the first to examine the relationship between Federal Reserve policy and equity markets. Using federal funds futures data to measure the unexpected component of rate announcements, a methodology first proposed by Kuttner (2000), the authors find a significant and highly negative correlation between unexpected changes in the federal funds rate and same-day returns on the S&P 500. To isolate the effect of Federal Reserve announcements from the effect of other economic releases occurring on the same day, Gürkaynak, Sack, and Swanson (2005) perform an identical regression for intraday S&P returns over a 30-minute window surrounding the announcement and find a large improvement in the explainable variance in equity returns.

More importantly, however, Gürkaynak et al. (2005) are one of the first to evaluate the informational content of Fed statements as separate from the explicit policy announcement contained at the beginning of each press release. Using principal components analysis, the authors identify an additional factor, besides the rate announcement, with a significant impact on financial markets and interpret this factor as capturing expectations about the Fed's "future path of policy."[2] They find that changes in the so-called "path factor" account for more than three-quarters of the explainable variation in intraday movements for five- and ten-year Treasury yields, and two-fifths of the explainable variation in intraday S&P 500 returns.

Lucca and Trebbi (2007) assess the effect of FOMC statements on the yield curve for Treasury securities by applying techniques from computational linguistics to come up with the "semantic orientation" of text fragments from Fed statements. The novel contribution of Lucca and Trebbi (2007) is in applying a methodology to assess each statement's textual content in the context of monetary policy and monetary policy expectations. To do this, the authors evaluate the "hawkishness" of sentences and phrases by measuring their relative frequency with the words "hawkish" and "dovish" in Google web searches and in a Factiva database of news articles in a three-day window around each statement's release. Using intraday changes in Treasury yields as the dependent variable, the authors find that changes in semantic scores account for a nearly 10 percentage point increase in R^2 for changes in two- and five-year yields, compared to a baseline regression with only the unexpected component of monetary policy as the right-hand side variable. The advantage of Lucca and Trebbi's approach is that by taking the difference in a statement's semantic orientation between articles published the day after its release and the day before, the authors are able to measure how the overall text of the statement, and not just the rate change, took markets by surprise.

[2] The authors apply principal components analysis to a matrix X whose rows correspond to monetary policy announcements and whose columns correspond to changes in asset prices around the time of each statement's release.

However, the issue of simultaneity arises in Gürkaynak et al. (2005) and Lucca and Trebbi (2007); in both methodologies, changes in financial markets following FOMC statements appear to have some effect on the measurement of the statement's textual content. The factor-based approach to quantifying monetary policy announcements characterizes each FOMC statement by its effects on prices in various financial markets. In the semantic orientation approach, it is unlikely that the post-statement scores being generated in Lucca and Trebbi (2007) have not already incorporated some of the fast-moving developments in market prices, since the frequency with which certain sentence fragments from a Fed statement appear online in conjunction with the word "hawkish" most likely reflects that day's market developments.

In this thesis, I aim to circumvent the potential endogeneity issues of past methods by directly extracting textual features. To do this, I represent text via the "bag of words" approach, a commonly-used technique in applications of NLP that transforms a corpus of raw text documents into a matrix of word and/or phrase frequencies. One notable use of this technique in the economics literature is by Matthew Gentzkow and Jesse Shapiro (2010) in a paper analyzing the media bias of U.S. newspapers. To date, however, there have been very few applications of textual analysis to the study of central bank communications in the academic literature. The unique and novel contribution of this thesis is the use of feature selection techniques to characterize Fed statements by a small number of "market-moving" phrases—a departure from previous methods that focus on coming up with an aggregate measure to characterize each Fed statement.

3 Exploratory Analysis of the FOMC Statement Dataset

The dataset I use contains 115 FOMC statements released between May 1999 and July 2013. I only examine statements published by the FOMC following its eight regularly-scheduled meetings each year; although the FOMC has released a number of other statements to announce policy changes in between its scheduled meetings, these are almost always written during times of particular economic or financial market turmoil (such as the September 17, 2001 statement). In addition, while past papers such as Gürkaynak et al. (2005) include in their dataset press releases dating back to 1990, I start my dataset at May 1999 because prior statements were only released after a decision by the FOMC to raise or lower the federal funds rate target. By restricting my analysis to statements beginning in May 1999, I focus my study of the relationship between Fed announcements and financial markets on an era of increased transparency surrounding the work of central bankers.

3.1 Labeling Sentiment: Intraday S&P Returns

Although most of the economic literature on Federal Reserve communications has focused on changes in the market for Treasury securities, I use changes in the S&P 500 index as the dependent variable because of its broad application in the financial media as a way to gauge market reaction to economic news. In doing so, I also follow the lead of previous papers on financial market sentiment analysis, which have used the index as an indicator of positive or negative sentiment in markets (Paul Tetlock, 2007; Michel Généreux, Thierry Poibeau, and Moshe Koppel, 2011). Since the sparse regression methods I use require a binary dependent variable, I measure each statement's sentiment based only on the direction of returns. Finally, because changes in the S&P 500 index encapsulate more than just the market's interest rate expectations, in this paper I will refer to statements as positive and negative rather than as hawkish or dovish.

To classify the sentiment of Fed statements, I use the SPDR S&P 500 ETF (ticker symbol: SPY) as a proxy for changes in the S&P 500 index. Using high-frequency trade data for SPDR contracts taken from the New York Stock Exchange Trade and Quote (TAQ) database, I calculate the price at each minute as a weighted average of all SPDR trades recorded during that 60-second interval (weighting the average price by the size of each trade). Using the same time window as Gürkaynak et al. (2005) and Lucca and Trebbi (2007), I measure the change in ETF price 10 minutes prior to and 20 minutes after each statement's release.

Summary of Post-Statement Reactions			
	Positive	Negative	*Total*
a) May 1999 - October 2008	28	49	*77*
**before quantitative easing*			
b) December 2008 - July 2013	22	16	*38*
c) Full sample	*50*	*65*	*115*

Table 1: **Summary of Post-Statement Reactions:** This table reports the number of positive and negative changes in SPDR prices in a 30-minute window around each statement's release. Subsample *(a)* contains post-statement SPDR changes during times of conventional monetary policy; subsample *(b)* covers statements released since the beginning of the central bank's quantitative easing programs.

3.1.1 Measuring the Monetary Policy "Surprise"

Each FOMC statement prior to the start of quantitative easing in December 2008 begins with an announcement regarding either a change in monetary policy, or a lack thereof. Before beginning my analysis of the text of FOMC statements, I first demonstrate that even during times of conventional monetary policy, the majority of variance in post-release equity returns cannot be explained by the monetary policy change announced. In Figure 1, I plot the path of SPDR returns as measured from the time of the Fed statement's release to show how press releases in a given year containing identical monetary policy announcements, or belonging to the same monetary policy cycle, can generate very different reactions in stock prices. Next, I empirically examine the relationship between post-release stock returns and monetary policy changes to provide further motivation for a text-based analysis of central bank communications.

To measure the sensitivity of equity index returns to changes in the federal funds target rate during times of conventional monetary policy, I regress the percent change in SPDR price, Δy_i, on the unexpected and expected changes in the federal funds rate target, $\Delta \tilde{r}_i^u$ and $\Delta \tilde{r}_i^e$, respectively:

$$\Delta y_i = \alpha + \beta_1 \Delta \tilde{r}_i^u + \beta_2 \Delta \tilde{r}_i^e + \varepsilon_i \tag{1}$$

To determine the unexpected or "surprise" component of each monetary policy decision, I employ the methodology proposed by Kuttner (2001) that calculates unexpected changes in the target rate using the one-day change in the spot federal funds futures rate according to the formula

$$\Delta \tilde{r}_t^u = \frac{m}{m-t}(f_{s,t}^0 - f_{s,t-1}^0) \tag{2}$$

where $\Delta \tilde{r}_t^u$ is the unexpected change in the federal funds rate on date t of the monetary policy announcement and $f^0_{s,t}$ is the federal funds futures rate for the spot month s on date t. m denotes the number of days in month s (multiplying by $\frac{m}{m-t}$ scales the change in the futures rate by remaining days until the contract's maturity).[34]

The results for the ordinary least-squares regression are shown below in Table 2.

	Full Sample (1999-2008)	Excluding Sept.-Dec., 2008
Intercept	-0.134**	-0.131**
	(-2.14)	(-2.13)
Expected change	0.003	0.003
	(0.81)	(1.06)
Surprise change	-0.009	-0.031*
	(-0.43)	(-1.68)
R^2	0.021	0.105

Table 2: **SPDR Returns and the Components of Monetary Policy: Regression Results**
This table reports the results from regressions of 30-minute SPDR returns on the unexpected and expected components of monetary policy. The first column reports the results from a regression using the full set of data from 1999 to 2008, which includes monetary policy announcements from 78 Fed statements; the second column shows the results from a regression excluding three observations from the peak of the financial crisis. Parentheses contain t-statistics calculated using heteroskedasticity-consistent estimates of the standard errors, and asterisks denote statistical significance at the .05 level (**) and 0.1 level (*).

Excluding observations from the peak of the financial crisis in the fall of 2008, we find that an unexpected hike in the federal funds target rate of 10 basis points is associated with a 0.3 percent decline in the S&P index. More importantly, our regression shows that even during times of "conventional" monetary policy, changes in the target rate can only explain approximately 10 percent of the variance in intraday S&P returns.[5] Given that no other significant market-moving news is regularly scheduled for release within the thirty-minute window around the Fed announcement, the remaining variance in equity market prices must be explained by information contained in the rest of the FOMC statement, which includes the Fed's assessment of and outlook for economic growth and its plan for the path of future

[3] When $t = 1$, the rate on the spot futures contract on the last day of the previous month, $f^0_{s-1,m}$, is used to calculate the change in the futures rate.

[4] The expected component of monetary policy is calculated as the difference between the change in the target rate announced by the FOMC and the monetary policy surprise as described in equation (2).

[5] In a regression of S&P returns against the unexpected component of monetary policy, using an identical 30-minute window for equity index returns, Gürkaynak et al. (2005) find an R^2 of .36. One potential reason for this discrepancy is the different range of observations used in the regression; Gürkaynak et al. (2005) use the set of all monetary policy annoncements from January 1990 to December 2004.

policy. In the next chapter, I present the method used to represent the textual content of FOMC statements and examine the semantic peculiarities of the FOMC statement dataset, which I will refer to from now on as the Fed statement corpus.

Fig. 1: **SPDR Returns over a 30-minute Response Window:** Here I plot the percentage change in SPDR prices relative to the time of each statement's release, typically 2:15 pm EST, over a 30-minute response window.
(Left) Price movements are labeled by the monetary policy announced in each statement. Blue lines show reactions to statements in which no change in the federal funds rate was announced, red (green) lines show price changes after a 25 basis point rate hike (cut), and yellow lines show price changes after a rate cut of more than 25 basis points.
(Right) Price movements are labeled by the monetary policy announcement relative to easing/tightening cycles. Red (green) lines denote "hawkish" ("dovish") changes in policy, i.e. from no rate change to a rate hike, or from a rate cut to no change (from no rate change to a rate cut, or from a rate hike to no change). Blue lines indicate price movements after a monetary policy announcement identical to the one preceding it.

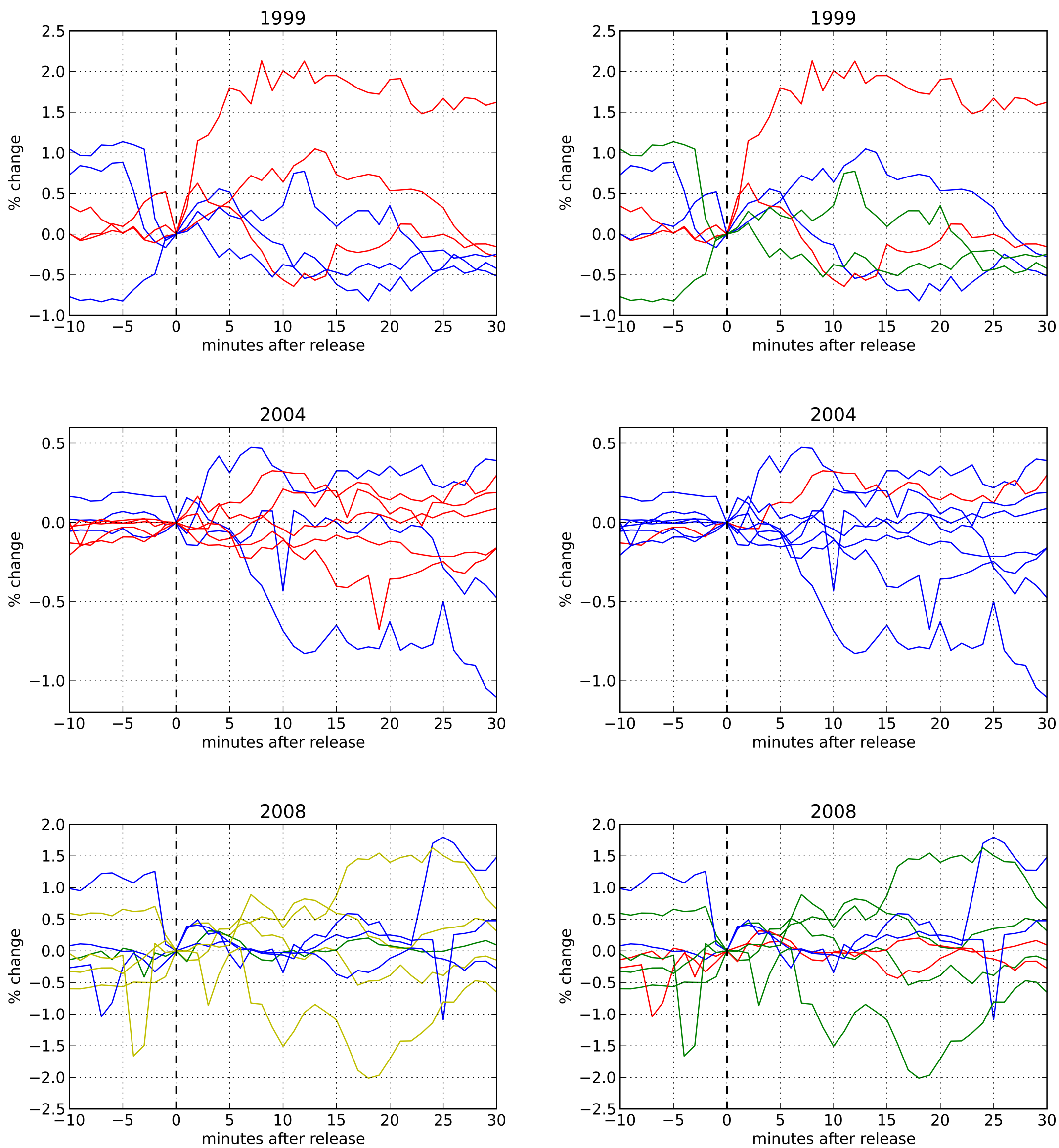
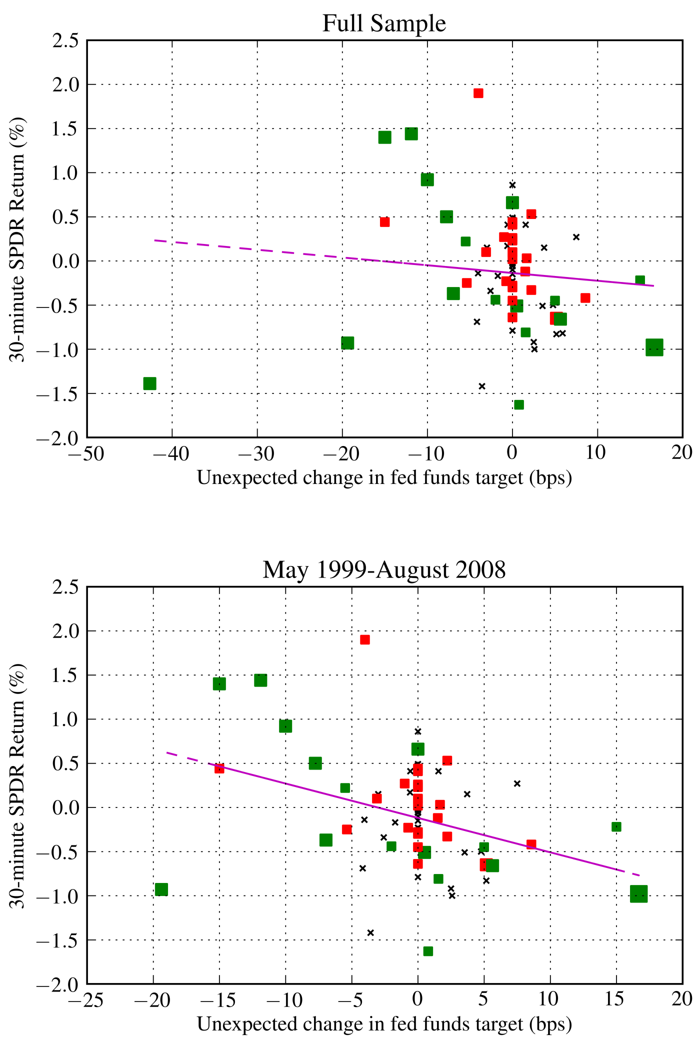


Fig. 2: **SPDR Returns and the Monetary Policy "Surprise" (1999 - 2008):** These two scatterplots graph the 30-minute SPDR response against the unexpected component of monetary policy for the full sample of observations from 1999 to 2008 *(top)* and for the subsample excluding observations from September to December 2008 *(bottom)*. An 'x' denotes a statement in which the Fed announced no change in the target federal funds rate. Rate cuts are shown in green, rate hikes are shown in red, and the size of each square corresponds to the magnitude of the announced rate change for that date.
(*N.B.: The large number of completely anticipated rate hikes shown in the graphs below are observations almost exclusively from the 2004 - 2006 tightening cycle.)*



3.2 The Fed Statement Corpus

3.2.1 "Bag of Words" Representation

The majority of machine learning classification techniques rely upon what is often dubbed the "bag of words" approach to text representation, in which a given document or sentence is transformed into a vector of word frequencies. Consider the two sentences below:

Sentence 1: "The Committee continues to see a risk of heightened inflation pressures."

Sentence 2: "Inflation and inflation expectations remain contained."

The word frequency vectors for the two sentences corresponding to the word dictionary

{"contained", "continues", "expectations", "heightened", "inflation", "pressures", "remain", "risk"}

are as follows:

Sentence 1: (0, 1, 0, 1, 1, 1, 0, 1)

Sentence 2: (1, 0, 1, 0, 2, 0, 1, 0)

Any corpus (a collection of text documents) can therefore be represented as a document-term matrix whose columns correspond to the frequency of a particular word in each document and whose rows indicate the frequency of each word in a particular document.

The example above illustrates unigram text representation, in which each entry in the word frequency vector corresponds to the presence of a single-word string. However, unigram representation is inherently limited because it fails to preserve information about words' order, position in the sentence or text, and relationship with surrounding words. Consider the use of the word "elevated" to describe inflation in two separate FOMC statements:

June 2004: *"Although incoming inflation data are somewhat ELEVATED, a portion of the increase in recent months appears to have been due to transitory factors."*

June 2006: *"Readings on core inflation have been ELEVATED in recent months."*

According to the first sentence, price stability is not a major concern despite what seems to be abnormally high levels of inflation; the second sentence offers no such qualification about high inflation data and, at first glance, might appear to an investor as a more "hawkish" stance by the Fed. Nevertheless, unigram word frequency vectors for both sentences contain an identical value for the term "elevated." Although NLP techniques that use a probablistic approach to classification have been able to achieve remarkably high accuracy without considering the more subtle aspects of sentence structure by training on thousands of documents (Bo Pang and Lillian Lee, 2002), the small size of the Fed statement corpus presents a major obstacle to machine learning.

3.2.2 Sparsity of *N*-Gram Features

In the "bag of words" framework, the simplest way to better capture word sense and preserve the information conveyed in a sentence's structure is to increase the size of textual features. Bigram features, and to a lesser extent, trigram features, are common alternatives to unigram text representation in NLP applications (Dave et al., 2007). Nevertheless, these alternatives still only preserve information contained in consecutive two- or three-word chunks. Further increasing the size of n-grams is not feasible in most NLP applications because of the sparsity of the resulting document-term matrix; the vast majority of large n-grams are unlikely to appear more than once in a collection of documents.

However, extracting features of longer textual chunks is possible for the Fed statement corpus because the percentage of larger n-gram features that occur more than once in the corpus is much higher. What distinguishes Federal Reserve statements from most other text corpora is the similar (and sometimes identical) content of many statements and the repeated usage of particular phrases. For example, the 5-gram "inflation expectations remain well contained" appears in a total of eight Fed statements. As a result, significantly increasing the minimum feature length provides a feasible way to bypass the low informational content of most individual unigrams, bigrams, and trigrams and to capture the information conveyed by more complex syntatical structures underlying central bank language. The following figure shows the greater frequency of higher n-gram features in the Fed statement corpus compared to a sample corpus taken from the R "Text Mining" extension package[6].

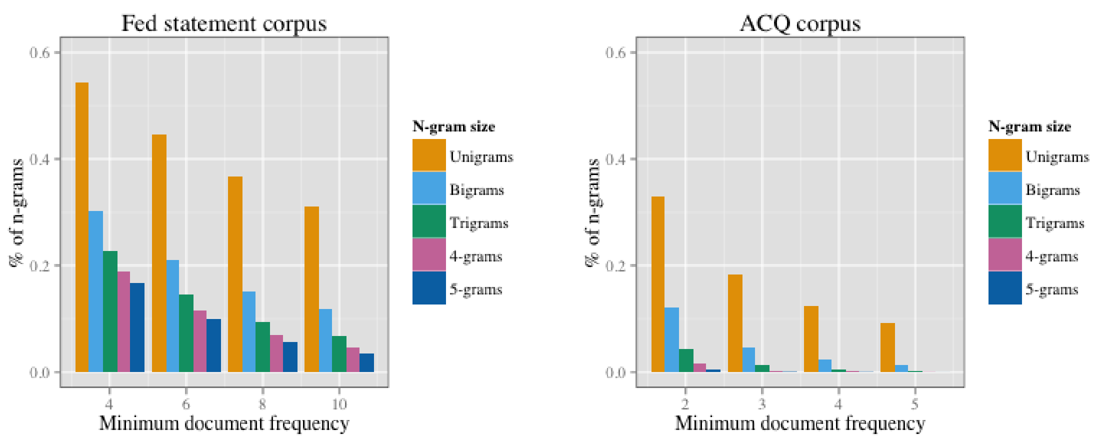


Fig. 3: **Sparsity of the Fed Statement Corpus:** The two figures illustrate the sparsity of the Fed statement corpus *(left)* relative to a baseline corpus *(right)*. The bars indicate the percentage of n-grams of a given size that appear in at least x documents.

[6] The "acq" corpus provided as part of R's Text Mining package contains 50 Reuters news articles on the subject of corporate acquisitions (Ingo Feinerer, 2014).

4 Feature Selection for Text Categorization

Nevertheless, the treatment of each n-gram as a separate variable creates an extremely high dimensional document-term matrix; the Fed statement corpus contains 1204 unigrams, 5131 bigrams, and over 10,000 5-grams. Limiting the text dataset to n-grams that occur at least five times in the corpus still leaves 589 different unigrams, 1251 bigrams, and 1299 5-grams to be considered. Only a small percentage of all possible n-gram features in Fed statements are likely to be meaningful to investors. As a result, machine learning classification techniques commonly used in the NLP field that analyze the informational content of each n-gram (e.g., naive Bayes classifiers and decision trees) are unlikely to be helpful.

In the statistical NLP toolbox, topic models offer one approach to dimension reduction of textual data. In one of the few applications of textual analysis to the study of central bank communications, Ellyn Boukus and Joshua Rosenberg (2006) apply singular value decomposition to the document-term matrix for a collection of Fed announcements to come up with different factors, or "topics," by which to represent each text—a technique known as Latent Semantic Analysis. Another approach to topic modeling known as latent Dirichlet allocation characterizes each text as a collection of underlying or latent topics with a Dirichlet prior (David Blei, Andrew Ng, and Michael Jordan, 2003). However, the topic modeling approach presupposes a structure to documents inside a corpus which may not necessarily be helpful in discerning sentiment. Given that we are interested in individual phrases, rather than topics, the need arises for feature selection techniques to parse through a high-dimensional "bag of n-grams" from the Fed statement corpus.

4.1 Concise Comparative Summarization

Generating subsets of words and phrases that summarize contrasting text corpora via n-gram feature selection—an approach which Jinzhu Jia et al. (2014) call "Concise Comparative Summarization" (CCS)—has a variety of potential applications in the social sciences. Burt Monroe, Michael Colaresi, and Kevin Quinn (2008) evaluate the results of using different regularized word-count methods to characterize speeches by Democrats and Republicans on contentious political topics. Jia et al. (2014) generate n-gram summaries for text corpora of international news articles from the New York Times using four different feature selection methods—co-occurence screening, correlation screening, L1-regularized Logistic Regression, and the LASSO (short for "least absolute shrinkage and selection operator"). Because the authors find that the LASSO yields the best overall results for text summarization[7], I follow that approach to find sources of market sentiment from within the texts of positive and negative Fed statements.

[7] To evaluate the effectiveness of different feature selection techniques, the authors ask human subjects to rate the summaries generated by each method according to their own understanding of a given news article.

In this chapter, I describe the specifications of the LASSO technique used to narrow the set of potential n-gram features to those that are most strongly correlated with financial market sentiment. Then, I introduce an algorithm written by Georgiana Ifrim et al. (2008) and modified by Luke Miratrix that reduces the computational cost of feature selection.

4.2 The Least Absolute Shrinkage and Selection Operator (LASSO)

The LASSO is a variation of the OLS regression that coerces a subset of the least-squares coefficients to zero by imposing an upper bound on the sum of the absolute value of all coefficients (Robert Tibshirani, 1996). From a large set of variables, the LASSO generates a much smaller feature subset of influential variables with non-zero regression coefficients.

Definition 4.2a *(LASSO): Given* $(\mathbf{x_i}, y_i)$*,* $i = 1, 2, ...N$*, where* $\mathbf{x_i} = (x_{i1}, \ldots, x_{ip})^T$*, the set of predictor variables corresponding to observation i, the LASSO estimator* $(\hat{\alpha},\hat{\beta})$*,* $\hat{\beta} = (\hat{\beta}_1, \ldots, \hat{\beta}_p)$ *is determined by:*

$$(\hat{\alpha}, \hat{\beta}) = argmin \left\{ \sum_{i=1}^{N} (y_i - \alpha - \textstyle\sum_j \beta_j x_{ij})^2 \right\} \text{ s. t. } \textstyle\sum_j |\beta_j| \leq t$$

$$\text{or, alternatively : } (\hat{\alpha}, \hat{\beta}) = argmin \left\{ \sum_{i=1}^{N} (y_i - \alpha - \sum_j \beta_j x_{ij})^2 - \lambda \sum_j |\beta_j| \right\}$$

The tuning parameter, λ, determines the extent to which the coefficients of the LASSO estimator are coerced to 0. The intercept of the regression, α, is exempt from "shrinkage."[8]

In the context of text categorization, $\mathbf{x_i}$ is the row of the document-term matrix that corresponds to document i and β_j is the regression coefficient for the column of the document-term matrix that corresponds to term j. The LASSO is used to pick a small number of n-grams by which to characterize texts of opposing sentiment. Because in the context of CCS, texts are categorized into contrasting categories using a binary dependent variable (in this case, positive and negative), Jia et al. (2014) minimize the sum of squared hinge loss, a commonly used function in machine learning for training statistical classifiers.

Definition 4.2b *(LASSO with Squared Hinge Loss): Given* $(\mathbf{x_i}, y_i)$*,* $i = 1, 2, ...N$*, where* y_i *is the ith observation,* $y_i = \{+1, -1\}$*, and* $\mathbf{x_i} = (x_{i1}, \ldots, x_{ip})^T$*, the set of predictor variables corresponding to observation i, the LASSO estimator* $(\hat{\alpha},\hat{\beta})$*,* $\hat{\beta} = (\hat{\beta}_1, \ldots, \hat{\beta}_p)$ *is determined by:*

[8] In Section 1 of the Appendix, I discuss the procedure for choosing an optimal tuning parameter, and in the following chapter, I use high values of λ to test for the significance of the resulting feature subset.

$$(\hat{\alpha}, \hat{\beta}) = argmin \left\{ \sum_{i=1}^{N} H_i(\alpha, \beta)^2 - \lambda \sum_j |\beta_j| \right\}$$

$$Hinge\,loss: \; H_i(\alpha, \beta) = \max \left[0, 1 - y_i \cdot (\alpha + \sum_j \beta_j x_{ij}) \right]$$

When $y_i \cdot (\alpha + \sum_j \beta_j x_{ij}) < 1$, the hinge loss function penalizes the residual of the prediction, and the squared hinge loss penalty is the same as the penalty from the squared-residual in the OLS setup. However, when $y_i \cdot (\alpha + \sum_j \beta_j x_{ij}) > 1$, the hinge loss penalty is zero; there is no penalty for overshooting a prediction in either direction.

4.3 Pre-processing with L2 Normalization

Stop words—functional words such as "the", "a", or "it"—contain little or no informational value, and their inclusion in the dcoument-term matrix generates noise that can to their selection as features (Jia et al., 2014). One approach to reducing the influence of these words is to remove words according to a predefined stop word list, but different corpora may necessitate different stop word lists depending on their content (Monroe et al., 2008). As an alternative to manual, list-based removal, Jia et al. (2014) explore methods of rescaling term frequency vectors in order to reduce the influence of stop words as well as other high-frequency words and phrases. The authors find that for shorter units of texts such as paragraphs, the L2 normalization pre-processing scheme yields summary feature subsets most commonly chosen by humans as consistent with their understanding of particular text corpora.

Definition 4.2 *(L2 Normalization): Given a document-term frequency matrix C where c_{ij} = the number of times term j occurs in document i, X is an L^2 rescaled version of C if:*

$$x_{ij} = \frac{c_{ij}}{\sqrt{z_j}}, \text{ where } z_j = \sum_{i=1}^{N} c_{ij}^2 \qquad \text{(Jia et al., 2014)}$$

The effect of L2 normalization is to penalize features according to their overall frequency in the corpus; words and phrases that are used in almost every document have a high L2 norm, $\sqrt{z_j}$, and thus a lower x_{ij}. Although by restricting the feature space to longer n-grams (as I describe in Chapter 5 and Section 2 of the Appendix), the influence of individual stop words is automatically eliminated, the reweighting of high-frequency "stop phrases" is still desired.

4.4 *N*-gram Search Algorithm

The computation of a solution to n-gram feature selection in this thesis relies on an algorithm originally published by Ifrim et al. (2008) and modified by Miratrix that obtains a sparse regression solution via gradient descent. The adaptation of the n-gram feature selection algorithm for the LASSO developed by Miratrix minimizes the sum of squared hinge loss function $L(\alpha, \beta) = \sum_{i=1}^{N} H_i(\alpha, \beta)^2$ by updating the feature coefficients according to the following rule at each iteration of the gradient descent process:

$$\alpha^{new} = \alpha^{old} - \epsilon \cdot \frac{\partial L\left(\alpha^{old}, \beta^{old}\right)}{\partial \beta_j}$$

$$\beta^{new} = \beta^{old} - \epsilon \cdot \frac{\partial L\left(\alpha^{old}, \beta^{old}\right)}{\partial \beta_j}$$

where $\frac{\partial L\left(\alpha, \beta\right)}{\partial \beta_j}$ is the partial derivative of the sum of squared hinge loss function with respect to the jth n-gram coordinate, defined as:

$$\frac{\partial L}{\partial \beta_j} = -2 \cdot \textstyle\sum_{\{i: y_i \cdot (\alpha + \beta x_i) < 1\}} y_i x_{ij}$$

The n-gram coordinate j used to calculate the gradient $\frac{\partial L\left(\alpha, \beta\right)}{\partial \beta_j}$ for each step is computed by finding the n-gram feature that returns the maximum gradient value.

The search algorithm originally developed by Ifrim et al. (2008) for sparse logistic regression drastically reduces the computational cost of the n-gram search and gradient calculation step by incrementally growing a given n-gram coordinate and pruning large areas of the *(n+1)*-gram subspace that do not improve on its gradient. Instead of computing the full gradient vector for each possible n-gram feature, for each n-gram coordinate of minimum length n^{min}, the algorithm calculates whether extending the sequence can improve the gradient. If so, the sequence is further extended until no more improvement in the gradient is reached; if not, all further subsequences of the original n-gram are removed from the search space, and the gradient of the original n-gram is compared against the best gradient obtained by the algorithm. This process is called repeatedly for each n-gram coordinate of minimum length n^{min} to find the best feature at each iteration.[9] Ifrim et al. (2008) find that this algorithm succeeds in pruning large areas of the feature subspace (up to 90%).

[9] A derivation of the bounds on the gradient that are necessary for pruning, as well as a more explicit formulation of the n-gram search algorithm, can be found in Ifrim et al. (2008).

5 Identifying Market-Moving Language

In this chapter, I report the results of feature selection on the full Fed statement corpus, using the LASSO with a squared hinge loss objective function. I label observations as +1 or -1 according to the direction of intraday SPDR returns and use L2 normalization to pre-process the document-term matrix. I omit from the text sentences containing only interest rate announcements to focus on uncovering sources of market sentiment in the Fed's communication of its policy rationale and "balance of risks" outlook,[10] rather than in its announcement of changes in the target federal funds rate.

As noted previously, the application of feature selection to text categorization in this chapter relies on the work of Professor Luke Miratrix of the Harvard University statistics department; Jinzhu Jia of the Peking University statistics department; Georgiana Ifrim of the Max-Planck Institute for Informatics; and others.

5.1 Parameter Selection

Having described in Chapter 3 the information gain of extracting longer features, I limit the set of possible n-gram features to those with a minimum length of five words. To avoid choosing features highly-correlated with market sentiment, but that appear infrequently in the corpus, I limit the set of possible features to n-grams that occur at least five times. I provide further motivation for the selection of these parameters, as well as examples of feature subsets generated under slightly less restrictive parameters, in Section 2 of the Appendix.

To set the tuning parameter, I follow the procedure described by Tibshirani (1996) that chooses λ to minimize prediction error under k-fold cross-validation.[11] Finally, I cap the number of iterations of the gradient descent algorithm by an arbitrarily large number in order to maximize the number of features selected.[12]

5.2 Positive and Negative *N*-Grams

The following table presents the n-gram features selected by the LASSO model as predictors of positive and negative market sentiment. Figures 4 and 5 plot the frequency of each feature in positive and negative texts over the entire corpus. In the next section, I discuss why the phrases themselves hold little meaning and look to each n-gram's context for evidence of its significance to financial markets.

[10] The "balances of risks" sentence, present in the vast majority of Fed statements, typically concludes the statement with an assessment of whether high inflation or sluggish economic growth poses the greater risk to the economic outlook.

[11] See Section 1 of the Appendix for a detailed description of the procedure for selecting λ.

[12] Ifrim et al. (2008) use cross-validation to pick a number of iterations that minimizes cross-validation error. However, given that the goal of this paper is text exploration rather than prediction accuracy, I run the algorithm until convergence each time to maximize the size of the n-gram feature subset.

Positive Features	Negative Features
longer-term inflation expectations stable inflation	*monetary policy coupled with still-robust*
economic outlook the FOMC also	*for the foreseeable future against*
the FOMC will continue to	*the FOMC perceives that*
and financial developments and will	*in recent months and the*
growth and price stability for	
is likely to be measured	
however investment in non-residential structures is	
and longer-term inflation expectations remain	
the federal funds rate for	

Table 3: **Positive and Negative Features:** Feature subset obtained by minimizing squared hinge loss for the LASSO with $\lambda = 2.5$. Set of all possible features restricted to n-grams containing at least five words and n-grams with a minimum support of five.

The fewer number of n-grams correlated with negative market sentiment, despite the greater number of negative observations, indicates that sources of negative market sentiment are harder to discern from within the text of the corpus than sources of positive sentiment. Given that part of the purpose of the FOMC's post-conference press releases is to address public concerns about the economic outlook, we can reasonably expect the Fed to use language that it suspects will be interpreted favorably by the public; by the same token, we can expect the Fed to avoid the repeated use of language that appears to cause negative sentiment. Even when conveying a more hawkish stance, policymakers may want to soften the blow by framing their position in a particular context. Another possible explanation for this result is that a significant determinant of negative market sentiment is the absence of particular language that Fed-watchers have come to expect. Other aspects of the informational content of FOMC statements less readily captured by the presence of specific words or phrases would remain undetected by an n-gram selection algorithm.

In addition, the n-gram feature subset appears to be skewed towards phrases used in more recent Fed statements (particularly from the financial crisis onwards). This property of the LASSO estimator is consistent with the tendency of post-crisis Federal Reserve statements to contain many of the same phrases and sentences from meeting to meeting. Figure 6 compares the sparsity of n-gram text features for statements written before and after the beginning of quantitative easing, and for statements written during the leadership of different Fed chairmen.

Fig. 4: **Timeline of Positive Feature Frequencies**
This heatmap plots the occurences of each positively-correlated n-gram in positive (blue) and negative (red) statements. Darker bars indicate that a particular feature occurred twice in a particular statement.

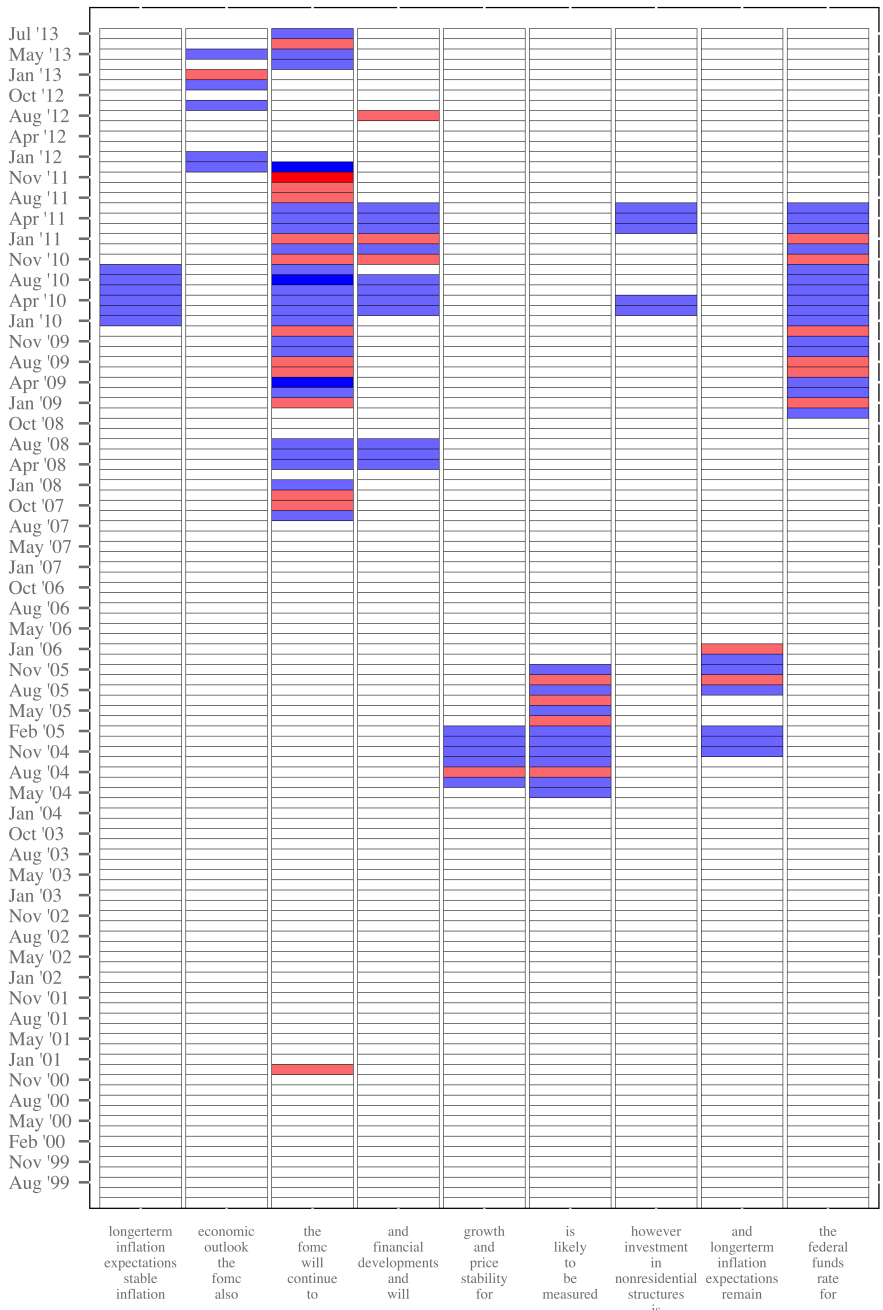


Fig. 5: **Timeline of Negative Feature Frequencies**
This heatmap plots the occurences of negatively-correlated n-grams in positive (blue) and negative (red) statements.

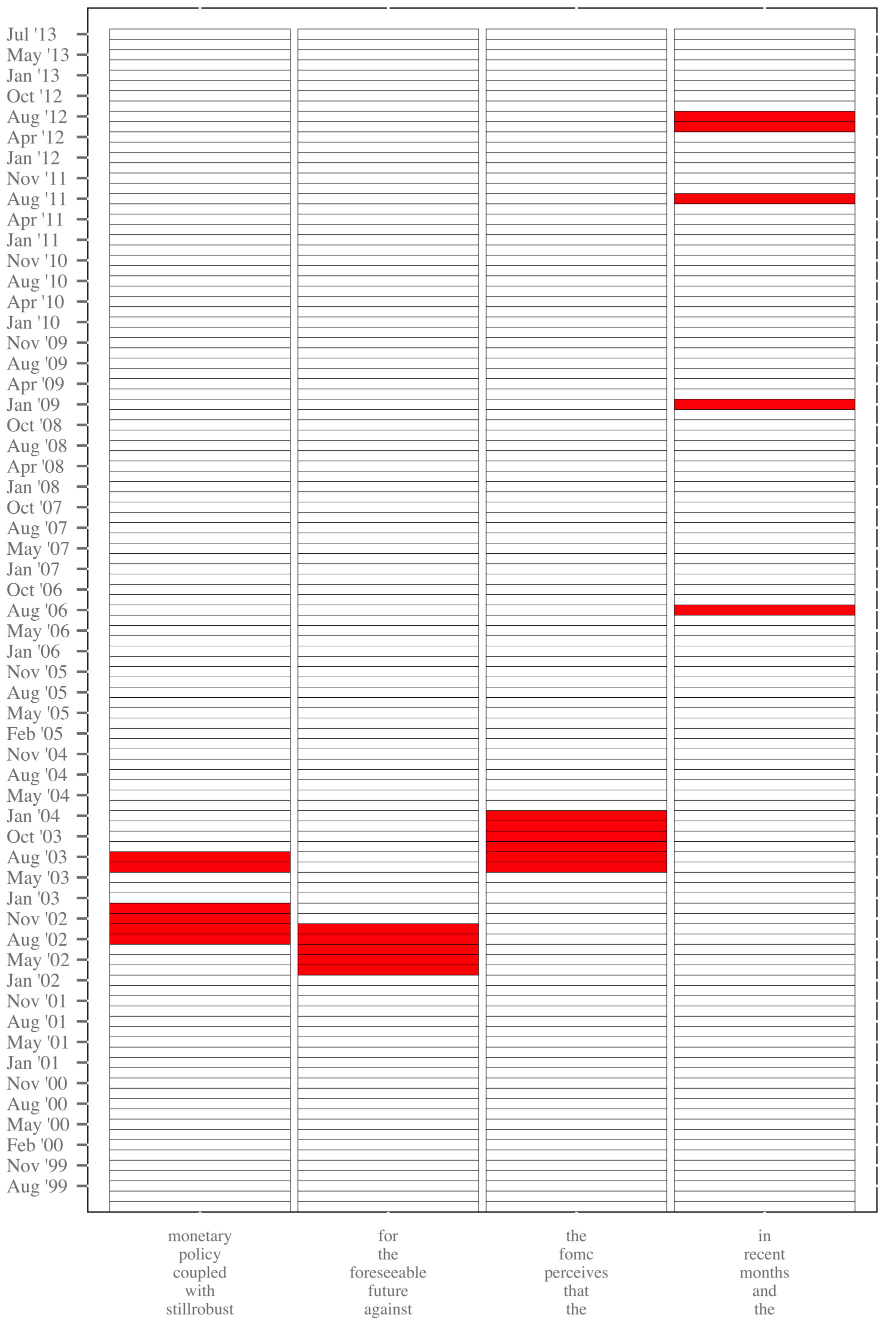
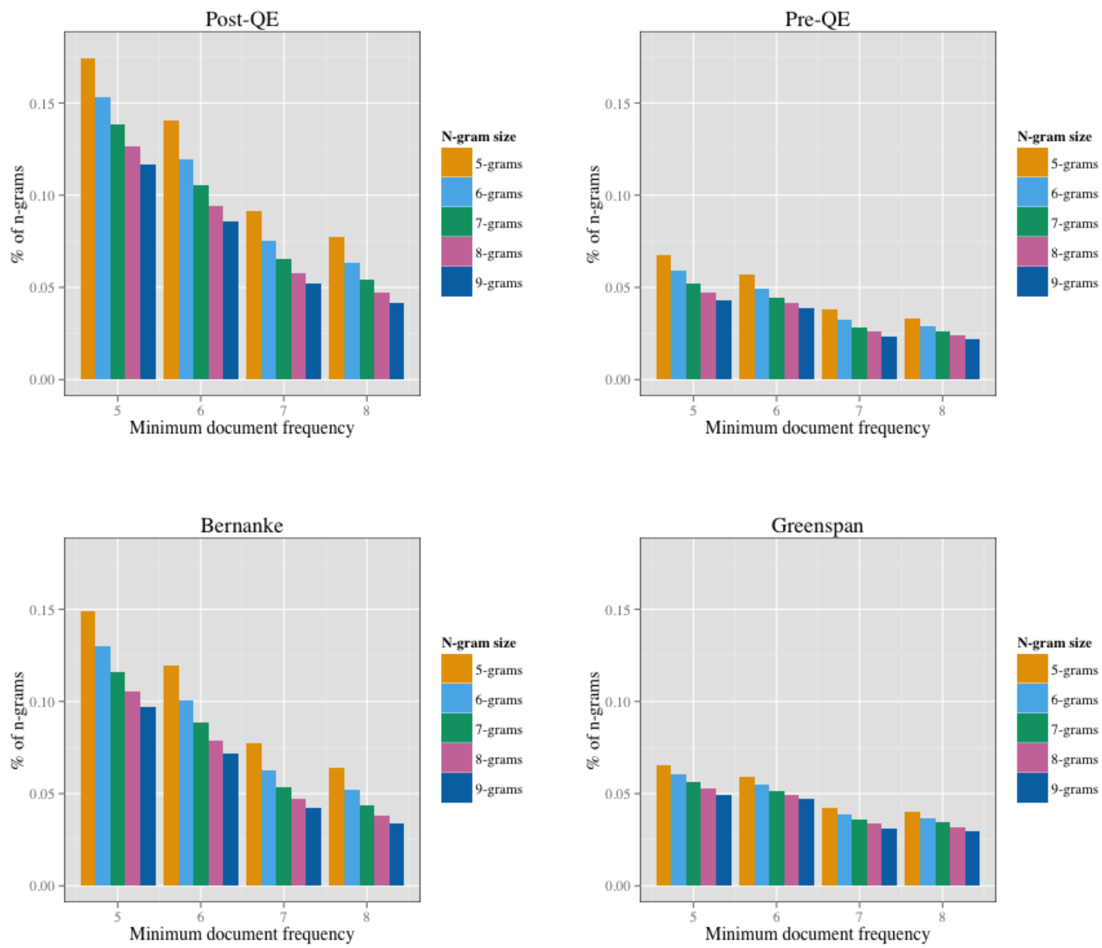


Fig. 6: **Sparsity of the Fed Statement Corpus (2):** The figures below compare the sparsity of different subsamples of the Fed statement corpus. The two figures at the top compare the sparsity of higher n-gram freatures for statements released before and after the beginning of quantitative easing, which began just before the December 2008 meeting. The two figures at the bottom compare the sparsity of statements written under the chairmanship of Alan Greenspan (statements from May 1999 to January 2006) and Ben Bernanke (statements from March 2006 to July 2013).



5.2.1 Interpretation of Features

Although each feature on its own holds little meaning, an inspection of the context of n-grams turns up key phrases from Fed statements. Because the goal of regression in this paper is to use text classification as a tool to identify sources of market sentiment, rather than to predict stock market responses generated by future statements, the lack of interpretability of the n-gram features on their own is not overly concerning, if those features are consistently embedded in potentially influential language.

The difficulty of reading the features themselves is in part due to the machine learner's lack of a human learner's a priori linguistic knowledge. Even if the LASSO estimator is able to detect signals of market sentiment embedded in the Fed's language, it may not be able to locate the precise source of the signal from within the surrounding series of words and phrases. This difficulty is further complicated by the tendency of more interpretable components of positive or negative language to be tweaked from meeting to meeting; as a result, neutral-sounding phrases are selected because they are the only parts of the sentence with that occur a sufficient number of times in the overall corpus.

For example, the phrase "economic outlook the FOMC also" is always mentioned in the context of "downside risks to the economic outlook" and the Fed's belief that inflation will remain low; however, even though the n-gram "economic outlook the FOMC also" appears a total of six times, the wording of its context varies slightly. Consider the following usages of the phrase:

- "Strains in global financial markets continue to pose significant downside risks to the **economic outlook. The FOMC also** anticipates that inflation will settle over coming quarters at levels at or below the FOMC's dual mandate." (December 2011)
- "Strains in global financial markets continue to pose significant downside risks to the **economic outlook. The FOMC also** anticipates that over coming quarters inflation will run at levels at or below the FOMC's dual mandate." (January 2012)
- "The FOMC continues to see downside risks to the **economic outlook. The FOMC also** anticipates that inflation over the medium term likely will run at or below its 2 percent objective." (May 2013)

In this section, I present a brief summary for each feature in the context of present and future monetary policy.[13]

The features correlated with positive market sentiment are:

1. "longer-term inflation expectations stable inflation": Used in an identical context each time to express the view that high inflation is unlikely to be a concern in the future.

[13] A full concordance of n-gram features is included in section 3 of the Appendix.

- "With substantial resource slack continuing to restrain cost pressures and **longer-term inflation expectations stable, inflation** is likely to be subdued for some time." (January 2010 - September 2010)

2. "economic outlook the FOMC also": Used in the context of "downside risks to the economic outlook" and followed by the FOMC stating it does not expect inflation to be a concern in the future.

   - "Strains in global financial markets continue to pose significant downside risks to the **economic outlook. The FOMC also** anticipates that inflation will settle over coming quarters at levels at or below the FOMC's dual mandate." (December 2011)

3. "the FOMC will continue to": Almost always the start of a sentence in which the FOMC pledges to monitor economic conditions and support the economic recovery, particularly in the context of asset purchases since the start of QE.

   - "**The FOMC will continue to** monitor the economic outlook and financial developments and will employ its policy tools as necessary to support the economic recovery and to help ensure that inflation over time is at levels consistent with its mandate." (November 2010 - April 2011)

4. "and financial developments and will": (see feature 3)

   - **"**The FOMC will continue to monitor the economic outlook **and financial developments and will** employ its policy tools as necessary to support the economic recovery and to help ensure that inflation over time is at levels consistent with its mandate." (November 2010 - April 2011)

5. "growth and price stability for": Used in the first six announcements of a monetary tightening cycle beginning in June 2004 to describe equal "upside and downside risks" to the goals of economic growth and price stability.

   - "The FOMC perceives the upside and downside risks to the attainment of both sustainable **growth and price stability for** the next few quarters are roughly equal." (June 2004 - February 2005)

6. "is likely to be measured": Also used in the monetary tightening cycle beginning June 2004 to describe the removal of policy accommodation, given "low" or "contained" levels of underlying inflation.

   - "With inflation low and resource use slack the FOMC believes that policy accommodation can be removed at a pace that **is likely to be measured**. Nonetheless the FOMC will respond to changes in economic prospects as needed to fulfill its obligation to maintain price stability." (June 2004 - November 2005)

7. "however investment in non-residential structures is": Used in 2010-2011 to describe nonresidential real estate investment as "weak" or "still declining."

    - "Household spending and business investment in equipment and software continue to expand. **However, investment in nonresidential structures is** still weak and the housing sector continues to be depressed." (March - June 2011)

8. "and longer-term inflation expectations remain": Used in the 2004-2006 tightening cycle and in post-financial crisis statements to describe inflation expectations as "contained," "stable," or "well-anchored."

    - "Core inflation has stayed relatively low in recent months **and longer-term inflation expectations remain** contained." (December 2005, January 2006)

9. "the federal funds rate for": Used in a total of 21 instances since the establishment of the 0-1/4 percent target federal funds rate range as part of the Fed's pledge to keep rates low.

    - "The FOMC anticipates [continues to anticipate] that weak economic conditions are likely to warrant exceptionally low levels of **the federal funds rate for** some time." (December 2008)

For several of the positively correlated features, there are intuitive explanations for how their usage may impact investor sentiment. In feature 9, for example, a pledge to keep the federal funds rate in its target zero percent range very explicitly conveys a dovish stance by the Fed. If the Fed expects inflation to remain low, a viewpoint expressed in the contexts of features 1, 6, and 8, it is also unlikely to adopt a more hawkish monetary policy.

The contexts for features negatively correlated with market sentiment are as follows:

1. "monetary policy coupled with still-robust": Used in 2000-2001 to describe the Fed's view that it was providing sufficiently accommodative policy to help boost the economy in the context of "still-robust underlying growth in productivity."

    - "The current accommodative stance of **monetary policy coupled with still-robust** underlying growth in productivity should be sufficient to foster an improving business climate." (August 2002)

2. "for the foreseeable future against": Used in 2000-2001 to describe the "balance of risks" as either balanced or "weighted mainly towards conditions that may generate economic weakness."

- "Although the stance of monetary policy is currently accommodative, the FOMC believes that **for the foreseeable future, against** the background of its long-run goals of price stability and sustainable economic growth and of the information currently available, the risks are balanced with respect to the prospects for both goals." (March - August 2002)

3. "the FOMC perceives that the": Used in 2003 to describe the "upside and downside risks" to economic growth as balanced.

    - "**The FOMC perceives that the** upside and downside risks to the attainment of sustainable growth for the next few quarters are roughly equal. In contrast, the probability, though minor, of an unwelcome substantial fall in inflation exceeds that of a pickup [rise] in inflation from its already low level." (June - October 2003)

4. "in recent months and the": Used to describe the state of different of economic indicators (energy prices, employment) "in recent months" leading up to the statement.

    - "Readings on core inflation have been elevated **in recent months and the** high levels of resource utilization and of the prices of energy and other commodities have the potential to sustain inflation pressures." (August 2006)

The contexts of features correlated with negative market sentiment offer less consistency, possibly for the same reasons that explain the smaller size of the negative n-gram subset.

5.3 Significance of N-gram Features

5.3.1 Feature Ranking

We can use the tuning parameter λ to evaluate the relative importance of n-grams selected by the model by determining the minimum value of λ required to generate a coefficient of zero for each feature. Since choosing higher levels of λ has the effect of decreasing the size of the n-gram subset, we can interpret a feature's presence under more heavily penalized models as a measure of its relative importance. Overall, the positive features have larger threshold values of λ than negative features, which is consistent with results so far indicating a stronger positive signal in the Fed statement corpus. The feature ranked as the most important is also the most interpretable n-gram in the feature subset: "longer-term inflation expectations stable inflation." Figure 7 plots the convergence of n-gram feature coefficients to 0 under increasing values of λ.

Positive n-grams	**Threshold λ**
longer-term inflation expectations stable inflation	5.54
and financial developments and will	5.08
however investment in nonresidential structures is	4.44
the FOMC will continue to	4.44
growth and price stability for	4.06
economic outlook the FOMC also	4.00
is likely to be measured	3.84
and longer-term inflation expectations remain	3.52
the federal funds rate for	3.52
Negative n-grams	
the FOMC perceives that	4.04
monetary policy coupled with still-robust	4.04
in recent months and the	3.81
for the foreseeable future against	3.45

Table 4: **Ranking the Feature Subset**

5.3.2 Hypothesis Testing

To test for the statistical significance of our feature subset, I use a permutation test to evaluate the feature subset against the null hypothesis, which is that there is no association between n-grams and market sentiment (i.e. that the feature selection algorithm should produce an empty subset). I find the distribution of λ_0, the minimum tuning parameter required to generate an empty feature subset, by calculating its value under a large number of random permutations of document labels, and then running the regression with a tuning parameter at the 95th percentile of the distribution for λ_0.[14] Because the resulting regression generates an empty feature subset, we are unable to reject the null hypothesis with 95% confidence.

5.4 Evaluating Feature Selection Performance

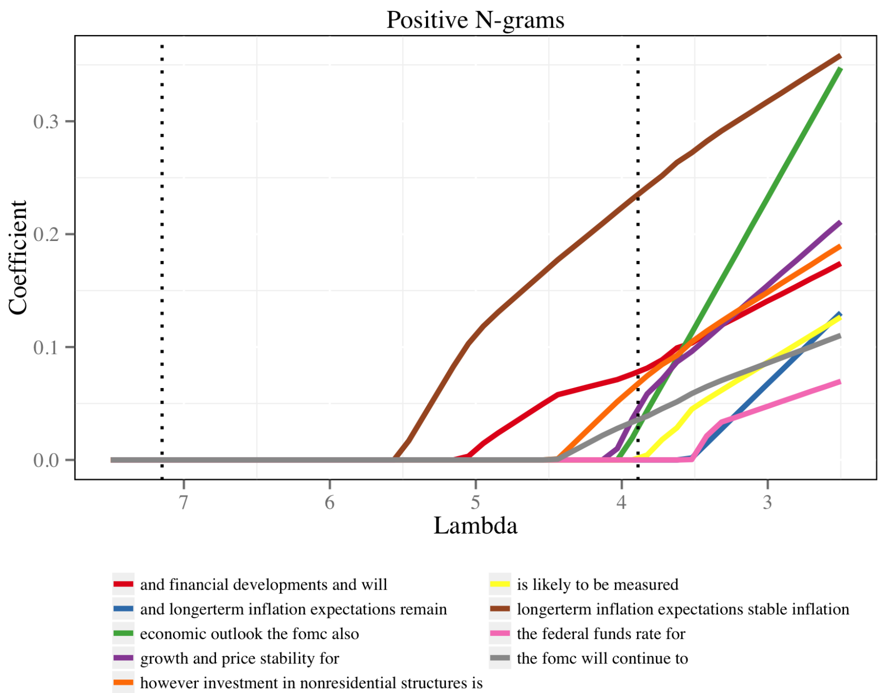
In the previous section, I discussed the extent to which n-gram feature selection can flag key passages from Fed statements based on their associated market sentiment, using longer tex-

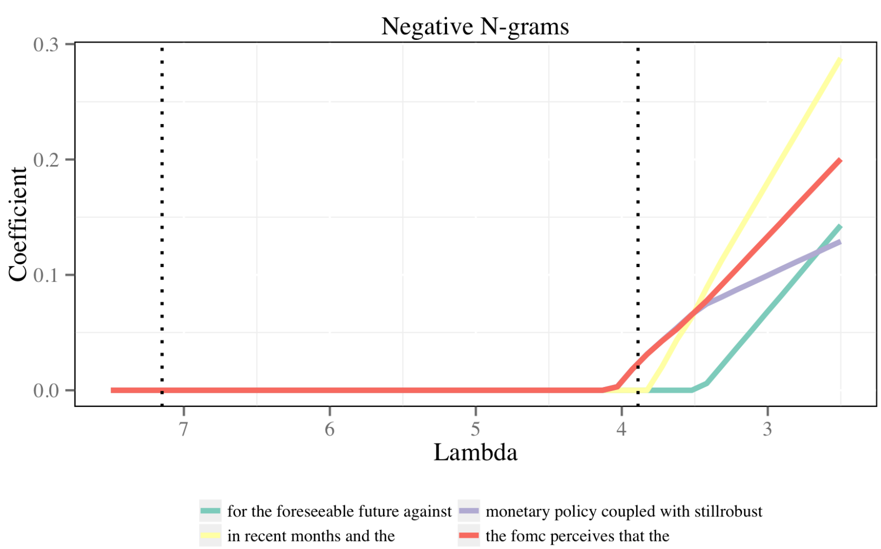
[14] Shuffling the labels of our observations should generate interchangeable datasets under the strong null hypothesis that there is no association between n-grams and market sentiment. The permutation test allows one to create a distribution for the test statistic λ_0 without assuming a normal distribution for the dataset.

tual fragments from documents to capture information contained in more complex semantic structures. Nevertheless, even with longer n-grams, the lack of "readability" in the feature subset generated by a statistical learner reveals the limitations of n-gram feature selection.

Moreover, the lack of significance of the n-gram feature subset indicates a more nuanced relationship between the textual content of Fed statements and market sentiment. Although identical language used from statement to statement allows the the extraction of longer phrases from the Fed corpus, the n-gram feature selection approach does not capture changes in language between consecutive statements that may catch the attention of investors. Furthermore, the response of market participants to Fed announcements is dependent on economic and financial market conditions, which is captured only partially by n-gram representation of the Fed's economic outlook in press releases. These factors, which remain latent variables in this thesis, offer further opportunities for text-based exploration of central bank communications that will become more feasible over the coming years with the inclusion of new observations and the return of "conventional" monetary policy.

Fig. 7: **Regularization Paths of *N*-gram Coefficients:** The two figures below plot the paths of n-gram coefficients under increasing levels of regularization. The region between the dotted lines is the range for the values of λ_0 obtained under 100 random shuffles of the document labels.





6 Conclusion

At the beginning of this thesis, I discussed the need for an unbiased way to measure the content of central bank communications. To that end, I set out to determine how particular language used by the Fed in its post-meeting press releases can generate positive or negative sentiment in equity markets. To do this, I first separately analyze the policy announcement regarding the target federal funds rate change to show how the rate change alone cannot explain the majority of variance in post-release stock returns. Then, treating the text of each Fed statement as a high-dimensional collection of n-grams, I use sparse regression to identify a small number of textual features associated with positive and negative reactions in the S&P 500 index. I find that sparse regression is able to identify frequently used phrases and sentences in Fed statements that are potentially associated with positive sentiment among investors; however, negative sentiment is harder to pin down from within the text of the statement itself. I also find that sparse regression is better at identifying sources of sentiment in statements published within the last half decade.

Although several positively correlated n-grams are contained in language indicating a dovish stance by the Fed, we are unable to establish a statistically significant relationship between the n-gram feature subset obtained using the LASSO and market sentiment. The relatively small number of observations and the difficulties of using n-gram representation to capture meaning in language are two possible reasons for the lack of significance in the results. Perhaps more importantly, however, this particular application of feature selection does not directly take into account changes in language between consecutive statements, nor does it control for the relative strength of macroeconomic and financial indicators around the time of each statement's release—two factors besides the presence of n-grams which are likely to have a significant effect on how texts are interpreted by markets.

However, in trying to study the relationship between Fed communications and market reactions, we should be careful not to treat this outcome as a rejection of the machine learning approach. The integration of NLP techniques used to discern sentence structure[15] with feature selection techniques has the potential to generate more readable and more informative features. In addition, the identification of changes in Fed language can generate a text-based dataset to supplement the "bag of n-grams" data used for feature selection in this paper. Finally, given its relative simplicity and relationship with ordinary least-squares regression, sparse regression offers a theoretical framework on which to develop future statistical methods tailored to the nature of the Fed statement corpus—or any other copus of central bank communications.

[15] "Parsing" and "chunking" (Manning and Schütze, 1999).

References

[1] Ben S. Bernanke and Kenneth N. Kuttner. "What Explains the Stock Market's Reaction to Federal Reserve Policy?". *The Journal of Finance*, 60(3):1221–1257, 2005.

[2] David M. Blei, Andrew Y. Ng, and Michael I. Jordan. "Latent Dirichlet Allocation". *The Journal of Machine Learning Research*, 3:993–1022, March 2003.

[3] Ellyn Boukus and Joshua V. Rosenberg. "The Information Content of FOMC Minutes". Technical report, Federal Reserve Bank of New York, July 2006.

[4] Kushal Dave, Steve Lawrence, and David M. Pennock. "Mining the Peanut Gallery: Opinion Extraction and Semantic Classification of Product Reviews". In *Proceedings of the 12th International Conference on World Wide Web*, WWW '03, pages 519–528, New York, NY, USA, 2003. ACM.

[5] Ingo Feinerer. *"tm - Text Mining Package"*, 2014.

[6] Michel Genereux, Thierry Poibeau, and Moshe Koppel. "Sentiment Analysis Using Automatically Labelled Financial News Items". In Khurshid Ahmad, editor, *Affective Computing and Sentiment Analysis*, volume 45 of *Text, Speech and Language Technology*, pages 101–114. Springer Netherlands, 2011.

[7] Matthew Gentzkow and Jesse M. Shapiro. "What Drives Media Slant? Evidence From U.S. Daily Newspapers". *Econometrica*, 78(1):35–71, 2010.

[8] Refet S. Gurkaynak, Brian Sack, and Eric T. Swanson. "Do Actions Speak Louder Than Words? The Response of Asset Prices to Monetary Policy Actions and Statements". *International Journal of Central Banking*, pages 55–93, 2005.

[9] Georgiana Ifrim, Gokhan Bakir, and Gerhard Weikum. "Fast logistic regression for text categorization with variable-length n-grams". In *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 354–362, 2008.

[10] Jinzhu Jia, Luke W. Miratrix, Bin Yu, Brian Gawalt, Laurent El Ghaoui, Luke Barnesmoore, and Sophie Clavier. "Concise Comparative Summaries (CCS) of Large Text Corpora With a Human Experiment". *Annals of Applied Statistics*, Submitted 2014.

[11] Kenneth N. Kuttner. "Monetary policy surprises and interest rates: Evidence from the Fed funds futures market". *Journal of Monetary Economics*, 47(3):523–544, June 2001.

[12] Roger Lowenstein. "The Education of Ben Bernanke". *The New York Times*, January 20, 2008.

[13] David O. Lucca and Francesco Trebbi. "Measuring Central Bank Communication: An Automated Approach with Application to FOMC Statements". Working Paper 15367, National Bureau of Economic Research, September 2009.

[14] Luke W. Miratrix, Jinzhu Jia, Brian Gawalt, Bin Yu, and Laurent El Ghaoui. "What is in the news on a subject: automatic and sparse summarization of large document corpora". Technical Report 801, UC Berkeley Dept. of Statistics, July 2011.

[15] Burt L. Monroe, Michael P. Colaresi, and Kevin M. Quinn. "Fightin' Words: Lexical Feature Selection and Evaluation for Identifying the Content of Political Conflict". *Political Analysis*, 16(4):372–403, October 2008.

[16] Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan. "Thumbs Up?: Sentiment Classification Using Machine Learning Techniques". In *Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing - Volume 10*, EMNLP '02, pages 79–86. Association for Computational Linguistics, 2002.

[17] Robert Tibshirani. "Regression Shrinkage and Selection via the Lasso". *Journal of the Royal Statistical Society (Series B)*, 58:267 – 288, 1996.

A Appendix

A.1 Tuning parameter selection

I follow the procedure described in Tibshirani (1996) that chooses λ to minimize prediction error under k-fold cross-validation (a method of testing a statistical model on out-of-sample data by partitioning the original dataset k times into a training set on which to fit the model and a test set on which to evaluate its performance). Although the purpose of the LASSO in this context is feature selection rather than the prediction of financial market sentiment, I use cross-validation as a heuristic to generate a λ that balances the amount of penalization in the model.

Given the coefficients of the LASSO estimator obtained using a tuning parameter of λ, $(\hat{\alpha},\hat{\beta})_\lambda$, the prediction error of the model on out-of-sample data defined using average squared hinge loss is:

$$PE(\lambda) = \frac{1}{N}\sum_{i=1}^{N} H_i[(\hat{\alpha},\hat{\beta})_\lambda]^2$$

Under the cross-validation procedure described by Tibshirani (1996), for each cross-validation fold $j = 1, 2, ...K$ I split the complete dataset T into training and test sets, $T - T^j$ and T^j. For each fold, I find the estimator $(\hat{\alpha},\hat{\beta})^j_\lambda$ and choose $\hat{\lambda}$ according to the following:

$$\hat{\lambda} = argmin\left\{\frac{1}{K}\sum_{j=1}^{K} PE^j(\lambda)\right\}$$

where $PE^j(\lambda)$ is the prediction error of the estimator $(\hat{\alpha},\hat{\beta})^j_\lambda$ on the test data T^j.

For the FOMC statement dataset, I use 10-fold cross-validation to search for an optimal λ over the range $\{0, .5, 1, 1.5, ..., \lambda_{max}\}$ where $\lambda_{max} = 6$, the smallest value of the tuning parameter that generates an empty feature subset, $\hat{\beta}_{\lambda_{max}} = 0$, and find that a tuning parameter of $\hat{\lambda} = 2.5$ minimizes cross-validation error.

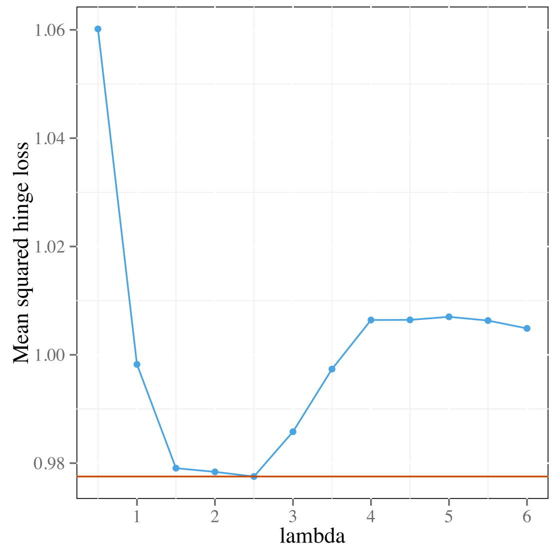


Fig. 8: **Cross-validation Hinge Loss:** The figure above plots the average squared hinge loss obtained under 10-fold cross-validation for different levels of the tuning parameter.

A.2 Relaxing minimum support and minimum n-gram length restrictions

The decision to limit the n-gram space to phrases containing at least 5 words or more and to phrases appearing in at least 5 statements was motivated by considerations discussed in Chapters 3 and 5. However, expanding the range of features under consideration to include all n-grams does not in itself preclude the selection of longer and/or more commonly used n-grams. Here I present two alternate feature subsets generated with different parameters for the n-gram selection algorithm and briefly discuss the impact of restricting the set of n-grams under consideration.

Lowering the minimum support for n-gram features generates a larger but very similar feature subset; all negative features from the original subset are present in their entirety, and seven out of the nine positive features are present either in their entirety or in a sligtly altered form.

Positive Features	Negative Features
*longer-term inflation expectations stable inflation***	*monetary policy coupled with still-robust***
*the FOMC will continue to***	*for the foreseeable future against***
*and financial developments and will***	*the FOMC perceives that***
*however investment in nonresidential structures is***	*in recent months and the***
*is likely to be measured***	*judges that some further policy*
*and longer-term inflation expectations remain well**	*run at or below the*
*financial developments and will act**	*the downside risks to growth*
*federal funds rate for an**	*to moderate over time reflecting*
contained the fomc perceives the	*commodity price increases dissipate further*
the federal reserve will continue	*in productivity is providing important*
in recent months the fomc	
as well as the extent	

Table 5: **Alternate Feature Subset (1):** $\lambda = 2.5$, minimum length = 5, minimum support = 3. (**) indicates an n-gram identical to one from the original feature subset; (*) indicates similarity.

Lowering the minimum length for n-gram features leads to a much more noticeable difference in the feature subset; only one feature from the original subset is present in its entirety. Three-word fragements from five out of the nine positive features are present in the new subset, which indicates that the longer n-grams in our original subset were pruned by the search algorithm because they did not improve on the gradient of the shorter n-gram. As a result, lowering the minimum length for n-gram features generates a larger subset that flags additional phrases from Fed statements.

For example, consider two similar n-grams generated under the original algorithm and under the algorithm with a lower minimum n-gram length:

Original n-gram: "and longer-term inflation expectations remain"

Alternate n-gram: "and longer-term inflation"

Whereas the original n-gram is used to describe inflation expectations as "[remaining] contained" or "[remaining] well-contained," the shorter n-gram flags additional sentences from Fed communications that describe inflation expectations as "stable" or "continue to be anchored."

If the shorter n-gram is used in very different contexts, however, manually preventing the pruning of n-grams with five or fewer words by setting a high minimum n-gram length is desired. Consider the following pair of phrases:

Original n-gram: "is likely to be measured"

Alternate n-gram: "likely to be"

Whereas the original n-gram is used exclusively in the 2004-2006 tightening cycle to describe the removal of policy accommodation, the second n-gram is also used to describe the possible expansion of credit facilities during the financial crisis, or the likelihood of a "subdued" or "moderate" economic recovery in the years following the financial crisis. Since the meaning of each n-gram by itself is very limited, I choose a more restrictive minimum n-gram length in order to generate a feature subset that flags consistent language that can be potentially interpreted in the context of monetary policy.

Neither lowering the minimum support nor lowering the minimum n-gram length generates a statistically significant feature subset with 95% confidence, according to the permutation test procedure described in Chapter 5, Section 4.

Positive Features	Negative Features
*however investment in nonresidential structures is***	*perceives that the**
*and longerterm inflation**	*foreseeable future against**
*outlook the fomc also**	*and the unemployment rate remains*
*will continue to**	*coupled with stillrobust**
*fomc will continue to**	*months and the*
*financial developments and**	*economic growth has*
*likely to be**	*an unwelcome fall*
quarters to be	*financial markets have*
its purchases of	
economic recovery is	
financial markets remain	
over the next	
price stability in	
in the housing	
the first quarter	

Table 6: **Alternate Feature Subset (2):** $\lambda = 2.5$, minimum length = 3, minimum support = 5. (**) indicates an n-gram identical to one from the original feature subset; (*) indicates similarity.

A.3 *N*-gram feature concordance

A.3.1 Positive *n*-grams:

"longer-term inflation expectations stable inflation"

- "With substantial resource slack continuing to restrain cost pressures and **longer-term inflation expectations stable, inflation** is likely to be subdued for some time." (January 2010 - September 2010) "economic outlook the FOMC also"

- "Strains in global financial markets continue to pose significant downside risks to the **economic outlook. The FOMC also** anticipates that inflation will settle over coming quarters at levels at or below the FOMC's dual mandate." (December 2011)
- "Strains in global financial markets continue to pose significant downside risks to the **economic outlook. The FOMC also** anticipates that over coming quarters inflation will run at levels at or below the FOMC's dual mandate." (January 2012)
- "Furthermore, strains in global financial markets continue to pose significant downside risks to the **economic outlook. The FOMC also** anticipates that inflation over the medium term likely would run at or below its 2 percent objective." (September 2012)
- "Furthermore, strains in global financial markets continue to pose significant downside risks to the **economic outlook. The FOMC also** anticipates that inflation over the medium term likely will run at or below its 2 percent objective." (December 2012)
- "Although strains in global financial markets have eased somewhat, the FOMC continues to see downside risks to the **economic outlook. The FOMC also** anticipates that inflation over the medium term likely will run at or below its 2 percent objective." (January 2013)
- "The FOMC continues to see downside risks to the **economic outlook. The FOMC also** anticipates that inflation over the medium term likely will run at or below its 2 percent objective." (May 2013)

"the FOMC will cotinue to"

- "**The FOMC will continue to** monitor closely the evolving economic situation against the background of its long-run goals of price stability and sustainable economic growth and of the information currently available." (December 2000)
- "**The FOMC will continue to** assess the effects of these and other developments on economic prospects and will act as needed to foster price stability and sustainable economic growth." (September 2007)
- "**The FOMC will continue to** monitor economic and financial developments and will act as needed to foster sustainable economic growth and price stability." (April - August 2008)
- "**The FOMC will continue to** assess the effects of financial and other developments on economic prospects and will act as needed to foster price stability and sustainable economic growth." (October, December 2007)

- "**The FOMC will continue to** monitor carefully the size and composition of the Federal reserve's balance sheet in light of evolving market developments . . . " (January - April 2009)
- "**The FOMC will continue to** evaluate the timing and overall amounts of its purchases of securities in light of the evolving economic outlook and conditions in financial markets." (April - November 2009)
- "**The FOMC will continue to** evaluate its purchases of securities in light of the evolving economic outlook and conditions in financial markets." (January 2010)
- "**The FOMC will continue to** monitor the economic outlook and financial developments and will employ its policy tools as necessary to promote economic recovery and price stability in light of improved functioning of financial markets." (March, April 2010)
- "**The FOMC will continue to** monitor the economic outlook and financial developments and will employ its policy tools as necessary to promote economic recovery and price stability." (June, August 2010)
- "**The FOMC will continue to** roll over the Federal Reserve's holdings of of Treasury securities as they mature." (August 2010)
- "**The FOMC will continue to** monitor the economic outlook and financial developments and is prepared to provide additional accommodation if needed to support the economic recovery and to return inflation over time to levels consistent with its mandate." (September 2010)
- "**The FOMC will continue to** monitor the economic outlook and financial developments and will employ its policy tools as necessary to support the economic recovery and to help ensure that inflation over time is at levels consistent with its mandate." (November 2010 - April 2011)
- "**The FOMC will continue to** pay close attention to the evolution of inflation and inflation expectations to promote the ongoing economic recovery and to help ensure that inflation over time is at levels consistent with its mandate." (June, August 2011)
- "**The FOMC will continue to** pay close attention to the evolution of inflation and inflation expectations to support a stronger economic recovery and to help ensure that inflation over time is at levels consistent with the dual mandate." (September - December 2011)
- "**The FOMC will continue to** assess the economic outlook in light of incoming information and is prepared to employ its tools to promote a stronger economic recovery in a context of price stability." (November, December 2011)

- “**The FOMC will continue to** take appropriate account of the likely efficacy and costs of such purchases as well as the extent of progress toward its economic objectives to support continued progress toward maximum employment and price stability. (March - July 2013)

“and financial developments and will”

- “The FOMC will continue to monitor economic **and financial developments and will** and act as needed to promote sustainable economic growth and price stability.” (April - August 2008)
- “The FOMC will continue to monitor economic **and financial developments and will** employ its policy tools as necessary to promote economic recovery and price stability in light of improved functioning of financial markets.” (March, April 2010)
- “The FOMC will monitor the economic outlook **and financial developments and will** act as needed to best foster maximum employment and price stability.” (June 2010, June 2011)
- “The FOMC will continue to monitor the economic outlook **and financial developments and will** employ its policy tools as necessary to promote economic recovery and price stability.” (August 2010)
- “The FOMC will continue to monitor the economic outlook **and financial developments and will** employ its policy tools as necessary to support the economic recovery and to help ensure that inflation over time is at levels consistent with its mandate.” (November 2010 - April 2011)
- “The FOMC will closely monitor incoming information on economic **and financial developments and will** provide additional accommodation as needed to promote a stronger economic recovery and sustained improvement in labor market conditions in a context of price stability.” (August 2012)

“growth and price stability for”

- “The FOMC perceives the upside and downside risks to the attainment of both sustainable **growth and price stability for** the next few quarters are roughly equal.” (June 2004 - February 2005)

“is likely to be measured”

- “With inflation low and resource use slack the FOMC believes that policy accommodation can be removed at a pace that **is likely to be measured**.” (May 2004)

- “With inflation low and resource use slack the FOMC believes that policy accommodation can be removed at a pace that **is likely to be measured**. Nonetheless the FOMC will respond to changes in economic prospects as needed to fulfill its obligation to maintain price stability.” (June 2004 - November 2011)

“however investment in nonresidential structures is”

- “Business spending on equipment and software has risen significantly**. However, investment in nonresidential structures is** declining, housing starts have been flat at a depressed level, and employers remain reluctant to add to payrolls.” (March 2010)

- “Business spending on equipment and software has risen significantly**. However, investment in nonresidential structures is** declining and employers remain reluctant to add to payrolls. Housing starts have edged up but remain at a depressed level.” (April 2010)

- “Household spending and business investment in equipment and software continue to expand. **However, investment in nonresidential structures is** still weak and the housing sector continues to be depressed.” (March - June 2011)

“and longer-term inflation expectations remain”

- “Output appears to be growing at a moderate pace despite the rise in energy prices and labor market conditions have improved [continue to improve gradually]. Inflation **and longer-term inflation expectations remain** well contained.” (November 2004 - Feburary 2005)

- “Core inflation has been relatively low in recent months **and longer-term inflation expectations remain** well contained but pressures on inflation have stayed elevated.” (August 2005)

- “Higher energy and other costs have the potential to add to inflation pressures. However, core inflation has been relatively low in recent months **and longer-term inflation expectations remain** contained.” (September 2005)

- “The cumulative rise in energy and other costs has the potential to add to inflation pressures. However, core inflation has been relatively low in recent months **and longer-term inflation expectations remain** contained.” (November 2005)

- “Core inflation has stayed relatively low in recent months **and longer-term inflation expectations remain** contained.” (December 2005, January 2006)

“the federal funds rate for”

- "The FOMC anticipates [continues to anticipate] that weak economic conditions are likely to warrant exceptionally low levels of **the federal funds rate for** some time." (December 2008, January 2009)

- "The FOMC will maintain the target range for the federal funds rate at 0 to 1/4 percent and anticipates [continues to anticipate] that economic conditions are likely to warrant exceptionally low levels of **the federal funds rate for** an extended period." (March 2009 - September 2009)

- "The FOMC will maintain the target range for the federal funds rate at 0 to 1/4 percent and continues to anticipate that economic conditions, including low rates of resource utilization, subdued inflation trends, and stable inflation expectations, are likely to warrant exceptionally low levels of **the federal funds rate for** an extended period." (November 2009 - April 2011)

- "The FOMC continues to anticipate that economic conditions, including low rates of resource utilization and a subdued outlook for inflation over the medium run, are likely to warrant exceptionally low levels for **the federal funds rate for** an extended period." (June 2011)

A.3.2 Negative n-grams:

"monetary policy coupled with still-robust"

- "The current accommodative stance of **monetary policy coupled with still-robust** underlying growth in productivity should be sufficient to foster an improving business climate [over time]." (August, September 2002)

- "The FOMC continues to believe that this [an] accommodative stance of **monetary policy coupled with still-robust** underlying growth in productivity is providing important ongoing support to economic activity." (November, December 2002; June, August 2003)

"for the foreseeable future against"

- "Although the stance of monetary policy is currently accommodative, the FOMC believes that **for the foreseeable future, against** the background of its long-run goals of price stability and sustainable economic growth and of the information currently available, the risks are balanced with respect to the prospects for both goals." (March - August 2002)

- "However, considerable uncertainty persists about the extent and timing of the expected pickup in production and employment, owing in part to the emergence of heightened geopolitical risks. Consequently, the FOMC believes that **for the foreseeable future, against** the background of its long-run goals of price stability and sustainable economic growth and of the information currently available, the risks are weighted mainly toward conditions that may generate economic weakness." (September 2002)

"the fomc perceives that the"

- "**The FOMC perceives that the** upside and downside risks to the attainment of sustainable growth for the next few quarters are roughly equal. In contrast, the probability, though minor, of an unwelcome substantial fall in inflation exceeds that of a pickup [rise] in inflation from its already low level." (June - October 2003)
- "**The FOMC perceives that the** upside and downside risks to the attainment of sustainable growth for the next few quarters are roughly equal. The probability of an unwelcome fall in inflation has diminished in recent months and now appears almost equal to that of a rise in inflation." (December 2003, January 2004)

"in recent months and the"

- "Readings on core inflation have been elevated **in recent months and the** high levels of resource utilization and of the prices of energy and other commodities have the potential to sustain inflation pressures." (August 2006)
- "In light of the declines in the prices of energy and other commodities **in recent months and the** prospects for considerable economic slack, the FOMC expects that inflation pressures will remain subdued in coming quarters." (January 2009)
- "Indicators suggest a deterioration in overall labor market conditions **in recent months and the** unemployment rate has moved up." (August 2011)
- "However, growth in employment has slowed **in recent months and the** unemployment rate remains elevated." (June 2012)
- "Gowth in employment has been slow **in recent months and the** unemployment rate remains elevated." (August 2012)