# Narratives in Finance

Master's Thesis

Submitted in partial fulfillment of the requirements for the degree of
Master of Arts in Quantiative Economics and Finance

Author
Corinne Knöpfel
Wiesen 2488, Herisau
11-613-676
corinne.knoepfel@student.unisg.ch

Supervisor
Dr. Peter Gruber
Institute of Finance
Università della Svizzera Italiana

Sunday 24th March, 2019

# Abstract

# Contents

# List of Figures

# List of Tables

Chapter 1

# Introduction

# Chapter 2

# Monetary Policy and Interest Rates

However little understood, the relationship between monetary policy and market interest rates is undeniable. Interest rates of all maturities react to changes in monetary policy, creating opportunities and risks for traders, challenges for policy makers, and puzzling effects for academics to study (Ellingsen & Söderström, 2001, p. 1594).

Target rate changes in particular have an impact on the bond market and on interest rates (Cook & Hahn, 1989, p. 332). Yet, the understanding of yield curve movements is incomplete at best. On average, the relationship between monetary policy and interest rates appears to be positive: An increase in the central bank's target rate leads to an increase in the interest rates of all maturities. However, there are many instances where this simple rule has proven false and interest rates of long maturities fell in response to an increase in the central bank's rate (Ellingsen & Söderström, 2001, p. 1594).

Chapter 2.1 gives an account of the puzzle posed by the inconsistent response of long-term rates, Chapter 2.2 touches on previous research and possible explanations, and Chapter 2.3 outlines how an investigation of narratives might be able to shed light on this puzzle.

## 2.1 Excess Sensitivity Puzzle

Cook and Hahn (1989) analyzed financial data from the late 70s and found that the U.S. Federal Reserve (Fed), by setting the target for the federal funds rate, had a strong influence on interest rate movements. While short-term rates reacted particularly strongly, changes in the target rate also caused small but significant movements in long-term rates.

It is not surprising that short-term rates follow the target rate closely, after all the Fed keeps the overnight rate close to the target and thus directly influences the one-month rate (Ellingsen et al., 2003, p. 1). The movements of the long-term rates are more ambiguous. Cook and Hahn (1989, pp. 343–346) interpret the fact that, on average, 10-year and 30-year bonds co-move with the short-term rates as evidence for the expectation theory of the term structure of interest rates. According to the expectation theory, long-term rates are equal to short-term rates over the same period of time plus a term premium. Thus, an increase in the short-term rates is expected to drive up long-term rates as well, but to a lesser extent (Ellingsen & Söderström, 2001, p. 1594).

To Romer and Romer (2000), on the other hand, the response of long-term rates presents a puzzle. They argue that standard theory predicts a drop in inflation as short-term rates rise, which ought to lead to a reduction in long-term rates. The opposite can be observed, however: Interest rates for all maturities typically rise following an increase in the target rate. Romer and Romer (2000) explain this anomaly with information-asymmetry between the Fed and the general public. They find evidence that the Fed is in possession of private information, which it reveals to other market participants through its monetary policy. In response, market participants adjust their inflation expectations upwards, causing long-term rates to rise.

Dissecting the interest rate response in more detail led Skinner and Zettelmeyer (1995) to paint an even complexer picture. While the yield curve shifts upwards on average, they found a number of occasions where an adjustment to the target rate caused the yield curve to tilt: Long and short rates responded by moving in opposite directions (as cited in Ellingsen et al., 2003, p. 1). Skinner and Zettelmeyer came to the conclusion that these were not singular occurrences, but that such tilts made up a considerable portion of the yield curve responses and could be observed in all four of the big economies they studied, that is in France, Germany, the United Kingdom, and the United States (as cited in Ellingsen & Söderström, 2001, p. 1594). An example is the yield curve movement in 1994, where interest rates of long maturities fell after the Fed announced an increase in its target rate (Ellingsen & Söderström, 2001, p. 1594). So not only is the positive response of long-term rates difficult to explain, the response is not even consistent in its direction: long-term rates may move up or down when the Fed increases the target rate.

Whether positive or negative, to Gürkaynak, Sack, and Swanson (2005b, p. 425) any response of long-term rates is in contradiction to standard macroeconomic models. They argue that models predict that short-term rates return quickly to their steady state and thus have only a transitory effect on the future path of interest rates. Therefore, one would expect long-term rates not to react to monetary policy changes. They refer to the fact that long-term rates move significantly in response to monetary policy decisions as *excess sensitivity* of long-term interest rates (Gürkaynak, Sack, & Swanson, 2003, p. 2).

Gürkaynak et al. (2005b, pp. 426–427) focus on the response of forward interest rates as a different way of expressing the yield curve. They find that long-term forward rates move in the opposite direction as the monetary policy actions. As they note, this stands in sharp contrast to the findings of Cook and Hahn (1989) and Romer and Romer (2000), who observed a movement of long-term rates in the same direction. Gürkaynak et al. put this down to their use of forward rates, which they consider a better measure for sensitivity. Additionally, they criticize previous research for the usage of raw change in the target rate, neglecting to differentiate between expected and unexpected policy moves. In their opinion, only the unexpected components of a monetary policy action can be expected to influence the term structure (Gürkaynak et al., 2005b, pp. 430–431).

Since Gürkaynak et al. observe a negative response of the long-term forward rates, they suggest that such a response is not an anomaly but has a very natural explanation. Standard macroeconomic models assume that long-run levels of inflations and real interest rates are relatively fixed and known by all market participants (Gürkaynak et al., 2005b, p. 425). Gürkaynak et al. argue that models might be misspecified and long-run inflation expectations are not as perfectly anchored as assumed. They see the most plausible explanation for the observed term structure movements in the fact that monetary policy surprises lead market participants to adjust their expectations of the long-run level of inflation (Gürkaynak et al., 2005b, pp. 434–435).

Even though Gürkaynak et al. (2005b) are able to account for the negative response of long-term forward rates to an increase in the target rate, Ellingsen and Söderström (2004, p. 2) maintain that their model is unable to explain the positive response of long-term yields observed by other researchers. Thus, Gürkaynak et al. (2005b) fail to solve the puzzle as to why the yield curve shifts on one occasion but tilts at another when provoked by apparently identical monetary policy actions. Ellingsen et al. address this shortcoming in their own theoretical model (2001) and provide empirical support for their hypotheses (2003).

## 2.2 Existing Research and Explanations

Ellingsen and Söderström (2001) use a simple dynamic macroeconomic model where shocks to output and inflation exhibit some persistence and monetary policy actions have a lagged effect on output and inflation. The central bank is assumed to minimize deviations of inflation and output from their long-run averages, while market participants form rational expectations concerning the future target and short rates. On the basis of this model, Ellingsen and Söderström (2001, p. 1599–1602) make several predictions:

- *Proposition 1*: If there is symmetric information, economic shocks are observed by all market participants and affect interest rates directly. Policy actions by the central bank reveal no new information and thus will not affect the term structure of interest rates.
- *Proposition 2*: If the central bank has private information about shocks to supply or demand, market participants will infer this information from the central bank's policy actions. Thus, the yield curve of market interest rates will respond by moving in the same direction as the target rate change.
- *Proposition 3*: If the central bank has private information about changes in its own inflation preferences, market participants will infer these changes by observing the central bank's reaction to an economic shock. Consequently, they will adjust their expectations about future interest rate targets. This causes the yield curve to tilt as long rates move in the opposite direction as the target rate change.

Thus, the yield curve moves for two reasons: either the Fed reacts to new, possibly private information about the economy (what Ellingsen and Söderström call *endogenous*, outlined in proposition 2), or the Fed's policy preferences change (what Ellingsen and Söderström call *exogenous*, outlined in proposition 3). They predict that interest rates of all maturities move in the same direction after an endogenous policy action, but that short and long-rates move in opposite directions after an exogenous change (2001, p. 1594–1595).

In a second paper, Ellingsen et al. (2003) analyze empirical data to find evidence for their model. In order to determine whether a policy action is endogenous or exogenous, they analyze reports on U.S policy in the *Credit Market* column of the *Wall Street Journal*. This text basis is supposed to capture the traders' opinions to a policy move and not the central bank's intention behind it, as it is the traders' opinions that move the bond prices (Ellingsen et al., 2003, p. 2). Ellingsen et al. used the articles on the day of the Fed move, as well as on the day before and the day after. They found publications on the days following a policy action to be the most informative (2003, p. 8).

They estimate the following regression (Ellingsen et al., 2003, p. 13):

$$\Delta i_t^n = \alpha + (\beta_n^{NP} d_t^{NP} + \beta_n^{Ex} d_t^{Ex} + \beta_n^{End} d_t^{End}) \Delta i_t^{3m} + v_t^n, \tag{2.1}$$

where $\Delta i_t^n$ is the change in the interest rate of maturity $n$ on day $t$; $d_t^{NP}$, $d_t^{Ex}$, and $d_t^{End}$ are dummies for non-policy, exogenous policy, and endogenous policy days respectively; and $\Delta i_t^{3m}$ is the change in the 3-month treasury bill rate on day $t$.

The one-day change in the 3-month T-bill rate is taken as a measure of unexpected monetary policy action (regressor in eq.2.1). Ellingsen et al. (2003, p. 13) argue that the 3-month rate is sufficiently short to be determined by policy actions, but also sufficiently long to avoid noise from expectation errors. If the target rate remains unchanged, that is on non-policy days, the change in the 3-month rate measures the adjustment of expectations about future monetary policy actions provoked by the day's new information. If the target rate changes, that is on policy days, any change in the 3-month rate is interpreted as the unexpected component of the policy action (Ellingsen et al., 2003, p. 12).

The main hypothesis of Ellingsen et al.'s model is that long-term interest rates respond positively to endogenous policies and negatively to exogenous policies:

$$H_0: \text{ For large } n\text{: } \beta_n^{Ex} < 0 < \beta_n^{End} \tag{2.2}$$

Using data from October 1988 to December 2001, Ellingsen et al. (2003, p. 16) find significant positive responses of the the 6-month and 1-year rate to endogenous and exogenous policy actions. For the 10-year and the 30-year rate, on the other hand, the coefficients are significant and positive for endogenous changes, and negative for exogenous changes. Ellingsen et al. conclude that their model finds strong support in U.S. data.

Yet, the author of this thesis cannot help but note that the explained variation ($R^2$) is rather small for long rates. While the model is able to account for up to 60% of the variation in short rates, this ratio drops to 15% for 10-year rates and 10% for 30-year rates (Ellingsen et al., 2003, p. 16). Additionally, Ellingsen et al. (2003, p. 20) admit that their results might be dependent on the classification of a few pivotal events. Since the classification was done manually, it is quite subjective. This could explain why von Krosigk (2017) was not able to replicate their results using text mining techniques. Von Krosigk analyzed data for the time period of January 2002 to June 2017 and found only positive coefficients, especially for exogenous events, with the only significant effect pertaining to the 6-month rate (2017, p. 36). This stands in sharp contrast to Ellingsen et al.'s results and raises doubts concerning the robustness of their findings.

## 2.3 New Insights Through Narrative Research

It is striking that Ellingsen et al. (2003), in order to find data in support of their model, naturally chose a narrative approach. In their model, they explicitly assume that the yield curve's response depends "on market participants' interpretation of the policy move" (Ellingsen & Söderström, 2001, p. 1603). They aim at classifying policy events as they are perceived by financial investors "since it is the investors' beliefs that determine the interest rate response" Ellingsen and Söderström (2001, p. 1604). They analyze newspaper articles not to determine the central bank's intentions underlying a policy move, but rather the traders' opinions. In their view it is "irrelevant whether a target change is in fact driven by policy preferences or by economic events. At any given point in time it is traders, and not the Fed, that determine the price of long-term bonds" (Ellingsen et al., 2003, p. 2). Thus, the effect on market interest rates is not driven by policy actions, but by the opinions and views market participants form about such actions. In other words, it's not the target rate change that influences the yield curve, but the stories surrounding it.

Likewise, Cook and Hahn (1989) used newspaper articles to analyze the reaction of the yield curve to target rate movements. They focused on perceived changes in the target rate as reported by the Wall Street Journal on the day after a target rate change to determine its magnitude and direction. Interestingly, Cook and Hahn (1989, p. 337) mention that the Journal sometimes used "speculative language" which hampered their ability to isolate the bare facts of the policy action. In their quest to determine the facts of the policy move, they did their best to strip the articles of all other information, including the manner in which the facts were presented and the interpretative value of the "speculative language."

Gürkaynak, Sack, and Swanson (2005a, pp. 86–87) drive the point home by saying that "previous studies estimating the effects of changes in the federal funds rate on bond yields [...] have been missing most of the story." Their research revealed that reactions on the financial market were at least partially driven by the statements accompanying a policy action. Announcements of the FOMC, the Federal Open Market Commitee of the Fed which regulates the funds rate target, account for at least three quarters of the variation in the movement of longer term Treasury yields around a FOMC meeting.

Similarly, Goetzmann, Kim, and Shiller (2016) support the view that market participants are highly influenced by narrative statements, especially by the financial press. A survey over a 26 year period revealed that investors generally hold an exaggerated assessment of the risk of a stock market crash and that their assessments were influenced by the news stories, in particular the front page stories, they have read. According to Goetzmann et al. (2016), newspaper articles make market returns, especially negative developments, more salient and thus influence investor behavior. Other researchers, such as Engelberg and Parsons (2011), Kräussl and Mirgorodskaya (2014), and Yuan (2015), support the view that the financial press plays an important role in focusing investor attention and thus influences their behavior.

Consequently, the author of this thesis hypothesizes that it is the interpretation of a policy event by the market participants, that is the narrative surrounding a target rate change, that determines the response of the financial markets and thus the movement of the long-term interest rate. Even though Ellingsen et al. (2003) employ a narrative approach, it remains closely tied to a macroeconomic model and only allows for certain predetermined aspects of a potentially much bigger narrative. Thus, it stands to reason that opening the focus of the analysis to include any type of narrative that could potentially influence a market participant's action will yield more robust results. To this end, the author proposes the following model:

$$\Delta i_t^n = \alpha + (\beta_n^{NP} d_t^{NP} + \beta_n^{N_1} d_t^{N_1} + \beta_n^{N_2} d_t^{N_2}) \Delta i_t^{3m} + v_t^n, \tag{2.3}$$

where $\Delta i_t^n$ is the change in the interest rate of maturity $n$ on day $t$; $d_t^{NP}$ is a dummy for non-policy days, $d_t^{N_1}$ and $d_t^{N_2}$ are dummies for policy days that were classified as being dominated by either narrative one ($N_1$) or narrative two ($N_2$); and $\Delta i_t^{3m}$ is the change in the 3-month treasury bill rate on day $t$.

Ideally, an examination of newspaper articles with regards to narratives surrounding a target rate change will allow the identification of two distinct narratives that are able to account for the inconsistent reaction of the long-term rates. Thus, the main hypothesis stipulates that narrative one leads to a negative reaction of the long-term rate while narrative two provokes a positive reaction:

$$H_0 : \text{ For large } n\text{: } \beta_n^{N1} < 0 < \beta_n^{N2} \tag{2.4}$$

Chapter 3 outlines what narratives are and why there is reason to believe that they have a strong influence on human behavior and thus warrant more attention in financial and economic research. To circumvent the problem of subjectivity faced by previous research when it comes to the identification of narratives, this thesis uses Natural Language Processing techniques rather than manual evaluation of text data. Chapter 4 gives an account of topic modeling methods and introduces Probabilistic Latent Semantic Analysis, which will be used to identify different narratives.

# Chapter 3

# Narratives and Decision Making

## 3.1 What Narratives Are

### 3.1.1 McAdams Research on Narratives

### 3.1.2 Social Psychology Background

## 3.2 How Narratives can help

### 3.2.1 Bayesian Brain and Predictive Coding

Here, there could be a direct link to the algorithms that are used in Machine Learning, AI, and NLP.

### 3.2.2 Influence and Change on Human Beings

Akerlof and Shiller understand narratives as a convention, but it is more than that, it changes how people think and perceive the world. Akerlof and Snower (2016)

## 3.3 Narrative Research

Chapter 4

# Topic Analysis

In this chapter, I take a closer look at topic mining and analysis, a field of natural language processing that aims at analyzing the content of text data. To that end, I introduce unsupervised text mining techniques called probabilistic topic models that discover latent topics in the text data. Zhai and Massung (2016, p. 329) define a topic as "the main idea discussed in the text data, which may also be regarded as a theme or subject of a discussion or conversation."

The reason for the use of topic analysis in this thesis is obvious: News articles that present different narratives on the target rate adjustment will do so by discussing different themes or topics concerning such an adjustment. On an analytical level, we can expect that articles reporting along the lines of different narratives will also employ a different vocabulary. This is exactly what a probabilistic topic modeling with its bag-of-words method is able to pick up on.

A well-known topic model is Latent Dirichlet Allocation (LDA) which was first introduced by Blei, Ng, and Jordan (2003). LDA has been further developed and adapted by Hoffman, Bach, and Blei (2010) to allow for the processing of massive text data collections that arrive continuously. Although LDA is a very popular topic model, this thesis focuses on Probabilistic Latent Semantic Analysis (PLSA), which preceded LDA and provided the foundation LDA was built on. Given that I analyze a large but fixed set of newspaper articles (see Chapter 5.2), PLSA is an appropriate model that offers sufficient functionality for the analysis.[1]

Probabilistic Latent Semantic Analysis (PLSA) was first introduced by Hofmann (1999).

[1] Unlike PLSA, LDA is able to assign probabilities to unseen documents. While this comes in handy when out of sample properties are to be assessed, it offers no advantage in the current analysis.

Blei et al. (2003, p. 994) called Hofmann's model a "significant step forward" in probabilistic topic modeling. The main achievement of PLSA was to supplement the theory of Latent Semantic Analysis with a sound statistical foundation and to introduce a proper generative model of the data (Hofmann, 1999, p. 289). It is based on the idea that documents are mixtures of topics and the generative model specifies a probabilistic procedure by which documents are assumed to be generated (Steyvers & Griffiths, 2007, p. 2). As Blei et al. (2003, p. 994) explain, PLSA "models each word in a document as a sample from a mixture model, where the mixture components are multinomial random variables that can be viewed as representations of 'topics'." Let's take a closer look at the generative process via such mixture models.

## 4.1 Generative Model

As input a topic modeling task takes a collection of text documents and a specification of the number of topics, as output it produces the topics as well as the coverage of the topics in every document (Zhai & Massung, 2016, pp. 330–331), as illustrated in Table 4.1.[2]

| Input | |
|---|---|
| • A collection of $N$ text documents | $C = \{d_1, ..., d_N\}$ |
| • Number of topics | $k$ |
| **Output** | |
| • Coverage of topics in each document $d_i$ | $\{\pi_{i1}, ..., \pi_{ik}\}$, with $\sum_{j=1}^{k} \pi_{ij} = 1$ |
| • $k$ topics | $\{\theta_1, ..., \theta_k\}$ |

**Table 4.1** – Formal definition of topic modeling task.

In the context of topic modeling, a topic is represented by a probability distribution over words (or terms). That is, a topic is a distribution that assigns each word in the vocabulary set a probability. Words closely associated with the topic are given a higher probability weight so that they are more likely to come up if we were to sample from the distribution. In general, all words in the vocabulary set may have a non-zero probability in any given word distribution. This accounts for the eventuality that a word that is completely unrelated to a topic shows up in a text on the topic (Zhai and Massung 2016, pp. 335–337; Blei et al. 2003, p. 994). Figure 4.1 illustrates four topics, for each of which the 16 words with the highest probability weights are

[2] Please note that this thesis follows the notation of Zhai and Massung (2016).

**Figure 4.1** – An illustration of four topics (Steyvers & Griffiths, 2007, p.2).

| word | prob. | word | prob. | word | prob. | word | prob. |
|---|---|---|---|---|---|---|---|
| DRUGS | .069 | RED | .202 | MIND | .081 | DOCTOR | .074 |
| DRUG | .060 | BLUE | .099 | THOUGHT | .066 | DR. | .063 |
| MEDICINE | .027 | GREEN | .096 | REMEMBER | .064 | PATIENT | .061 |
| EFFECTS | .026 | YELLOW | .073 | MEMORY | .037 | HOSPITAL | .049 |
| BODY | .023 | WHITE | .048 | THINKING | .030 | CARE | .046 |
| MEDICINES | .019 | COLOR | .048 | PROFESSOR | .028 | MEDICAL | .042 |
| PAIN | .016 | BRIGHT | .030 | FELT | .025 | NURSE | .031 |
| PERSON | .016 | COLORS | .029 | REMEMBERED | .022 | PATIENTS | .029 |
| MARIJUANA | .014 | ORANGE | .027 | THOUGHTS | .020 | DOCTORS | .028 |
| LABEL | .012 | BROWN | .027 | FORGOTTEN | .020 | HEALTH | .025 |
| ALCOHOL | .012 | PINK | .017 | MOMENT | .020 | MEDICINE | .017 |
| DANGEROUS | .011 | LOOK | .017 | THINK | .019 | NURSING | .017 |
| ABUSE | .009 | BLACK | .016 | THING | .016 | DENTAL | .015 |
| EFFECT | .009 | PURPLE | .015 | WONDER | .014 | NURSES | .013 |
| KNOWN | .008 | CROSS | .011 | FORGET | .012 | PHYSICIAN | .012 |
| PILLS | .008 | COLORED | .009 | RECALL | .012 | HOSPITALS | .011 |

listed. Based on these words, the topics could be labeled drug use, colors, memory, and health care (Steyvers & Griffiths, 2007, p.2). The aim of probabilistic topic models is to discover such word distributions in the text data.

Going back to the formal definition in Table 4.1, each $\theta_i$ is a word distribution. Thus,

$$\sum_{w \in V} p(w|\theta_i) = 1,$$

where $p(w|\theta_i)$ is the probability of word $w$ in the word distribution $\theta_i$ with $w$ being a unique word in the vocabulary set $V = \{w_1, w_2, ..., w_M\}$. The vocabulary set is the set of unique words in a collection $C$ (Zhai & Massung, 2016, pp. 338).

The generative model underlying topic models is based on probabilistic sampling rules and specifies how the words in documents are generated on the basis of latent variables. Given the definition of topics as word distributions, the model assumes that every word in a document is drawn from a topic. From which topic a word is to be drawn is determined by a distribution over topics. As such, documents with different content are generated by choosing different distributions over topics (Steyvers & Griffiths, 2007, p. 2–3). Figure 4.2 illustrates the process of generating three documents from two topics. While two of the documents are generated exclusively from one topic, one document is an equal mixture of both topics. Note that words can be part of more than one topic, allowing the model to capture the multiple meanings of polysemies (Steyvers & Griffiths, 2007, p. 2–3).

The generative process can be inverted to appear as a problem of statistical inference (see

**Figure 4.2** – The generative model and the problem of statistical inference (Steyvers & Griffiths, 2007, p.3).

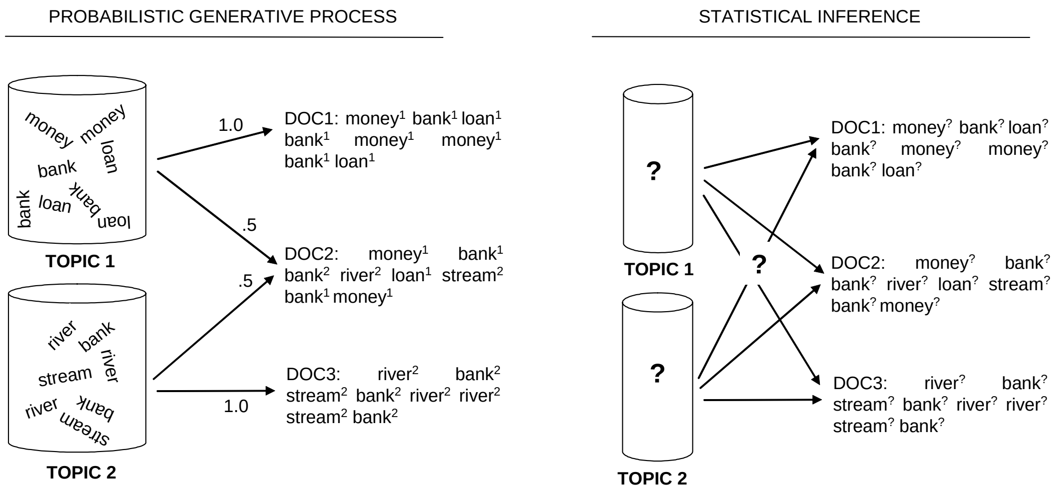



Figure 4.2): When fitting the generative model, the aim is to identify the latent variables, here the latent topics, that are best able to explain the observed data. Thus the model is able to infer the probability distributions over words and the distribution over topics for each document from the text data (Steyvers & Griffiths, 2007, p. 3). While the inferred distributions over words are interpreted as topics, the probability distribution over topics can be usefully thought of as the extent to which a topic accounts for the content of a document. That is, the fraction of words in document $d_i$ generated by topic $\theta_j$. Note that the probabilities of document $d_i$ covering topic $\theta_j$ sum to one across all topics:

$$\sum_{j=1}^{k} \pi_{ij} = 1.$$

Thus, the topics $\{\theta_1, ..., \theta_k\}$, as inferred by the topic model, account for every word in the document, with no words being attributed to other topics (Zhai & Massung, 2016, pp. 338).

There is one key assumption underlying the generative model: It is assumed that the order of the words in a document as well as the order of the documents in a collection can be neglected. The model picks up on the frequencies of words, but completely ignores the position of the words in the text. Should the order of the words contain useful information, it is discarded by probabilistic topic models. This assumptions is known as the “bag-of-words” assumption, an assumption of exchangeability of the the words in a document (Blei et al. 2003, p. 994; Steyvers and Griffiths 2007, p. 3). Hofmann (1999, p. 290) himself admits that the "key assumption is that the simplified 'bag-of-words' [...] representation of documents will in many cases preserve most of the relevant information." In Figure 4.2, this principle is captured by illustrating topics as bags containing different distribution over words. The position of the words is arbitrary.

## 4.2 Model Fitting with Expectation-Maximization

As outlined above, a topic model is a mixture model that consists of several (latent) variables, i.e. word distributions. An estimation of the model's variables is obtained via Maximum Likelihood (ML) estimation. If a mixture model consists of several components, that is more than just one topic, there is no analytical solution to the ML estimation and a numerical optimization algorithm must be employed. The Expectation-Maximization algorithm (EM) is a family of useful algorithms to obtain an ML estimate of mixture models (Zhai & Massung, 2016, p. 359) According to Hofmann (1999, p. 290; 2001, p. 181), EM is the "standard procedure for maximum likelihood estimation in latent variable models."

In a first step, the ML estimation problem is to be defined. Based on the generative model outlined in Chapter 4.1, the likelihood function is derived. Recall that $p(w|\theta_j)$ is the probability of word $w$ in the word distribution $\theta_j$ and that $\pi_{d,j}$ is the coverage of topic $\theta_j$ in document $d$. When faced with the problem of statistical inference, $\pi_{d,j}$ can be thought of as the probability of topic $\theta_j$ generating word $w$. Then, the probability of observing the sequence of words that make up document $d$ is:

$$p(d) = \prod_{w \in V} \left( \sum_{j=1}^{k} \pi_{d,j} \; p(w|\theta_j) \right)^{c(w,d)} \tag{4.1}$$

$$\log p(d) = \sum_{w \in V} c(w,d) \, \log \left( \sum_{j=1}^{k} \pi_{d,j} \; p(w|\theta_j) \right) \tag{4.2}$$

with $w$ being a unique word in the vocabulary set $V = \{w_1, w_2, ..., w_M\}$ and $c(w,d)$ being the count of word $w$ in document $d$. If we consider the entire collection of $N$ documents

$C = \{d_1, ..., d_N\}$, the log likelihood function is:

$$\log p(C|\pi_{d,j}, \theta_j) = \sum_{d \in C} \sum_{w \in V} c(w, d) \log \left( \sum_{j=1}^{k} \pi_{d,j}\ p(w|\theta_j) \right) \tag{4.3}$$

A few words of explanation. $\sum_{j=1}^{k} \pi_{d,j}\ p(w|\theta_j)$ denotes the probability of word $w$ occurring once in document $d$. It is the probability of observing the word regardless of which topic is used. We assume independence in generating each word, hence the probability of the document equals the product of the probability of each word in the document (see Eq. 4.1). Eq. 4.2 is the log likelihood function for document $d$, just as Eq. 4.3 is the log likelihood for the entire collection $C$ (Hofmann, 1999; 2001; Steyvers & Griffiths, 2007; for notation see Zhai & Massung, 2016, p. 340–377).

With the likelihood function defined, we are faced with a constrained optimization problem:

$$\underset{\pi_{d,j}\ \theta_j}{\arg\max}\ p(C|\pi_{d,j}, \theta_j) \tag{4.4}$$

$$\text{s. t.} \sum_{w \in V} p(w|\theta_j) = 1,\ j \in [1, k] \qquad \sum_{j=1}^{k} \pi_{dj} = 1,\ d \in C \tag{4.5}$$

All that is left to do is to solve the maximization problem. If we knew from which topic a word is generated it would be straightforward to calculate the ML estimate. In such a case, the word distributions would simply be the normalized word frequencies (see Chapter 4.3 for elaborations on unigram language models). However, the partitioning of the words among the topics is not known, that is we do not know from which topic a word is generated. Therefore, we have to fall back on an iterative algorithm to solve the maximization problem. The EM algorithm proceeds by guessing from which distribution a word is generated using a tentative estimate of the parameters. Based on the assumed partitioning, the estimate of the parameters is updated. This, in turn, allows an improved inference of the partitioning, leading to an iterative hill-climbing algorithm that improves the estimate of the parameters until it reaches a local optimum (Zhai & Massung, 2016, p. 360). The algorithm proceeds in two steps: In the expectation step (E), posteriori probabilities for the latent variables are computed based on current estimates. In the maximization step (M), the parameters are updated (Hofmann, 1999, p. 290; 1999, pp. 181–182).

In the E-step, the algorithm uses Bayes rule to infer the topic that has been used to generate a word. We introduce a hidden variable $z_{d,w} \in \{1, 2, ..., k\}$ to capture this information. The value of this hidden variable is inferred based on a tentative set of parameters (Hofmann, 1999, p. 290; 2001, p. 182; Zhai & Massung, 2016, pp. 362–374):

$$p(z_{d,w} = j) = \frac{\pi_{d,j}^{(n)} p^{(n)}(w|\theta_j)}{\sum_{j=1}^{k} \pi_{d,j}^{(n)} p^{(n)}(w|\theta_j)} \tag{4.6}$$

Eq. 4.6 shows the E-Step of the EM algorithm for PLSA, where $p(z_{d,w} = j)$ is the probability that word $w$ in document $d$ is generated from topic $\theta_j$. Note that this probability depends on the document $d$, i.e. whether a word has been generated from a specific topic depends on the document. As a result, each document will have a different topic distribution, $\pi_{d,j}$, indicating the varying emphasis on specific topics across documents. I will use this fact to classify articles according to different narratives (see Chapter 5.3.3). The probability value of $z_{d,w}$ is calculated for every unique word in each document by computing the product of the probability of word $w$ given the selected topic and the probability of selecting a topic. The product is normalized to ensure the constraints in Eq. 4.5 hold. The superscript $^{(n)}$ indicates the generation of parameters in the iteration (Zhai & Massung, 2016, pp. 374–376).

In the M-step we calculate an ML estimate of the parameters $\pi_{d,j}$ and $p(w|\theta_j)$ using the estimate obtained in step E. We use the estimated partitioning of the words among the topics to adjust the word count $c(w, d)$, that is we obtain a discounted word count $c(w, d)p(z_{d,w} = j)$ which can be understood as the expected count for the event that word $w$ is generated from $\theta_j$. Remember that the ML estimation could not be solved analytically because we did not know the partitioning of the words. Having obtained a tentative guess of the partitioning, we can easily estimate $\pi_{d,j}$ (the probability of document $d$ covering topic $\theta_j$) and $p(w|\theta_j)$ (the probability of word $w$ for topic $\theta_j$) (Hofmann, 1999, p. 290; 2001, p. 182; Zhai & Massung, 2016, pp. 364–375):

$$\pi_{d,j}^{(n+1)} = \frac{\sum_{w \in V} c(w, d)\ p(z_{d,w} = j)}{\sum_{j \in k} \sum_{w \in V} c(w, d)\ p(z_{d,w} = j)} \tag{4.7}$$

$$p^{(n+1)}(w|\theta_j) = \frac{\sum_{d \in C} c(w, d)\ p(z_{d,w} = j)}{\sum_{w \in V} \sum_{d \in C} c(w, d)\ p(z_{d,w} = j)} \tag{4.8}$$

As Eq. 4.7 and Eq. 4.8 illustrate, the computation of the parameter estimates amounts to no more than aggregating the the expected word counts and normalizing them among all topics or among all words, respectively. The normalization must ensure that the constraints in Eq. 4.5 hold. The new generation of parameters is used to adjust the probabilities of the $z_{d,w}$ values, which subsequently can be used to compute a next generation of parameters. The EM algorithm continues to iterate over the E- and M-step until the likelihood converges. Instead of using a convergence condition, a stopping rule can be employed. In this case, the algorithm stops updating the parameters once a sufficient performance, as defined in the stopping criterion, is reached (Hofmann, 2001, pp. 182–183).

As mentioned before, EM is a hill-climbing algorithm that starts with an initial guess of the parameter values and then successively improves it. In the process of improving the estimate, the EM algorithm is guaranteed to converge to a local, but not a global maximum. To account for this, Zhai and Massung recommend applying the algorithm repeatedly with changing initial values for the unknown parameters and using the run with the highest likelihood value (2016, pp. 363–368).[3]

## 4.3 Adding a Background Language Model

In the original PLSA (Hofmann, 1999; 2001) as it was introduced in Chapter 4.2 there is no background topic. The model contains $k$ topics, $\{\theta_1, ..., \theta_k\}$, that are treated as unknown and inferred via a probabilistic process. As a result, very common words will show up prominently in the learned topics. Yet, common words often carry very little relevant information, such as stop words (more on stop words in Chapter 5.3.1). The aim of a background topic is to represent the common words in the text data so that the learned topics can capture the content-bearing words and the statistical semantic word patterns of interest are revealed (Darling, 2011, pp.7–8).

An example of a probabilistic topic model that includes background distributions is Chemudugunta, Smyth, and Steyvers's (2007) special words with background model (SWB). It is an extension to LDA which allows words in a document to be modeled as either originating from general topics, or from document-specific "special" word distributions, or from a corpus-wide background distribution. However, a simpler approach, where only one background distribution is added

[3] For more details on the properties of EM, see Chapter 9 in Bishop (2006).

to LDA, is possible as well, as Darling (2011, pp.7–8) demonstrates. Analogously to Darling (2011), I add a single background distribution to my PLSA model.

Let's look at the behavior of a mixture model to understand the usefulness of a background model. Topics are word distributions with the simplest word distribution being the unigram language model. We assume that words in a text are generated independently. Thus, the likelihood of a sequence of words is equal to the product of the probability of each word (Zhai & Massung, 2016, pp. 51–54). Assume that $\theta$ is the single topic to be inferred from document $d$. Then the ML estimation problem is given by

$$p(d|\theta) = \prod_{w \in V} p(w|\theta)^{c(w,d)} \tag{4.9}$$

$$\text{s. t.} \sum_{w \in V} p(w|\theta) = 1 \tag{4.10}$$

the solution to which is

$$p(w_i|\hat{\theta}) = \frac{c(w_i, d)}{|d|} \tag{4.11}$$

where $c(w, d)$ is the word count in document $d$ and $|d|$ is the total number of words in document $d$ (Zhai & Massung, 2016, pp. 341-343).

Eq. 4.11 illustrates that the ML estimator of a unigram language model gives each word a probability equal to its relative frequency in the document. Consequently, only observed words have a non-zero probability while words that do not show up have zero probability. Further, a very high probability is given to words that show up frequently. The more prominent a word, the higher its probability in the word distribution. However, usually functional words and stop words are used the most frequently in a text and thus will have high probabilities, while content-carrying words will have a much lower probability. Thus, a topic can be completely dominated by uninformative words. Introducing a background language model can help filter out such words (Zhai & Massung, 2016, pp. 51–54).

To change the predominance of common words, the generative process needs to be adjusted so that the learned topics do not have to generate these words. Specifically, another word distribution that generates the common words, called the background topic ($\theta_B$), has to be introduced. A natural choice for the background topic is the unigram language model because it assigns high probabilities to frequent words. Including a background model in the mixture model will allow the learned topics to assign high probabilities to content-carrying words instead (Darling, 2011, pp.7–8; Chemudugunta et al., 2007).

Consider the optimization problem of a mixture model in Eq. 4.4 and Eq. 4.5. In order to maximize the overall likelihood, different topics tend to put a high probability on different words. Put differently, when a word, $w_1$, is assigned a high probability in one word distribution, $w_1$ will tend to have a low probability in another distribution. That is, if $p(w_1|\theta_j)$ is high, then $p(w_1|\theta_i)$ tends to be low, and vice versa. Thus, the behavior of a mixture model ensures that when a background model is introduced that puts high probabilities on common words, the other distributions are encouraged to put a low probability on common words and more probability mass on words that rank low in the background distribution. Of course, due to the nature of the ML estimate of a mixture model, if a word appears very frequently in the text, it will tend to have a high probability in all distributions as this optimizes the overall likelihood (Zhai & Massung, 2016, pp. 353–359).

If a background topic is introduced into the mixture model, a decision needs to be made as to whether a word is generated by the background model, $\theta_B$, or by one of the learned models, $\theta_i$. This choice is controlled by a probability distribution over the components of the mixture models: $\lambda_B$ denotes the probability that the background model was used to generate a word. The likelihood function is given by (Zhai & Massung, 2016, p. 372):

$$\log\, p(C|\pi_{d,j}, \theta_j) = \sum_{d \in C} \sum_{w \in V} c(w,d) \,\log \left( \lambda_B p(w|\theta_B) + (1-\lambda_B) \sum_{j=1}^{k} \pi_{d,j}\; p(w|\theta_j) \right) \tag{4.12}$$

The difference to the likelihood function in Chapter 4.2 is marked in blue. The ML estimation problem remains unchanged (see Eq. 4.4 and Eq. 4.5). Note that the probability of observing a word from the background distribution is $\lambda_B\; p(w|\theta_B)$ while the probability of observing a word from the learned topic $\theta_j$ is $(1-\lambda_B)\; p(w|\theta_j)$. The background model, $\theta_B$, is assumed to be known and can, for instance, be estimated as a unigram language model of the collection. This is powerful way to introduce domain knowledge about the text into the model (Zhai & Massung, 2016, pp. 352, 372–376).

Consider how the probability of choosing the background model $\lambda_B$ influences the maximization problem. The higher the probability of choosing the background model is, the higher is the probability mass on the words in the background model $\lambda_B p(w|\theta_B)$. The opposite is true for learned topics as $(1 - \lambda_B)p(w|\theta_j)$ will be lower the higher $\lambda_B$. As a result, inferred distributions will give a low probability to common words because they have a high probability in the background topic. Instead, the learned topics will give other, content-bearing words a higher probability. The higher $\lambda_B$, the more common words are filtered out from the learned topics $\{\theta_1, ..., \theta_k\}$ (Zhai & Massung, 2016, pp. 352–359).

$\lambda_B$ indicates the fraction of words generated from the background model. This parameter needs to be specified and can be understood as a prior in the process of Bayesian inference. It encodes a prior belief about the partitioning of the words between the background and the learned topics and thus again allows us to introduce domain knowledge that we may have about the text (Zhai & Massung, 2016, pp. 361, 372–376).

Based on Eq. 4.12 the EM algorithm looks as follows:

$$p(z_{d,w} = j) = \frac{\pi_{d,j}^{(n)} p^{(n)}(w|\theta_j)}{\sum_{j=1}^{k} \pi_{d,j}^{(n)} p^{(n)}(w|\theta_j)} \tag{4.13}$$

$$p(z_{d,w} = B) = \frac{\lambda_B \; p(w|\theta_B)}{\lambda_B \; p(w|\theta_B) + (1 - \lambda_B) \sum_{j=1}^{k} \pi_{d,j}^{(n)} p^{(n)}(w|\theta_j)} \tag{4.14}$$

Eq. 4.13 and Eq. 4.14 show the E-Step of the EM algorithm for PLSA with a background model, where $p(z_{d,w} = j)$ is the probability that word $w$ in document $d$ is generated by topic $\theta_j$ (conditional on not being generated by the background model) and $p(z_{d,w} = B)$ is the probability of $w$ being generated by the background model $\theta_B$. The M-step is:

$$\pi_{d,j}^{(n+1)} = \frac{\sum_{w \in V} c(w,d) \; (1 - p(z_{d,w} = B)) \; p(z_{d,w} = j)}{\sum_{j \in k} \sum_{w \in V} c(w,d) \; (1 - p(z_{d,w} = B)) \; p(z_{d,w} = j)} \tag{4.15}$$

$$p^{(n+1)}(w|\theta_j) = \frac{\sum_{d \in C} c(w,d) \; (1 - p(z_{d,w} = B)) \; p(z_{d,w} = j)}{\sum_{w \in V} \sum_{d \in C} c(w,d) \; (1 - p(z_{d,w} = B)) \; p(z_{d,w} = j)} \tag{4.16}$$

As before, in the M-step we use a tentative partitioning of the words, as derived in the E-step, to update the parameters. Now, however, words are partitioned among the $k$ inferred topics, $\{\theta_1, ..., \theta_k\}$, conditional on not being allocated to the background topic (Zhai & Massung, 2016, pp. 372–376).

Note that the PLSA model with a background topic is the more general case. By setting $\lambda_B = 0$ we arrive at the formulae for the PLSA model without background distribution. If $\lambda_B = 1$, the background topic is an infinitely strong prior that leaves no room for other topics. This is in line with the intuitive understanding of $\lambda_B$ as the fraction of words that are generated by the background topic.

# Chapter 5

# Data and Methodology

This chapter gives an overview of the data used in this thesis. The financial data set is outlined in Chapter 5.1, the text data set in Chapter 5.2. Further, Chapter 5.3 explains how the data was processed and applied to the topic modeling task.

## 5.1 Financial Data

Data on target rate adjustments and FOMC meetings was retrieved from the Fed's website (Federal Reserve System, 2018, 2013; Federal Open Market Committee, 2018a; 2018b) and summarized in table A.1 and A.2 in Appendix A.

The FOMC holds eight regularly scheduled meetings during the year. Additional, unscheduled meetings are called when necessary (Federal Open Market Committee, 2018b). In April 2011, the Fed has taken up the practice of holding quarterly press conferences where it comments on its policy decisions, including its treatment of the target rate. Usually, the press conferences take place after every other meeting. In June 2018, the Fed announced that starting January 2019 it will hold a press conference after every meeting (Federal Open Market Committee, 2018c). Table A.1 in Appendix A lists all 195 FOMC meetings that have taken place over the last 20 years. If a meeting lasted two days, only the last day is listed. Unsurprisingly, certain years necessitated more emergency meetings than others: In 2008, the FOMC held a total of 14 meetings, 6 more than usual. In 2001, the FOMC held 13 meetings, two of which were conference calls shortly after the events of 9/11. While the market anticipates the scheduled meetings and forms expectations about potential policy actions, the same is not possible for unscheduled emergency meetings.

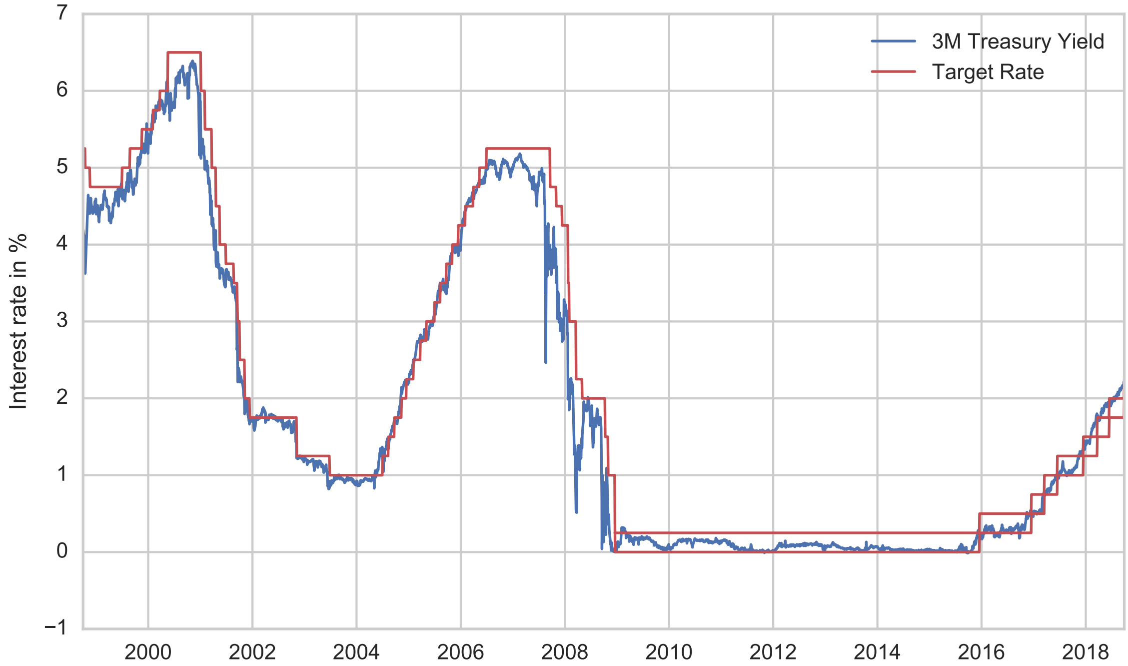
Table A.2 in Appendix A gives an overview of the changes in the target rate over the past 20 years. From January 1998 to September 2018, the target rate was adjusted on 57 occasions. Only six of the 57 adjustments happened after unscheduled meetings. Most notably, two surprise adjustments took place during 2008 and three during 2001, one of them shortly after 9/11. Until 2008, the Fed used to decide and implement the new target rate on the very same day. The practice was changed after 2008 and the target rate was subsequently adjusted on the day following the FOMC meeting.

The development of the target rate over time is characterized by periods of stark increase and decrease. In May 2000, the target rate was at its highest with 6.5%. From January 2001 to June 2003, the target rate decreased steadily until it reached a low point of 1%. Mid-2006, the target rate was again at 5.25% but was lowered drastically in 2007 and 2008 until it reached its all-time low of 0% in December 2008. At the same time, the FOMC introduced a target range and defined the target rate no longer as a discrete number but with the help of an upper and lower bound. For seven years, until december 2015, the FOMC kept the target rate in the range of 0% to 0.25%. Since then, it has slowly increased the target rate, announcing a range of 2%–2.25% in September 2018.

The daily yield curve of the US Treasury bills, notes, and bonds for the period of October 1, 1998 to September 30, 2018 was retrieved from Thomson Reuters Datastream. Figure 5.1 illustrates the development of the 3-month treasury yield as well as the target rate over the 20-year period. As explained in Chapter 2.2, Ellingsen et al. approximate unexpected monetary policy actions with the change in the 3-month T-bill rate. They argue, that the 3-month rate is sufficiently short to be determined by the target rate, but also sufficiently long to avoid noise from expectation errors (Ellingsen et al., 2003, p. 13). Indeed, Figure 5.1 shows that the 3-month T-bill rate and the target rate seem to move in tandem.

The treasury yields for all other maturities are depicted in Figure A.1 in Appendix A. Please note that for the 1-month rate the data series starts on July 31, 2001 and for the 7-year rate the series starts on May 26, 2009. Due to data availability, the 20-year rate is given as a constant maturity rate while the rates of the other maturities are given as bid yields. As expected, while

**Figure 5.1** – Target rate and 3M treasury yield.




the short term yields (up to 1 year) move quite closely with the target rate, the longer the maturity the more it emancipates itself from the target rate.

## 5.2 Text Data

Articles are collected from the Dow Jones Factiva Database[1] by use of the search terms *Federal Reserve* and *interest rate*. Only articles that appeared in the *United States* on the subject of *interest rates* in a window of three days around each target rate adjustment (the day before, of, and after an adjustment as listed in Table A.2) are taken into account. The articles are mainly taken from the *The Wall Street Journal*, *Financial Times*, *Reuters*, *The Associated Press*, *Market News International*, *Dow Jones Institutional News*, *Agence France*, and *AFX*, as these newspapers seem to publish the most articles on the topic.

From October 1, 1998 to September 30, 2018, 56 target rate adjustments have taken place, for which a total of 2'318 articles have been extracted. Since the FED changed its policy from undertaking a target rate change on the same day as a FOMC meeting to only doing so on the following day, articles ±1 day around the official target rate change have been collected. This

[1] https://global.factiva.com

**Figure 5.2** – Wordcloud illustrating the most common words across all articles.

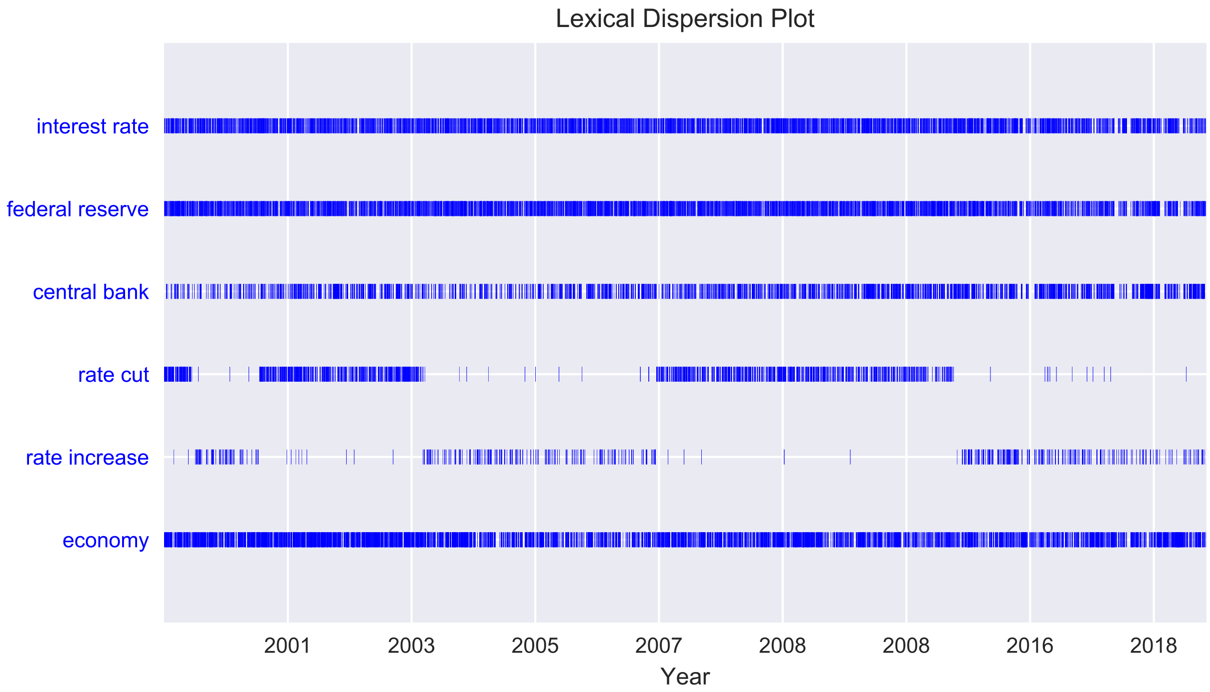

ensures that the articles capture any information, speculation, and interpretations that abound directly after the meeting, on the day of the target rate change as well as on the following day. On rare occasions, the number of articles ran in the several hundreds and only the most relevant have been selected.

The final sample comprises between 14 and 97 publications for every policy day. In total, over 1′600′000 tokens are extracted from these articles. After stop words have been removed, the token count drops to roughly 850′000. For more information on text preprocessing, see Chapter 5.3.1. Figure B.1 in Appendix B illustrates the spread of the text data across policy days and shows that the token count fluctuates quite drastically. More data is available for recent target rate adjustments, less for more historic adjustments.

Looking at the most common words in the entire text data base, as illustrated in Figure 5.2, yields no surprises. As one might expect, the articles make heavy use of context specific vocabulary such as ”interest rate”, ”federal reserve”, ”economy”, but also of terms appropriate for a financial environment such as ”basis point”, ”percentage point”, or ”investor”. Even analyzing just the headlines of the articles yields an almost identical word usage.

A closer look at the dispersion of the most common words reveals that their usage is con-

**Figure 5.3** – Dispersion of most common words across time.




sistent over time (see Figure 5.3[2]). The frequently used terms are just as likely to be used in older articles as in newer publications. As one might expect, the articles talk about ”rate cut” in times where the Fed is lowering the target rate, and about ”rate increase” when the Fed is in the process of raising the target rate.

## 5.3 Methodology

This Chapter explains the preprocessing of the text data (Chapter 5.3.1), the engenieering of the feature space (Chapter 5.3.2) and the application of the PLSA model with and without background topic (Chapter 5.3.3 and 5.3.4).

[2] A remark on the label of the x-axis: The x-axis tracks the token count of the entire text data and the diversion plot indicates at which token count the word of interest has been found. However, displaying the token count at which a word is found has little meaning. Thus, the x-axis displays the progression of years in which the articles were published. Since the tokens are not spread smoothly across the years, the intervals are irregular.

### 5.3.1 Text Preprocessing

Text data in its raw form contains all the ambiguities that are inherent in natural language. Thus, before any NLP tasks can be performed, the text data needs to be prepared. Text preprocessing is the process of converting raw text into "a well-defined sequence of linguistically meaningful units." This part of the NLP process is essential, as any analysis that follows is based on the linguistic units (characters, words, and sentences) identified at this stage (Palmer, 2010, p. 9). Topic models in particular are sensitive to preprocessing as they rely on a sparse vocabulary (Schofield & Mimno, 2016, p. 288).

Recall that topic models rely on the "bag-of-words" assumption and neglect the order of words in a text as well as the order of documents in a corpus. Therefore, the main task of the preprocessing stage is the correct identification of words as the linguistically meaningful units to be used in topic modeling. Palmer (2010, p. 10) refers to the process of converting text data into its components as text segmentation. In particular, the process of breaking up a sequence of characters at the word boundaries is called word segmentation. Words identified by word segmentations are known as tokens and the process of identifying them is also known as tokenization (Palmer, 2010, p. 10).

Palmer (2010, pp. 16-19) rightly points out that there exists plenty of tokenization ambiguity. In a language where words are generally space-delimited, a natural approach to tokenization is identifying any sequence of characters preceded and followed by a space as a token. However, this leaves a certain amount of ambiguity when it comes to punctuation and multi-word expressions. Apostrophes, for instance, are used to mark contractions, the genitive form of nouns, and even the plural form of nouns. Tokenization needs to implement a rule regarding the treatment of such ambiguous cases. Consider the example word *Peter's*. In this case, *'s* could indicate a possession as well as a contraction of the verb *is* and it could be tokenized as *Peters* or as two tokens, *Peter* and *'s*, or even as *Peter* and *is*, if that was the supposed meaning (Palmer, 2010, pp. 16-19).

Bird, Klein, and Loper (2009) supply a series of word tokenizers. However, even they admit that when it comes to tokenization "no single solution works well across-the-board, and we must decide what counts as a token depending on the application domain." The problem of tokenization gets compounded by the problem of stop words. As outlined in Chapter 4.3, functional words occur frequently in texts and thus can clutter up the information we are trying to discover. Such stop words carry little information and obscure the analysis by crowding-out more meaningful words. It can therefore be useful to remove stop words from the text data altogether.

To remove stop words, one can make use of publicly available lists of stop words. However, Nothman, Qin, and Yurchak (2018) point out the difficulties and incompatibilities of such lists. Investigated the variation and consistency of several lists, they come to the conclusion that stop words are often included based on frequency statistics but also based on manual selection and filtering. Thus, lists vary greatly in the words they include. Apart from uncertainty regarding the selection of the stop words, stop lists may also assume a particular tokenization strategy. Yet, as Nothman et al. (2018) point out, the stop lists provided by some open source libraries do not always go with the tokenizer provided by the same libraries. For example, while a tokenizer might split the word *hasn't* into the tokens *has* and *n't*, the accompanying stop word list only contains *has* and *not*, omitting the token *n't*. Further, several lists include the words *has* and *have*, but fail to list *had*, letting the omission and inclusion of stop words appear arbitrary (Nothman et al., 2018, pp. 7–11).

By general assumption, topic models benefit from stop word removal (Schofield, Magnusson, & Mimno, 2017a, p. 432). Blei et al. (2003), Chemudugunta et al. (2007), Steyvers and Griffiths (2007), and Hofmann (2001) removed stop words during text preprocessing as so keep these words from impacting their analysis. According to popular belief, topic models benefit from manually constructed stop word lists (Schofield et al., 2017a, p. 432). Darling (2011, p. 7), for instance, claims that stop lists must often be domain-specific. Contrarily, Schofield et al. (2017a) show that a domain specific approach to stop word removal is not necessary. While the removal of very frequent terms improves the quality of topic models, the removal of text specific stop words yields little benefit. If the removal of common stop words proves insufficient, unwanted terms can be removed after model training, yielding better results than ex-ante removal (Schofield et al., 2017a, p. 432; Schofield, Magnusson, & Mimno, 2017b).

Another common step in the stage of preprocessing is stemming. Stemming removes the suffixes of words in order to reduce related words to an identical stem. The words *happily*, *happier*, and *happy*, for instance, might be reduced to the same token (Schofield & Mimno, 2016,

p. 287). Potential advantages of stemming are an improved model fit due to a reduced feature space, improved stability in the topics due to elimination of small morphological differences, and improved interpretability due to word equivalence. However, stemming may also render words unrecognizable (e.g. *stai* as the stem of *stay*), or it may lead to the incorrect conflation of unrelated words (e.g. *absolutely* and *absolution*), and it does not deal well with morphologically similar words that carry different meanings (Schofield & Mimno, 2016, p. 287). Schofield and Mimno (2016) analyze the effect of stemming in topic modeling and find that it yields little benefit and might actually be harmful. For one, stemming leads to no likelihood gains after controlling for vocabulary size. Further, the coherence of the topics is not improved by stemming. As Schofield et al. (2017b) point out, topic models are already placing morphologically similar words in the same topic, thus making stemming for the purpose of morphological conflation redundant. Lastly, strong stemming increases the model's sensitivity to random initialization as it decreases clustering consistency, that is stemming does not ensure that related words are spread across fewer topics (Schofield & Mimno, 2016, pp. 293–295).

In conclusion, Schofield et al. (2017b) recommend to pre-process lightly, that is to remove only the most frequent stop words and to omit stemming altogether, in order to avoid discarding useful text information.

The text data used in this thesis is collected from the Down Jones Factiva Database and converted to .txt files for further processing. The metatext of all articles has been discarded so that only the main body of text and the headline remains. To segment the text into words, I apply the tokenizer provided by Bird et al. (2009) in the nltk library. This tokenizer uses regular expressions to split the text at the word boundaries, generally treating punctuation characters as separate tokens and splitting standard contractions, i.e. *don't* is tokenized as *do* and *n't* and *children's* as *children* and *'s.*

Subsequently, I filter out stop words using the stop word list for the English language provided by the nltk library (Bird et al., 2009). The list encompasses 179 words, including common verbs, articles, prepositions, and contractions. Nothman et al. (2018, pp. 8–9) caution against stop lists that fail to complement the accompanying tokenizer and mention that many stop lists also tend to include controversial words, like *system* and *cry.* A closer examination of the nltk list reveals that the token *n't* is not included in the list, even though the nltk tokenizer clearly produces it. Further, the stop word lists contains many contractions without the preceding apostrophe, even though the tokenizer joins the apostrophe and the enclitic in one token. For instance, the stop word list carries *ll*, *s*, *d*, and *ve* while the tokenizer produces *'ll*, *'s*, *'d*, and *'ve*. On the upside, the nltk list appears to carry no obviously controversial terms. I manually add the missing tokens to the stop word list before applying it.

In line with the recommendation from Schofield et al. (2017b), I refrain form applying a stemmer. However, I covert all characters to lower case, as to ensure that word equivalence is not hindered by capitalization. Also, I remove all non-alphabetic characters as well as tokens of length one, as numbers, punctuation marks, and single characters are of little use when it comes to topic generation and interpretation.

After tokenization, the text data collection encompasses more than 1'600'000 tokens. After the removal of stop words, the token count drops to roughly 1'100'000 tokens. Subsequently, the non-alphabetic characters, especially the punctuations marks, are removed and the final count amounts to approximately 850'000 tokens. Thus, even though a conservative preprocessing strategy has been chosen, the text data is reduced by almost half during preprocessing (see Figure B.1 in Appendix B).

### 5.3.2 Feature Engineering

Selecting the features is the most important factor in determining the success of a statistical learning process. While machine learning models are often general, feature engineering is domain-specific and a lot of effort in any machine learning projects has to be dedicated to the construction of the feature space. At the same time, feature engineering can appear as "black art" in the sense that it requires intuition, creativity, and an understanding of the domain (Domingos, 2012). Thus, exercising the necessary discretion, I make a few adjustments to feature space to improve and speed up the learning process.

Considering the most common words in the tokenized text data (see Figure 5.2) makes apparent that some tokens show up in tandem, such as *interest rate*, *federal reserve*, or *basis point*. It stands to reason that these terms should be tokenized as one token instead of two separate words, as they have a conjoint meaning. Processing the text data as bigrams, I select the 50 most common bigrams and add them to the feature space (see Table 5.1[3]).

| bigram | count | bigram | count | bigram | count |
|---|---|---|---|---|---|
| federal reserve | 3'353 | economic growth | 804 | last week | 560 |
| interest rate | 3'342 | wall street | 803 | federal open | 547 |
| interest rates | 3'338 | percent percent | 759 | central banks | 540 |
| central bank | 2'263 | fed funds | 759 | rate hike | 540 |
| funds rate | 2'209 | next year | 712 | fed rate | 522 |
| per cent | 2'202 | open market | 691 | long term | 494 |
| new york | 1'176 | rate increases | 679 | raise rates | 492 |
| rate cut | 1'151 | rate percent | 677 | point cut | 486 |
| percentage point | 1'142 | financial markets | 672 | stock market | 485 |
| basis points | 1'139 | chief economist | 650 | cut rates | 485 |
| federal funds | 1'139 | market committee | 633 | discount rate | 472 |
| short term | 1'119 | quarter percentage | 619 | two year | 461 |
| monetary policy | 1'108 | basis point | 589 | half point | 453 |
| dow jones | 1'106 | term interest | 578 | rate hikes | 453 |
| quarter point | 980 | rate increase | 573 | oil prices | 435 |
| rate cuts | 827 | fed officials | 569 | | |

**Table 5.1** – The top bigrams included in the feature space.

Including the tokens in table 5.1, the feature space consists of over 23'300 unique tokens, some of which occur only once or twice. However, tokens that occur only rarely increase the sparsity of the feature space and slow down computation without having any impact on topic generation. Thus, tokens that appear only once (about 7'900 tokens), twice (about 3'400 tokens), or thrice (around 1'900 tokens) are removed. This reduces the feature space from 23'300 unique tokens to approximately 10'150 tokens.

### 5.3.3 Application of PLSA

Using the preprocessed and prepared text data, I apply the PLSA algorithm without a background topic. The model specifications are presented in Table 5.2.

As explained in Chapter 2.3, we are looking for two narratives according to which policy days can be classified. Thus, we set $k = 2$ which will yield two topics, $\{\theta_1, \theta_2\}$. As recommended by Zhai and Massung (2016, p. 363), I run the algorithm repeatedly (six times, to be specific)

[3] The attentive reader will have noticed that the table lists only 47 bigrams. From the top 50 bigrams, the expressions *wsj com*, *fed said*, and *year treasury* have been excluded as they have no clear conjoint meaning. Note that the author set the limit at 50 because these bigrams are mostly idiosyncratic expressions featuring prominently in the text data while lower ranked bigrams appear more random with a less specific conjoint meaning.

| Input | |
|---|---|
| • A collection of 2'318 text documents | $C = \{d_1, ..., d_{2318}\}$ |
| • Number of topics | $k = 2$ |
| **Output** | |
| • Coverage of topics in each document $d_i$ | $\{\pi_{i1}, \pi_{i2}\}$, with $\pi_{i1} + \pi_{i2} = 1$ |
| • 2 topics | $\{\theta_1, \theta_2\}$ |

**Table 5.2** – Specification of PLSA model.

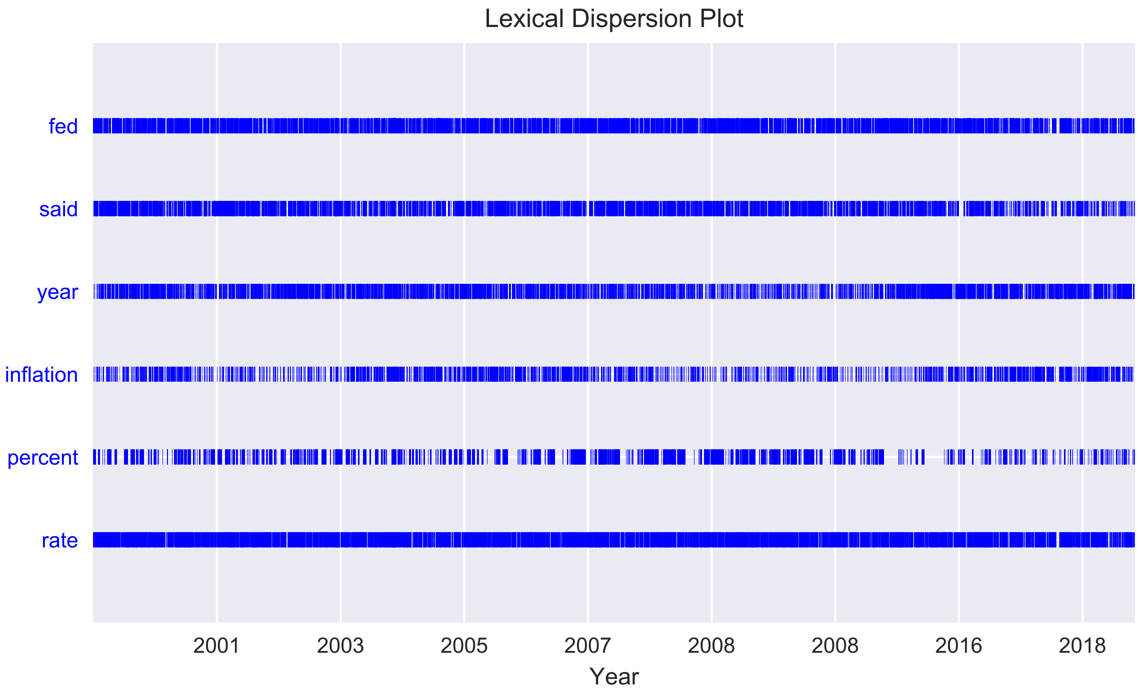
using different random initializations. In the end, I use the run with the highest likelihood value. The top 20 words of each topic are listed in Table 5.3.

| Topic 1 | | Topic 2 | |
|---|---|---|---|
| word | prob | word | prob |
| said | .011 | fed | .024 |
| year | .011 | said | .013 |
| percent | .009 | inflation | .011 |
| market | .009 | economy | .010 |
| rate | .008 | rate | .009 |
| dollar | .007 | would | .007 |
| fed | .007 | year | .006 |
| cut | .007 | rates | .006 |
| bank | .006 | economic | .006 |
| rates | .006 | percent | .005 |
| investors | .006 | growth | .005 |
| interest rate | .005 | policy | .005 |
| federal reserve | .005 | market | .005 |
| per cent | .004 | federal reserve | .005 |
| markets | .004 | interest rates | .005 |
| bond | .004 | could | .005 |
| expected | .004 | meeting | .004 |
| yield | .004 | time | .004 |
| interest rates | .004 | statement | .004 |
| us | .004 | interest rate | .004 |

**Table 5.3** – The two topics produced by PLSA.

Considering the top words of the topics, no clear topical concentration becomes salient. Both topics attribute a high weight to words concerning interest rates, the federal reserve, and the economy. This is hardly surprising as the articles have been selected for their focus on interest rates and the federal reserve (see Chapter 5.2). It is to be expected that the articles will make heavy use of a related vocabulary, which, by the nature of PLSA, will come to dominate the inferred topics. A look at the the lexical dispersion of the top words as shown by Figure 5.4

**Figure 5.4** – Dispersion of the top words in the topics inferred by original PLSA across time.




illustrates that the top words appear frequently in all articles. As explained in Chapter 4.3, the introduction of a background topic could help shift the weight from frequent terms to more meaningful, content-carrying terms. For the implementation of such a model see Chapter 5.3.4.

The text data collection comprises between 14 to 97 articles for every policy day. The words of each article are partitioned between the two topics, resulting in a topic coverage probability that indicates to what extent an article is to be attributed to a certain topic. To classify a policy day as belonging either to narrative one ($N_1$) or narrative two ($N_2$), I average the topic coverage probabilities of all articles on a given policy day and attribute the day to the topic that dominates the articles. Figure C.1 in Appendix C.1 depicts the resulting classification of the policy days.

38 policy days have been attributed to topic 2 and only 18 days to topic 1. Also, quite often both topics command a large share of the articles on a given day, one topic dominating the other only by a small margin. Thus, the classification of the policy days is not as clear cut as one might wish. If we consider how the individual articles have been classified, we would expect that on a given day all articles are attributed to the same topic. That would be a clear indication that a specific narrative, as identified by our topic model, permeated the press reporting at that

time. However, as Figure C.2 and C.3 show, quite often the articles on a specific day are almost evenly split between the topics.

### 5.3.4 Application of PLSA with Background Topic

To improve the topics inferred by PLSA, I introduce a background topic that captures the most common words in the text data. The background topic is a unigram language model generated from the entire text data collection. Thus, every word has been given a probability weight that corresponds to its relative frequency in the collection (see Eq. 4.11). Table 5.4 lists the 20 top words of the background topic.

| Background Topic | | | |
|---|---|---|---|
| word | prob | word | prob |
| fed | .016 | federal reserve | .005 |
| said | .013 | cut | .005 |
| year | .008 | interest rate | .004 |
| rate | .007 | interest rates | .004 |
| percent | .007 | bank | .004 |
| economy | .007 | economic | .004 |
| inflation | .006 | expected | .004 |
| market | .006 | markets | .003 |
| rates | .006 | dollar | .003 |
| would | .005 | could | .003 |

**Table 5.4** – The background topic $\theta_B$ introduced in PLSA.

The most common word in the background topic is *fed*, followed by *said*, *year*, and *rate*. These words also feature prominently in the inferred topics of PLSA without a background distribution. Including the background topic should liberate the learned topics so that meaningful topics can be inferred. To that end, I also need to specify how many words are generated by the background topic (as indicated by the parameter $\lambda_B$). Since this quantity is unknown, I run the model several times with different specifications for $\lambda_B$. The model specifications are as presented in Table 5.5.

The top 20 words of every topic are listed in Tables C.1–C.3 in Appendix C.2. The higher $\lambda_B$, the more the count of common words in the topics decreases in favor of more interesting words. Topic 1 appears to focus on *investing*: With an increasing $\lambda_B$, terms like *investors*, *yield*, *rose*, *fell*, *trading*, *index*, and *stocks* climb to the top of the word distribution. Topic 2, on the

| Input | |
|---|---|
| • A collection of 2'318 text documents | $C = \{d_1, ..., d_{2318}\}$ |
| • Number of topics | $k = 2$ |
| • The background topic | $\theta_B$ |
| • Fraction of words generated by $\theta_B$ | $\lambda_B \in [0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9]$ |
| **Output** | |
| • Coverage of topics in each document $d_i$ | $\{\pi_{i1}, \pi_{i2}\}$, with $\pi_{i1} + \pi_{i2} = 1$ |
| • 2 topics | $\{\theta_1, \theta_2\}$ |

**Table 5.5** – Specification of PLSA model with background topic.

other hand, seems to focus on the Fed's *economic analysis*: With an increasing $\lambda_B$, terms like *economic*, *growth*, *greenspan*, *bernanke*, *statement*, *chairman*, and *economists* come to dominate the topic. Thus, the topics clearly pick up on different subjects (or narratives) in the articles and, with an increased use of the background topic, the topical focus of the word distributions becomes salient.

Figure C.4 in Appendix C.2 depicts the resulting classification of the policy days for $\lambda_B = 0.1$ and Figure C.7 for $\lambda_B = 0.9$. With $\lambda_B = 0.1$, 37 policy days have been attributed to topic 2 and only 19 days to topic 1. With $\lambda_B = 0.9$, 34 policy days have been attributed to topic 2 and only 22 days to topic 1. The introduction of a background model as well as the level of $\lambda_B$ has very little impact on the classification of the policy days. As Figures C.5–C.6 and C.8–C.9 show, the higher $\lambda_B$, the less ambiguous is the classification of single documents. This is hardly surprising: When $\lambda_B = 0.9$, only 10% of the tokens of a document are still split between topic 1 and 2, which will more commonly lead to an allocation of all tokens to the same topic.

There are several possible approaches to evaluate topic models and their specifications. First, a test can be set up in which a human judge must assess topic quality, in particular the coherence of the inferred topics. Second, traditional metrics, such as log-likelihood of the held-out data, can be employed to measure how well a topic model generalizes. However, such metrics are often negatively correlated with topic quality as assessed by human judges. A third approach is downstream task improvement: A model is assessed based on the statistically significant improvement it brings to another task (Zhai & Massung, 2016, pp.383–384).

For the purpose of this thesis, we have already outlined how an increase in $\lambda_B$ makes the topics more interpretable. In Chapter 6 we will also consider how the different topic model specifications affect the downstream task of the financial analysis.

Chapter 6

# Results

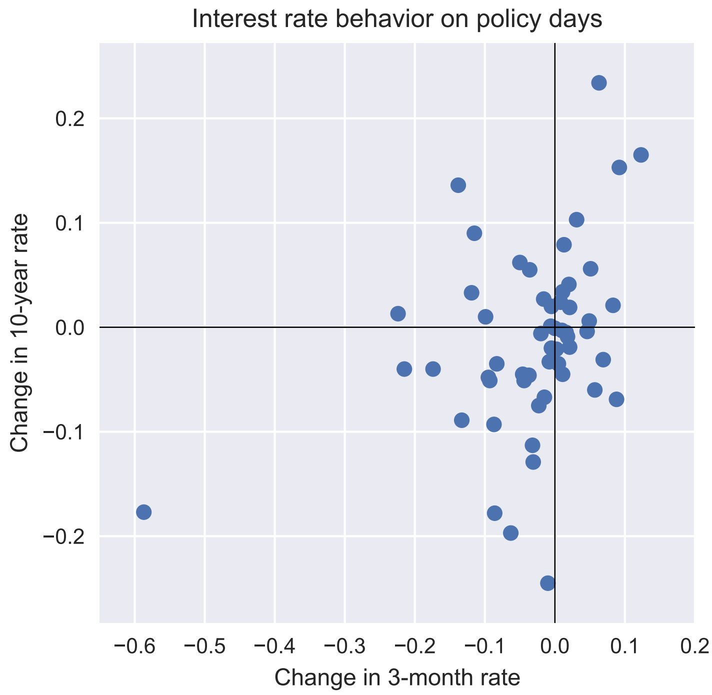
## 6.1 Classification of Policy Days

Following Ellingsen et al. (2003), a day on which the target rate has been adjusted by the Fed is considered a policy day. If the FOMC held a meeting but failed to change the target rate, the day is considered a non-policy day. In this thesis, a 20-year period from October 1998 to October 2018 is analyzed. During this time frame, there are 56 policy days and 5'161 non-policy days.

Analogously to Ellingsen et al. (2003, pp. 14 – 15), we take a look at the behavior of the interest rates. Figure 6.1 plots the change in the 3-month rate against the change in the 10-year rate on the policy days. Ellingsen et al. (2003, p. 14) found a clear positive relationship between the change of the 10-year and the 3-month rate, with some outliers. Similarly, Figure 6.1 shows a predominantly positive relationship, barring some odd observations, and a concentration around the origin. This indicates that the long-term rates tend to move in the same direction as the short term rates and the target rate change.

In Figures 6.2 and 6.3, the policy days are split up according to their classification by either Narrative one or Narrative two (as classified by PLSA without background topic, that is for $\lambda_B = 0$). A successful classification of policy days is expected to separate the days characterized by a co-movement of the yield curves from the days on which the rates move in opposite directions. As such, one figure should show a clear positive relationship, with the data points populating quadrants 1 and 3, while the other figure should display a negative

**Figure 6.1** – Response of the 10-year interest rate to a change in the 3-month rate on policy days.




relationship, with all data points populating quadrant 2 and 4. However, Figures 6.2 and 6.3 show a rather ambiguous picture. Both figures capture policy days with co-movement of the rates as well as days with counter-movement. It should be noted that Ellingsen et al. (2003, pp. 14–15), too, only found an ambiguous relationship.

Figures D.1–D.4 in Appendix D show the split of the policy days when $\lambda_B$ is increased to 0.1 and 0.9, respectively. Even though the output of the topic model changes notably in terms of topic quality with an increase in $\lambda_B$, the introduction of a background topic has little impact on the classification of the policy days. As such, Figures Figures D.1–D.4 present almost the same picture as Figures 6.2 and 6.3.

There is a small overlap in the time frame of Ellingsen et al. (2003) and this thesis. In Figures D.5 – D.7 in Appendix D.1 a comparison of the classification of Ellingsen et al. and this thesis is presented. However, the comparable sample is quite small due to the small overlap in the time frame as well as the fact that Ellingsen et al. chose no to classify a quarter of their data. Thus, only 12 of Ellingsen et al.'s endogenous policy days and 6 of their exogenous policy days have been classified in the course of this thesis. Figures D.5 – D.7 show that the classification according to narratives does not make a separation between the two groups as Ellingsen et al.

**Figure 6.2** – Response of the 10-year interest rate to a change in the 3-month rate on policy days influenced by Narrative one ($\lambda_B = 0$).

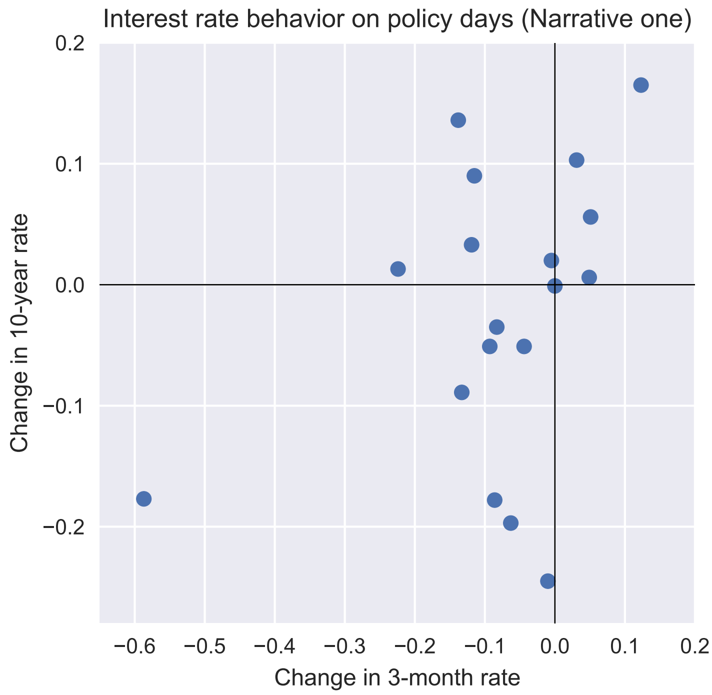
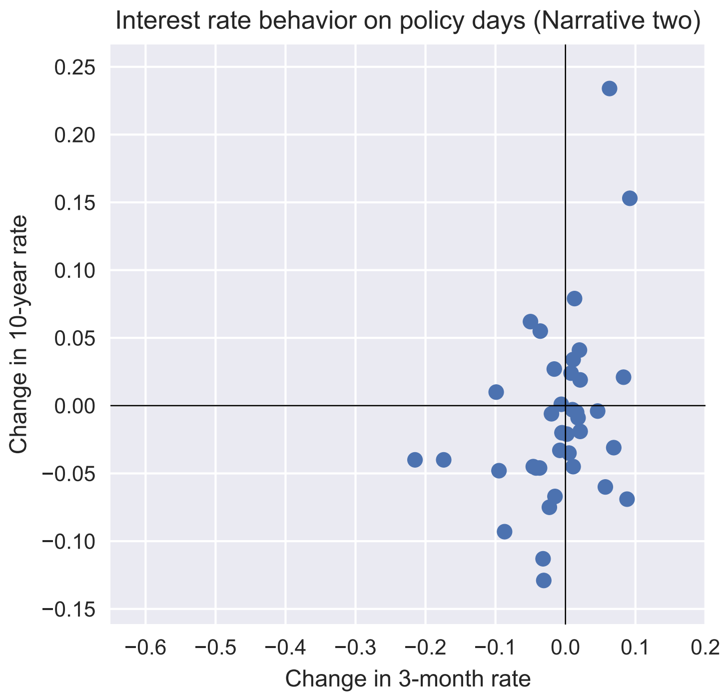



**Figure 6.3** – Response of the 10-year interest rate to a change in the 3-month rate on policy days influenced by Narrative two ($\lambda_B = 0$).



have done. Moreover, the classification of the selected policy days varies with a change in $\lambda_B$: Without background topic, the policy days are classified by narrative one, after the introduction of a background topic, they are classified by narrative two. This is rather unusual as $\lambda_B$ fails to have an impact on the classification of policy days in all other settings. Most likely, the sample is to small for the purpose of a stable and reliable comparison.

## 6.2 Regression Results

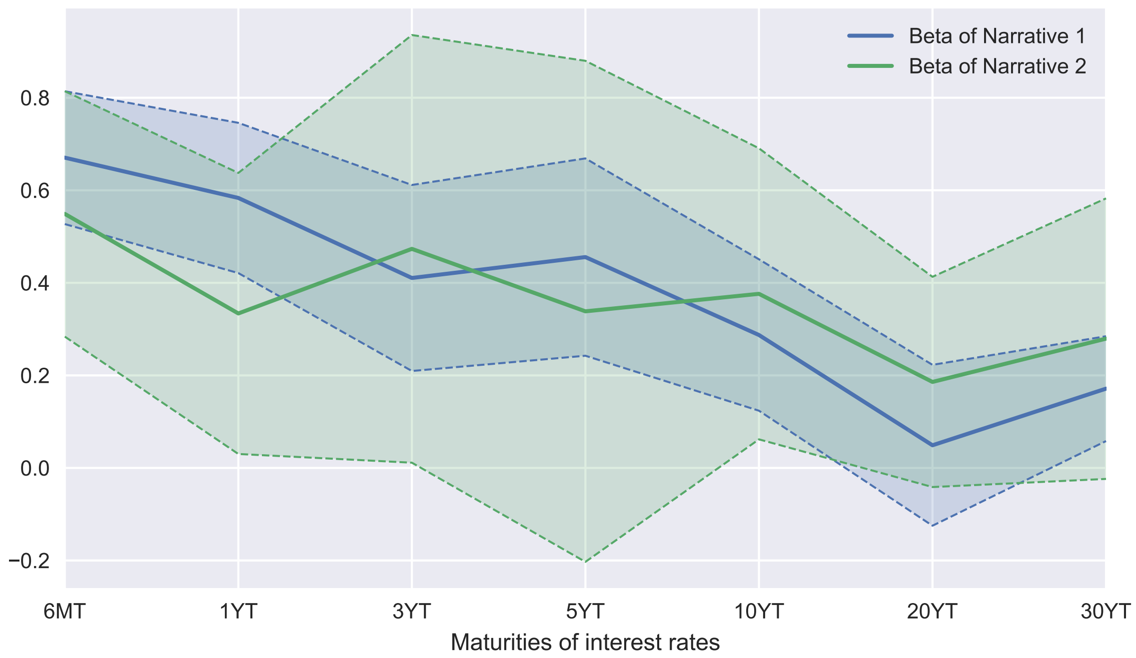
In a next step, we run the regression formulated in Eq. 2.3 for the classification with $\lambda_B = 0$. Table 6.1 reports the ordinary-least-square estimates for maturities from 6 month to 30 years. $\beta_n^{NP}$ measures the reaction of the $n$-maturity interest rate to the 3-month rate on non-policy days, $\beta_n^{N1}$ measures the reaction on policy days classified as being influenced by Narrative one and $\beta_n^{N2}$ on policy days classified as belonging to Narrative two. $\alpha_n$ is the constant term. **/* indicate significance on the 1% and 5% level, respectively. Heteroskedasticity robust White's (1980) standard errors are listed in brackets. The last row reports the test statistics from the test of equality between the estimated coefficients.

| | 6m | 1y | 3y | 5y | 10y | 20y | 30y |
|---|---|---|---|---|---|---|---|
| $\alpha_n$ | -0.00 | -0.00 | -0.00 | -0.00 | -0.00 | -0.00 | -0.00 |
| | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) |
| $\beta_n^{NP}$ | 0.36** | 0.30** | 0.28** | 0.25** | 0.18** | 0.13** | 0.12** |
| | (0.05) | (0.04) | (0.04) | (0.03) | (0.03) | (0.02) | (0.02) |
| $\beta_n^{N1}$ | 0.67** | 0.58** | 0.41** | 0.46** | 0.29** | 0.05 | 0.17** |
| | (0.07) | (0.08) | (0.10) | (0.11) | (0.08) | (0.09) | (0.06) |
| $\beta_n^{N2}$ | 0.55** | 0.33* | 0.47* | 0.34 | 0.38* | 0.19 | 0.28 |
| | (0.14) | (0.15) | (0.24) | (0.28) | (0.16) | (0.12) | (0.15) |
| $R^2$ | 0.30 | 0.20 | 0.08 | 0.05 | 0.03 | 0.02 | 0.02 |
| $\beta_n^{N1} = \beta_n^{N2}$ | 0.12 | 0.25 | -0.06 | 0.12 | -0.09 | -0.14 | -0.11 |

**Table 6.1** – Regression results (classification with $\lambda_B = 0$).

As expected, the intercept is always zero. However, the null hypothesis (see Eq. 2.4) finds little support in the data. The slope coefficients for the non-policy days are large and strongly significant for most maturities. Yet, in line with hypothesis 2.4, we would expect to find that $\beta_n^{N1}$ is positive for short maturities but negative for long maturities. Even though the slope coefficient $\beta_n^{N1}$ is decreasing with longer maturities and remains significant, it does not turn negative.

**Figure 6.4** – Regression coefficients for narrative one and two without background topic.




The slope coefficient $\beta_n^{N2}$, by contrast, decreases less in maturity. Figure 6.4 illustrates the change in the coefficients given the duration of the maturities. Indeed, while the difference in the betas is positive for short maturities, it turns negative for long maturities. This, at least, indicates that the classification according to narratives picks up on short- and long-term motivation. However, the differences between the coefficients are never significant. As the equality test shows, the two responses, $\beta_n^{N1}$ and $\beta_n^{N2}$, cannot be statistically separated for any maturity. In line with this finding, Figure 6.4 illustrates that the 95% confidence intervals of the two coefficients mostly overlap.

Tables D.1 and D.2 in Appendix D.2 show the regression result for policy day classifications with $\lambda_B = 0.1$ and $\lambda_B = 0.9$, respectively. However, the introduction of a background topic changes little in terms of policy day classification and, consequently, the regression results are very similar across all $\lambda_B$. Thus, also the relationship of the slope coefficients remains stable across $\lambda_B$, a fact that is once again illustrated by Figures D.8 and D.9 in Appendix D.2. Additionally, the introduction of a background topic has no effect whatsoever on the explained variation ($R^2$).

43

Chapter 7

# Discussion

# Chapter 8

# Conclusion

# References

Akerlof, G. A., & Snower, D. J. (2016). Bread and bullets. *Journal of Economic Behavior & Organization*, *126*, 58–71.

Bird, S., Klein, E., & Loper, E. (2009). *Natural language processing with python. Analyzing text with the natural language toolkit.* O'Reilly Media. Retrieved from `http://www.nltk.org/book/`

Bishop, C. M. (2006). *Pattern recognition and machine learning.* New York, NY: Springer.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, *3*(Jan), 993–1022.

Chemudugunta, C., Smyth, P., & Steyvers, M. (2007). Modeling general and specific aspects of documents with a probabilistic topic model. In *Advances in neural information processing systems* (pp. 241–248).

Cook, T., & Hahn, T. (1989). The effect of changes in the federal funds rate target on market interest rates in the 1970s. *Journal of Monetary Economics*, *24*(3), 331–351.

Darling, W. M. (2011). A theoretical and practical implementation tutorial on topic modeling and gibbs sampling. In *Proceedings of the 49th annual meeting of the association for computational linguistics: Human language technologies* (pp. 642–647).

Domingos, P. (2012). A few useful things to know about machine learning. *Commun. acm*, *55*(10), 78–87.

Ellingsen, T., & Söderström, U. (2001). Monetary policy and market interest rates. *American Economic Review*, *91*(5), 1594–1607.

Ellingsen, T., & Söderström, U. (2004). Why are long rates sensitive to monetary policy? *SSRN*

*Electronic Journal*.

Ellingsen, T., Söderström, U., & Masseng, L. (2003). Monetary policy and the bond market. *Stockholm School of Economics. Photocopy*.

Engelberg, J. E., & Parsons, C. A. (2011). The causal impact of media in financial markets. *The Journal of Finance*, *66*(1), 67–97.

Federal Open Market Committee. (2018a). *Historical materials by year.* Retrieved 2018-10-04, from `https://www.federalreserve.gov/monetarypolicy/fomc_historical_year.htm`

Federal Open Market Committee. (2018b). *Meeting calendars, statements, and minutes (2013-2018).* Retrieved 2018-10-04, from `https://www.federalreserve.gov/monetarypolicy/fomccalendars.htm`

Federal Open Market Committee. (2018c). *Transcript of chairman powell's press conference june 13, 2018.* Retrieved 2018-10-04, from `https://www.federalreserve.gov/mediacenter/files/FOMCpresconf20180613.pdf`

Federal Reserve System. (2013). *Open market operations archive.* Retrieved 2018-10-04, from `https://www.federalreserve.gov/monetarypolicy/openmarket_archive.htm`

Federal Reserve System. (2018). *Open market operations.* Retrieved 2018-10-04, from `https://www.federalreserve.gov/monetarypolicy/openmarket.htm`

Goetzmann, W., Kim, D., & Shiller, R. J. (2016). *Crash beliefs from investor surveys* (Working Paper No. w22143). National Bureau of Economic Research.

Gürkaynak, R. S., Sack, B. P., & Swanson, E. T. (2003). The excess sensitivity of long-term interest rates: Evidence and implications for macroeconomic models. *Finance and Economics Discussion Paper*.

Gürkaynak, R. S., Sack, B. P., & Swanson, E. T. (2005a). Do actions speak louder than words? the response of asset prices to monetary policy actions and statements. *International Journal of Central Banking*, *1*(1), 55–93.

Gürkaynak, R. S., Sack, B. P., & Swanson, E. T. (2005b). The sensitivity of long-term interest rates to economic news: Evidence and implications for macroeconomic models. *American Economic Review*, *95*(1), 425–436.

Hoffman, M., Bach, F. R., & Blei, D. M. (2010). Online learning for latent dirichlet allocation. In *Advances in neural information processing systems* (pp. 856–864).

Hofmann, T. (1999). Probabilistic latent semantic analysis. In *Proceedings of the fifteenth conference on uncertainty in artificial intelligence* (pp. 289–296). Morgan Kaufmann Publishers Inc.

Hofmann, T. (2001). Unsupervised learning by probabilistic latent semantic analysis. *Machine Learning*, *42*(1/2), 177–196.

Kräussl, R., & Mirgorodskaya, E. (2014). *News media sentiment and investor behavior* (Working Paper No. 492). Center for Financial Studies.

Nothman, J., Qin, H., & Yurchak, R. (2018). Stop word lists in free open-source software packages. In *Proceedings of workshop for nlp open source software (nlp-oss)* (pp. 7–12).

Palmer, D. D. (2010). Text preprocessing. In N. Indurkhya & F. J. Damerau (Eds.), *Handbook of natural language processing* (pp. 9–30). Boca Raton, FL: CRC Press.

Romer, C. D., & Romer, D. H. (2000). Federal reserve information and the behavior of interest rates. *American Economic Review*, *90*(3), 429–457.

Schofield, A., Magnusson, M., & Mimno, D. (2017a). Pulling out the stops: Rethinking stopword removal for topic models. In *Proceedings of the 15th conference of the european chapter of the association for computational linguistics: Volume 2, short papers* (pp. 432–436).

Schofield, A., Magnusson, M., & Mimno, D. (2017b). Understanding text pre-processing for latent dirichlet allocation. In *Proceedings of the 15th conference of the european chapter of the association for computational linguistics: Volume 2* (pp. 432–436).

Schofield, A., & Mimno, D. (2016). Comparing apples to apple: The effects of stemmers on topic models. *Transactions of the Association for Computational Linguistics*, *4*, 287–300.

Steyvers, M., & Griffiths, T. (2007). Probabilistic topic models. In T. Landauer, D. McNamara, & Dennis, S., & Kintsch., W. (Eds.), *Latent semantic analysis: A road to meaning* (pp. 424–440). Lawrence Erlbaum.

von Krosigk, M. (2017). *Classification of monetary policy decisions through text mining techniques* (Unpublished master's thesis). University of St. Gallen, St. Gallen, Switzerland.

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, *48*(4), 817–838.

Yuan, Y. (2015). Market-wide attention, trading, and stock returns. *Journal of Financial Economics*, *116*(3), 548–564.

Zhai, C., & Massung, S. (2016). *Text data management and analysis: A practical introduction* *to information retrieval and text mining* (First edition ed., Vol. 12). New York, NY and San Rafael: Association for Computing Machinery and Morgan & Claypool Publishers.

Appendix A

# Target Rate and Yield Curves

| Dates of all FOMC meetings since January 1, 1998 | | | | | | |
|---|---|---|---|---|---|---|
| 26.09.2018 | 18.03.2015 | 13.12.2011 | *07.02.2009 | 12.12.2006 | 12.08.2003 | 31.01.2001 |
| 01.08.2018 | 28.01.2015 | *28.11.2011 | 28.01.2009 | 25.10.2006 | 25.06.2003 | *03.01.2001 |
| 13.06.2018 | 17.12.2014 | 02.11.2011 | *16.01.2009 | 20.09.2006 | 06.05.2003 | 19.12.2000 |
| 02.05.2018 | 29.10.2014 | 21.09.2011 | 16.12.2008 | 08.08.2006 | *16.04.2003 | 15.11.2000 |
| 21.03.2018 | 17.09.2014 | 09.08.2011 | 29.10.2008 | 29.06.2006 | *08.04.2003 | 03.10.2000 |
| 31.01.2018 | 30.07.2014 | *01.08.2011 | *07.10.2008 | 10.05.2006 | *01.04.2003 | 22.08.2000 |
| 13.12.2017 | 18.06.2014 | 22.06.2011 | *29.09.2008 | 28.03.2006 | *25.03.2003 | 28.06.2000 |
| 01.11.2017 | 30.04.2014 | 27.04.2011 | 16.09.2008 | 31.01.2006 | 18.03.2003 | 16.05.2000 |
| 20.09.2017 | 19.03.2014 | 15.03.2011 | 05.08.2008 | 13.12.2005 | 29.01.2003 | 21.03.2000 |
| 26.07.2017 | *04.03.2014 | 26.01.2011 | *24.07.2008 | 01.11.2005 | 10.12.2002 | 02.02.2000 |
| 14.06.2017 | 29.01.2014 | 14.12.2010 | 25.06.2008 | 20.09.2005 | 06.11.2002 | 21.12.1999 |
| 03.05.2017 | 18.12.2013 | 03.11.2010 | 30.04.2008 | 09.08.2005 | 24.09.2002 | 16.11.1999 |
| 15.03.2017 | 30.10.2013 | *15.10.2010 | 18.03.2008 | 30.06.2005 | 13.08.2002 | 05.10.1999 |
| 01.02.2017 | *16.10.2013 | 21.09.2010 | *10.03.2008 | 03.05.2005 | 26.06.2002 | 24.08.1999 |
| 14.12.2016 | 18.09.2013 | 10.08.2010 | 30.01.2008 | 22.03.2005 | 07.05.2002 | 30.06.1999 |
| 02.11.2016 | 31.07.2013 | 23.06.2010 | *21.01.2008 | 02.02.2005 | 19.03.2002 | 18.05.1999 |
| 21.09.2016 | 19.06.2013 | *09.05.2010 | *09.01.2008 | 14.12.2004 | 30.01.2002 | 30.03.1999 |
| 27.07.2016 | 01.05.2013 | 28.04.2010 | 11.12.2007 | 10.11.2004 | 11.12.2001 | 03.02.1999 |
| 15.06.2016 | 20.03.2013 | 16.03.2010 | *06.12.2007 | 21.09.2004 | 06.11.2001 | 22.12.1998 |
| 27.04.2016 | 30.01.2013 | 27.01.2010 | 31.10.2007 | 10.08.2004 | 02.10.2001 | 17.11.1998 |
| 16.03.2016 | 12.12.2012 | 16.12.2009 | 18.09.2007 | 30.06.2004 | *17.09.2001 | *15.10.1998 |
| 27.01.2016 | 24.10.2012 | 04.11.2009 | *16.08.2007 | 04.05.2004 | *13.09.2001 | 29.09.1998 |
| 16.12.2015 | 13.09.2012 | 23.09.2009 | *10.08.2007 | 16.03.2004 | 21.08.2001 | *21.09.1998 |
| 28.10.2015 | 01.08.2012 | 12.08.2009 | 07.08.2007 | 28.01.2004 | 27.06.2001 | 18.08.1998 |
| 17.09.2015 | 20.06.2012 | 24.06.2009 | 28.06.2007 | 09.12.2003 | 15.05.2001 | 01.07.1998 |
| 29.07.2015 | 25.04.2012 | *03.06.2009 | 09.05.2007 | 28.10.2003 | *18.04.2001 | 19.05.1998 |
| 17.06.2015 | 13.03.2012 | 29.04.2009 | 21.03.2007 | 16.09.2003 | *11.04.2001 | 31.03.1998 |
| 29.04.2015 | 25.01.2012 | 18.03.2009 | 31.01.2007 | | 20.03.2001 | 04.02.1998 |

* indicates an unscheduled meeting/conference call

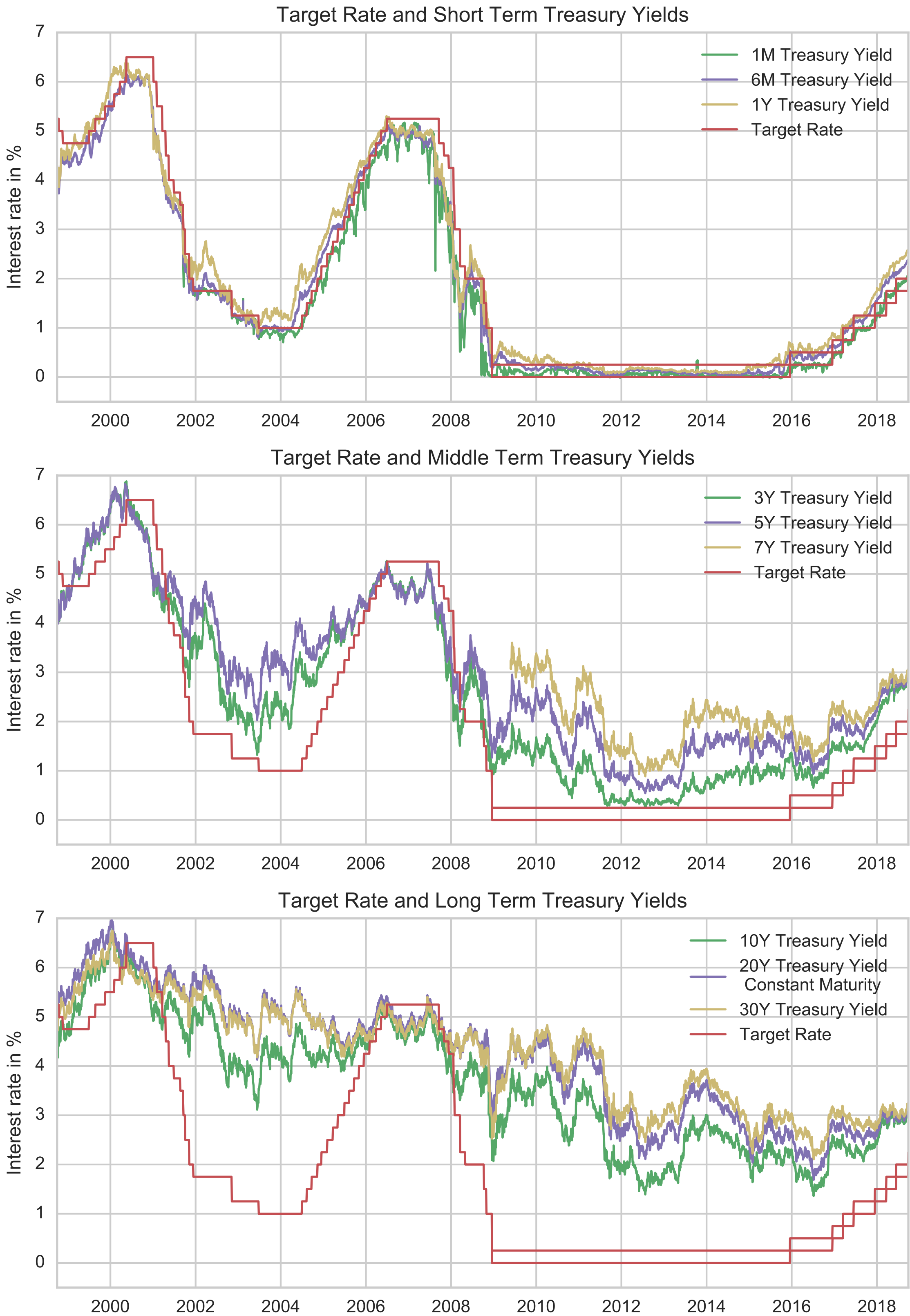
**Table A.1** – FOMC meetings.

| Date | $Tgt_{low}$ | $Tgt_{up}$ | $\Delta Tgt_{low}$ | $\Delta Tgt_{up}$ | FOMC Meeting on day prior | FOMC Meeting on same day | Date | $Tgt_{low}$ | $Tgt_{up}$ | $\Delta Tgt_{low}$ | $\Delta Tgt_{up}$ | FOMC Meeting on day prior | FOMC Meeting on same day |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 27.09.2018 | 2.00% | 2.25% | 25 | - | 1 | 0 | 14.12.2004 | 2.25% | - | 25 | - | 0 | 1 |
| 14.06.2018 | 1.75% | 2.00% | 25 | - | 1 | 0 | 10.11.2004 | 2.00% | - | 25 | - | 0 | 1 |
| 22.03.2018 | 1.50% | 1.75% | 25 | - | 1 | 0 | 21.09.2004 | 1.75% | - | 25 | - | 0 | 1 |
| 14.12.2017 | 1.25% | 1.50% | 25 | - | 1 | 0 | 10.08.2004 | 1.50% | - | 25 | - | 0 | 1 |
| 15.06.2017 | 1.00% | 1.25% | 25 | - | 1 | 0 | 30.06.2004 | 1.25% | - | 25 | - | 0 | 1 |
| 16.03.2017 | 0.75% | 1.00% | 25 | - | 1 | 0 | 25.06.2003 | 1.00% | - | -25 | - | 0 | 1 |
| 15.12.2016 | 0.50% | 0.75% | 25 | - | 1 | 0 | 06.11.2002 | 1.25% | - | -50 | - | 0 | 1 |
| 17.12.2015 | 0.25% | 0.50% | 25 | - | 1 | 0 | 11.12.2001 | 1.75% | - | -25 | - | 0 | 1 |
| 16.12.2008 | 0.00% | 0.25% | -75 | -100 | 0 | 1 | 06.11.2001 | 2.00% | - | -50 | - | 0 | 1 |
| 29.10.2008 | 1.00% | - | -50 | - | 0 | 1 | 02.10.2001 | 2.50% | - | -50 | - | 0 | 1 |
| *08.10.2008 | 1.50% | - | -50 | - | 1 | 0 | *17.09.2001 | 3.00% | - | -50 | - | 0 | 1 |
| 30.04.2008 | 2.00% | - | -25 | - | 0 | 1 | 21.08.2001 | 3.50% | - | -25 | - | 0 | 1 |
| 18.03.2008 | 2.25% | - | -75 | - | 0 | 1 | 27.06.2001 | 3.75% | - | -25 | - | 0 | 1 |
| 30.01.2008 | 3.00% | - | -50 | - | 0 | 1 | 15.05.2001 | 4.00% | - | -50 | - | 0 | 1 |
| *22.01.2008 | 3.50% | - | -75 | - | 1 | 0 | *18.04.2001 | 4.50% | - | -50 | - | 0 | 1 |
| 11.12.2007 | 4.25% | - | -25 | - | 0 | 1 | 20.03.2001 | 5.00% | - | -50 | - | 0 | 1 |
| 31.10.2007 | 4.50% | - | -25 | - | 0 | 1 | 31.01.2001 | 5.50% | - | -50 | - | 0 | 1 |
| 18.09.2007 | 4.75% | - | -50 | - | 0 | 1 | *03.01.2001 | 6.00% | - | -50 | - | 0 | 1 |
| 29.06.2006 | 5.25% | - | 25 | - | 0 | 1 | 16.05.2000 | 6.50% | - | 50 | - | 0 | 1 |
| 10.05.2006 | 5.00% | - | 25 | - | 0 | 1 | 21.03.2000 | 6.00% | - | 25 | - | 0 | 1 |
| 28.03.2006 | 4.75% | - | 25 | - | 0 | 1 | 02.02.2000 | 5.75% | - | 25 | - | 0 | 1 |
| 31.01.2006 | 4.50% | - | 25 | - | 0 | 1 | 16.11.1999 | 5.50% | - | 25 | - | 0 | 1 |
| 13.12.2005 | 4.25% | - | 25 | - | 0 | 1 | 24.08.1999 | 5.25% | - | 25 | - | 0 | 1 |
| 01.11.2005 | 4.00% | - | 25 | - | 0 | 1 | 30.06.1999 | 5.00% | - | 25 | - | 0 | 1 |
| 20.09.2005 | 3.75% | - | 25 | - | 0 | 1 | 17.11.1998 | 4.75% | - | -25 | - | 0 | 1 |
| 09.08.2005 | 3.50% | - | 25 | - | 0 | 1 | *15.10.1998 | 5.00% | - | -25 | - | 0 | 1 |
| 30.06.2005 | 3.25% | - | 25 | - | 0 | 1 | 29.09.1998 | 5.25% | - | -25 | - | 0 | 1 |
| 03.05.2005 | 3.00% | - | 25 | - | 0 | 1 | | | | | | | |
| 22.03.2005 | 2.75% | - | 25 | - | 0 | 1 | | | | | | | |
| 02.02.2005 | 2.50% | - | 25 | - | 0 | 1 | | | | | | | |

$\Delta Tgt$ are given in basis points * indicates a target rate adjustment following an unscheduled meeting

**Table A.2** – Target rate adjustments since January 1, 1998.

**Figure A.1** – Target rate and treasury yields of all maturities.

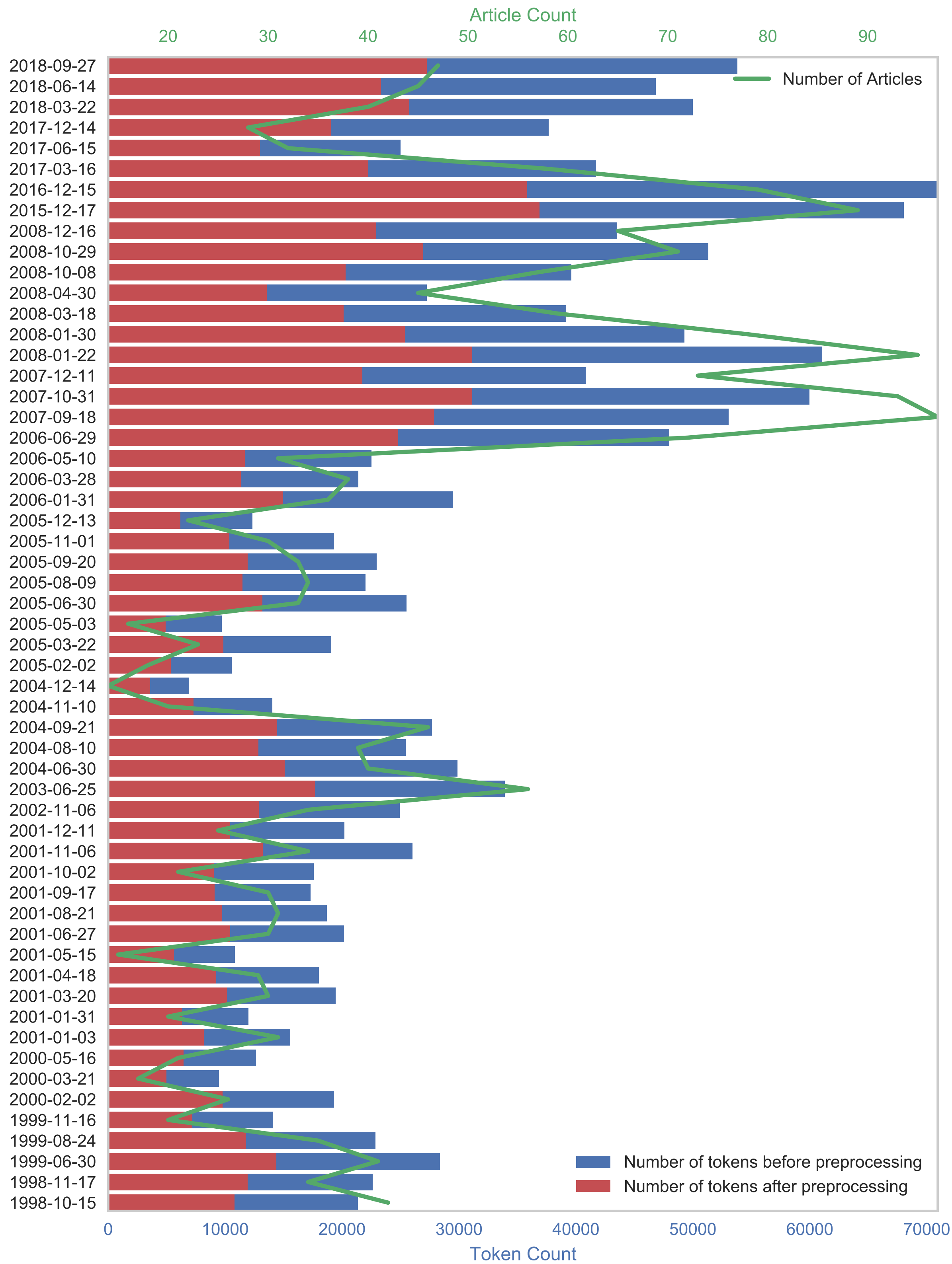


Appendix B

# Text Data

Figure B.1 illustrates the distribution of the token and article count of the text data across policy days. After tokenization, the text data collection encompasses more than 1'600'000 tokens (red bars). After the removal of stop words, the token count drops to roughly 1'100'000 tokens. Subsequently, the non-alphabetic characters, especially the punctuations marks, are removed and the final count amounts to approximately 850'000 tokens (blue bars). Note that the blue bars illustrate the token count after preprocessing but before feature engineering.

**Figure B.1** – Token and article count for every policy day.



Appendix C

# Topic Model Output

## C.1 PLSA without background topic

Figure C.1 shows the classification of policy days according to the weight of narrative one and narrative two across all articles of a specific policy day. The classification utilizes the topic model output without background topic, that is with $\lambda_B = 0$.

Figures C.2 and C.3 show the weight of narrative one of each document for $\lambda_B = 0$, split according to policy days. The number of documents attributed to each day vary across policy days, as indicated by the number below each graph. Ideally, all documents of a policy day would have either a very high or a very low weight of topic 1. This would indicate a clean split of days that concentrate on narrative one and days that are influenced by narrative two. As it is, the weights of he narratives are mixed across articles and days and the final classification of any given day is the result of averaging across the different weights.

## C.2 PLSA with background topic

Tables C.1–C.3 show the 20 top words of every topic depending on $\lambda_B$. The use of colors emphasizes the movement of words of interest given a varying $\lambda_B$.

Figures C.4 and C.7 illustrate the classification of policy days for $\lambda_B = 0.1$ and $\lambda_B = 0.9$, respectively. Analogously, Figures C.5–C.6 and C.8–C.9 show the weight of narrative one of each document for $\lambda_B = 0.1$ and $\lambda_B = 0.9$, split according to policy days.

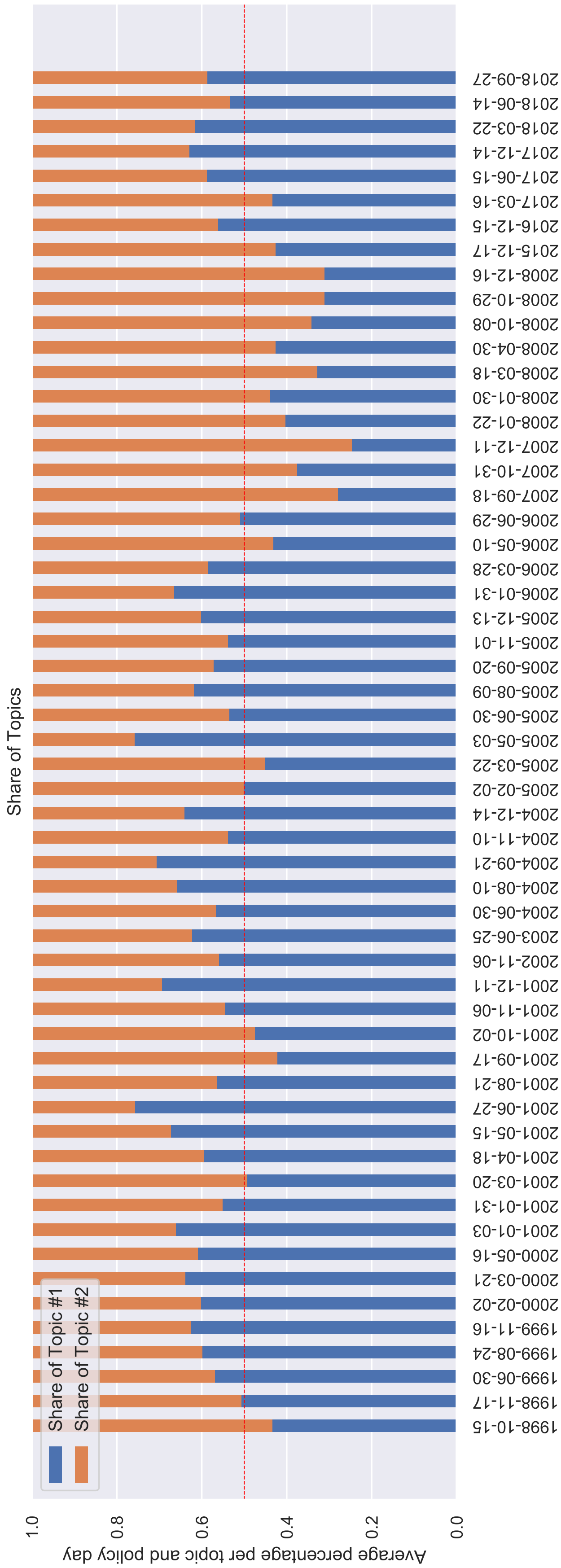



**Figure C.1** – Policy day classification according to PLSA with $\lambda_B = 0$.

**Figure C.2** – Document classification according to PLSA with $\lambda_B = 0$- Part I.

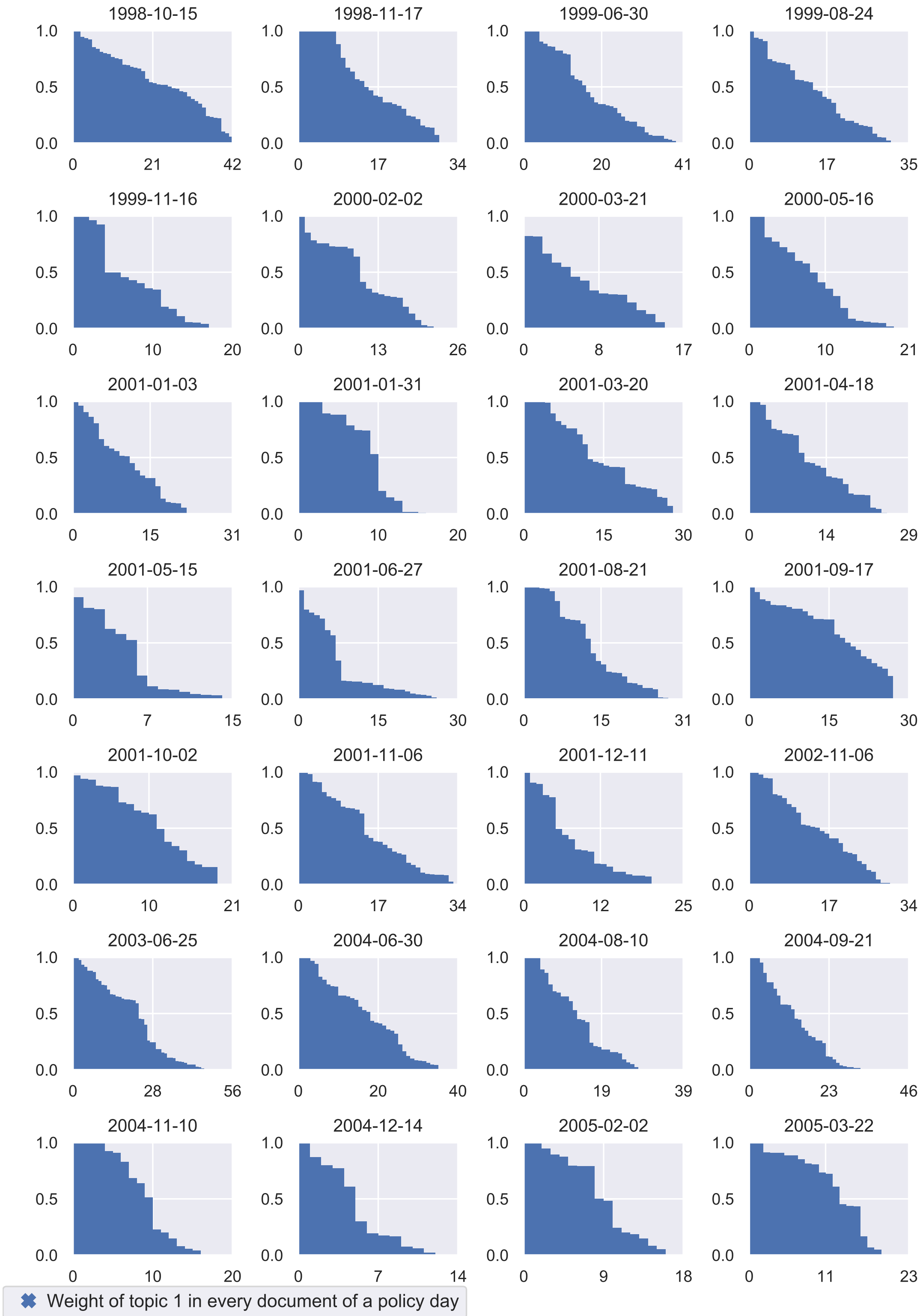
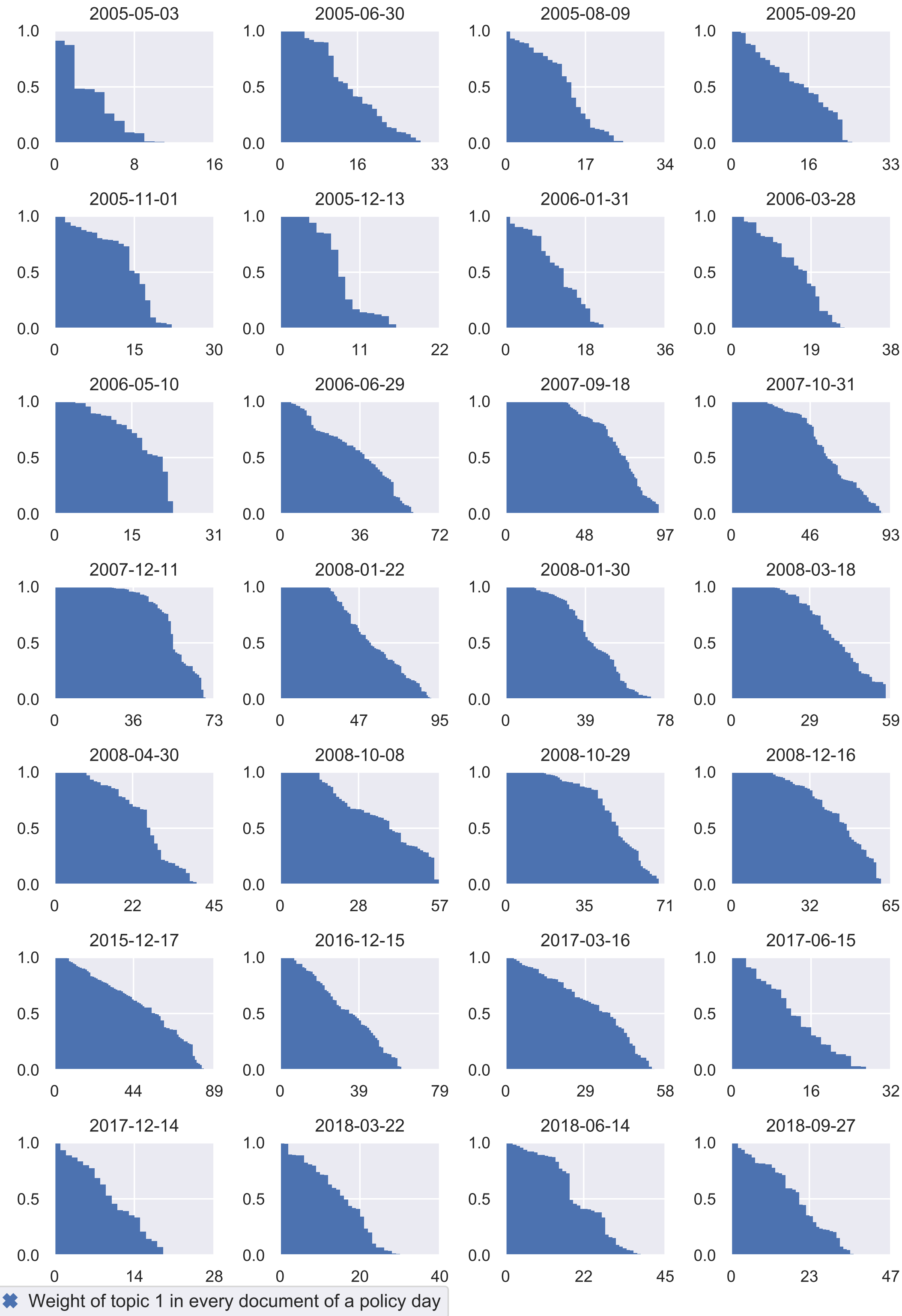


**Figure C.3** – Document classification according to PLSA with $\lambda_B = 0$ - Part II.



| Topic 1 | | | | | |
|---|---|---|---|---|---|
| $\lambda_B = 0.1$ | $\lambda_B = 0.2$ | $\lambda_B = 0.3$ | $\lambda_B = 0.4$ | $\lambda_B = 0.5$ | $\lambda_B = 0.6$ |
| said | year | year | year | year | year |
| year | said | said | said | percent | percent |
| percent | percent | percent | percent | market | market |
| market | market | market | market | said | dollar |
| rate | rate | dollar | dollar | dollar | said |
| dollar | dollar | rate | investors | investors | bank |
| cut | cut | bank | bank | bank | investors |
| bank | bank | cut | cut | cut | cut |
| rates | investors | investors | per cent | per cent | rate |
| investors | rates | rates | bond | bond | per cent |
| fed | interest rate | interest rate | interest rate | yield | yield |
| interest rate | fed | per cent | yield | rose | bond |
| per cent | per cent | bond | rate | interest rate | rose |
| federal reserve | bond | yield | us | us | interest rate |
| bond | federal reserve | markets | markets | markets | rates |
| markets | yield | federal reserve | rose | index | index |
| yield | markets | rose | expected | expected | fell |
| expected | rose | expected | index | fell | markets |
| interest rate | expected | index | fell | stocks | stocks |
| rose | index | us | stocks | month | month |

**Table C.1** – Topic 1 produced by PLSA depending on $\lambda_B$.

| Topic 1 | | | Topic 2 | | |
|---|---|---|---|---|---|
| $\lambda_B = 0.7$ | $\lambda_B = 0.8$ | $\lambda_B = 0.9$ | $\lambda_B = 0.1$ | $\lambda_B = 0.2$ | $\lambda_B = 0.3$ |
| year | dollar | dollar | fed | fed | fed |
| dollar | year | investors | said | said | said |
| percent | percent | yield | inflation | inflation | inflation |
| market | market | bank | economy | economy | economy |
| investors | investors | rose | rate | rate | rate |
| bank | bank | per cent | would | would | would |
| cut | yield | fell | year | economic | economic |
| yield | rose | trading | economic | rates | rates |
| said | cut | year | rates | year | growth |
| per cent | per cent | bond | growth | growth | year |
| rose | bond | index | percent | policy | policy |
| bond | fell | stocks | policy | percent | percent |
| index | index | yen | federal reserve | interest rates | could |
| fell | stocks | euro | interest rates | federal reserve | interest rates |
| us | trading | currency | could | could | federal reserve |
| stocks | euro | bonds | market | market | meeting |
| month | month | shares | meeting | meeting | statement |
| markets | us | cents | time | statement | market |
| euro | new york | bps | statement | time | time |
| trading | bonds | china | interest rate | interest rate | one |

**Table C.2** – Topic 1 and Topic 2 produced by PLSA depending on $\lambda_B$.

| Topic 2 | | | | | |
|---|---|---|---|---|---|
| $\lambda_B = 0.4$ | $\lambda_B = 0.5$ | $\lambda_B = 0.6$ | $\lambda_B = 0.7$ | $\lambda_B = 0.8$ | $\lambda_B = 0.9$ |
| fed | fed | fed | fed | fed | fed |
| said | said | inflation | inflation | inflation | inflation |
| inflation | inflation | economy | economy | economy | economy |
| economy | economy | said | said | said | economic |
| rate | rate | rate | would | economic | greenspan |
| rates | rates | would | rate | would | mr |
| would | would | economic | economic | growth | think |
| economic | economic | growth | growth | rate | growth |
| growth | growth | policy | policy | policy | would |
| policy | policy | rates | could | greenspan | bernanke |
| year | could | could | rates | statement | rate |
| percent | interest rates | meeting | meeting | meeting | chairman |
| interest rates | federal reserve | statement | statement | think | statement |
| federal reserve | statement | interest rates | time | could | policy |
| could | percent | time | think | mr | consumer |
| statement | time | federal reserve | mr | time | spending |
| meeting | year | mr | interest rates | bernanke | economists |
| time | mr | think | greenspan | chairman | washington |
| market | one | one | federal reserve | economists | time |
| one | think | greenspan | one | rates | fomc |

**Table C.3** – Topic 2 produced by PLSA depending on $\lambda_B$.

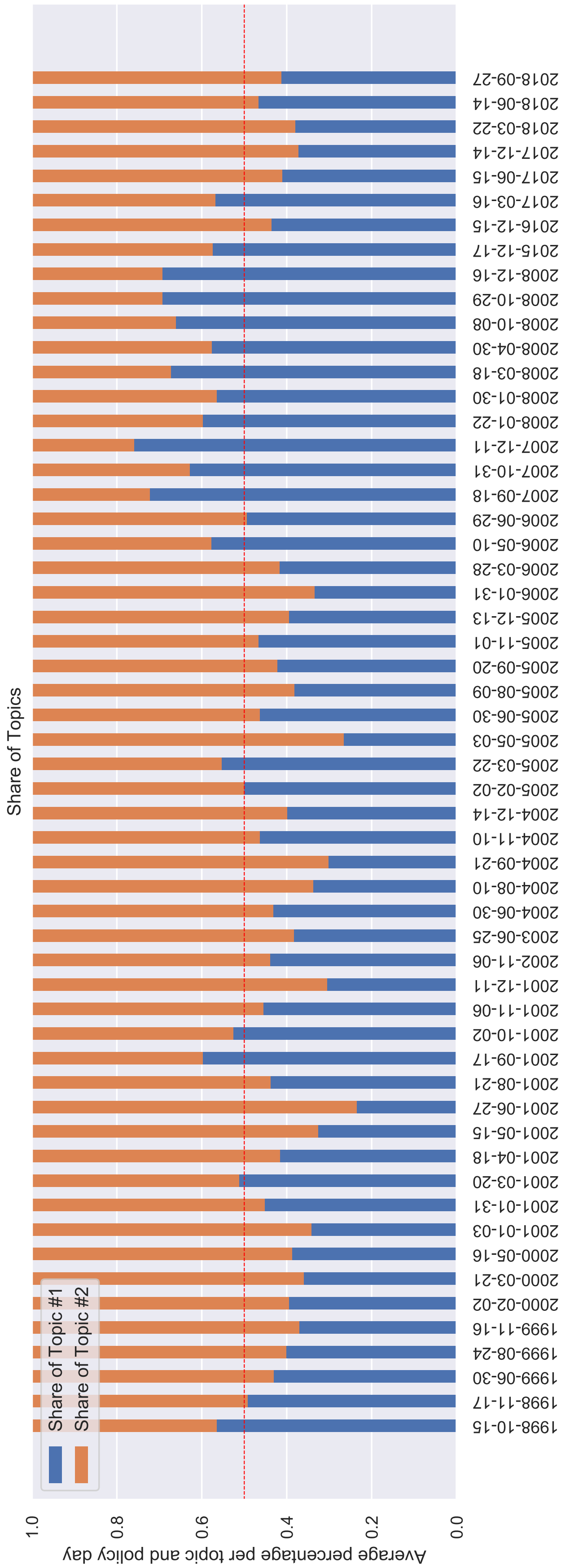



**Figure C.4** – Policy day classification according to PLSA with $\lambda_B = 0.1$.

**Figure C.5** – Document classification according to PLSA with $\lambda_B = 0.1$ - Part I.

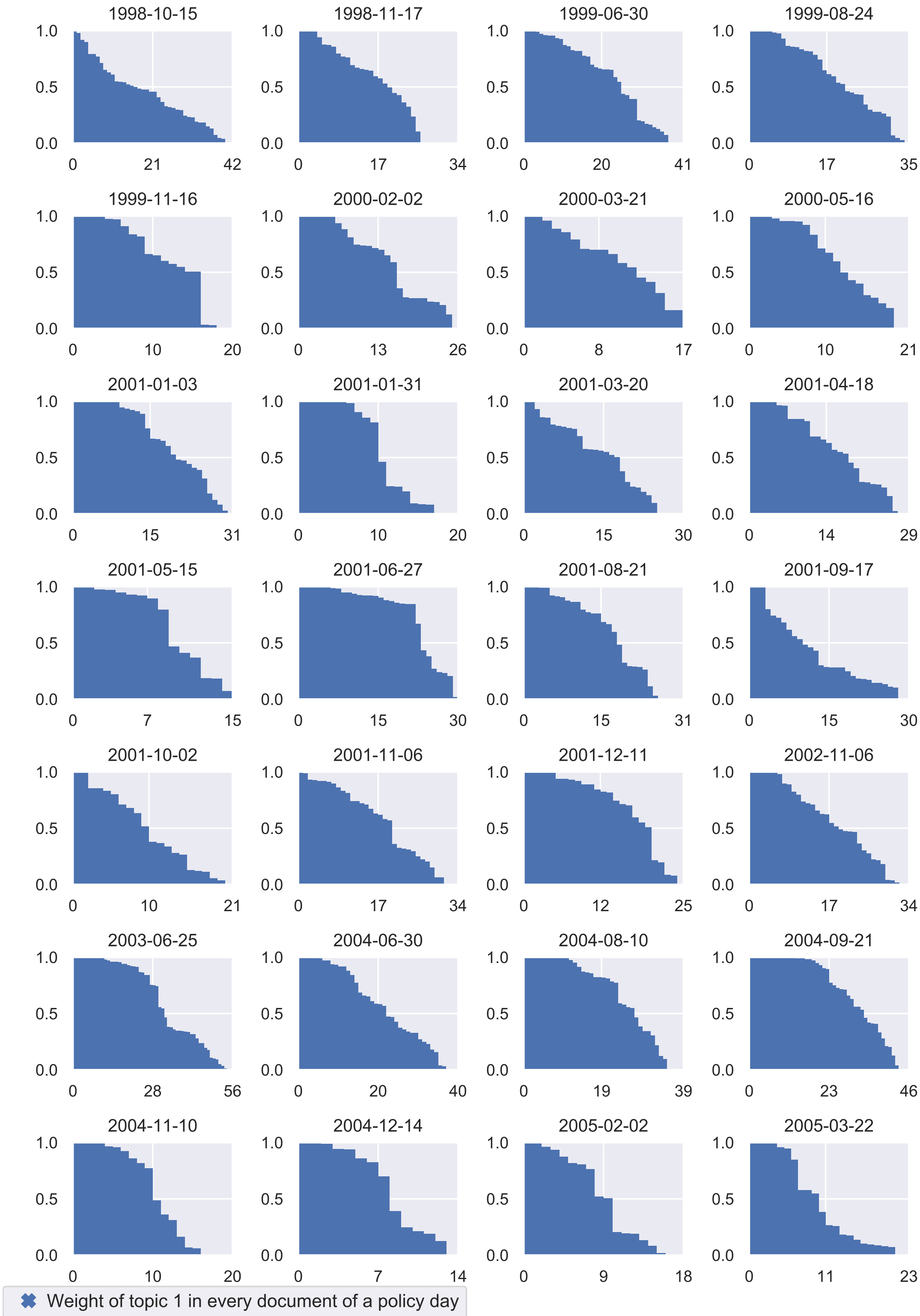
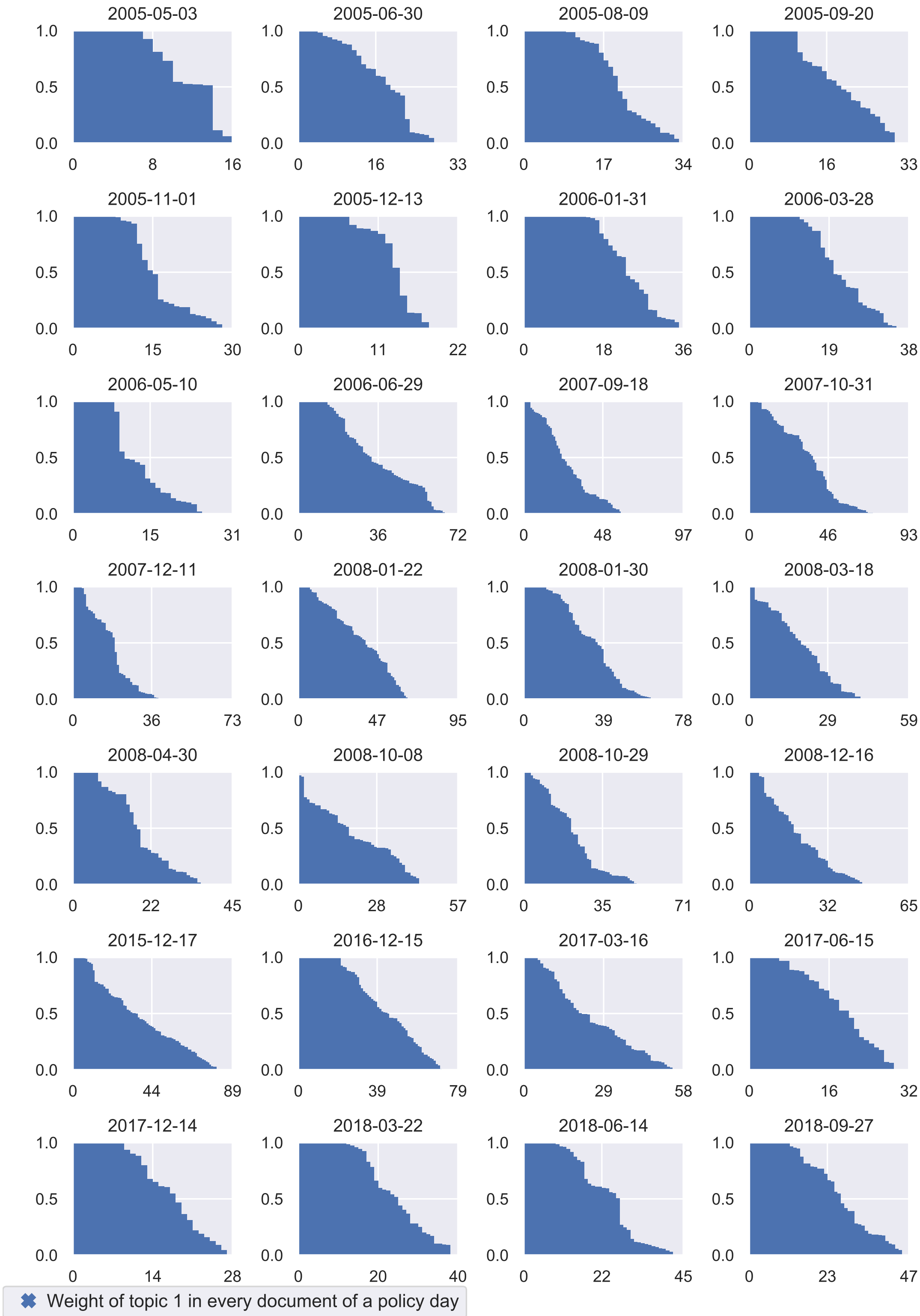


**Figure C.6** – Document classification according to PLSA with $\lambda_B = 0.1$ - Part II.



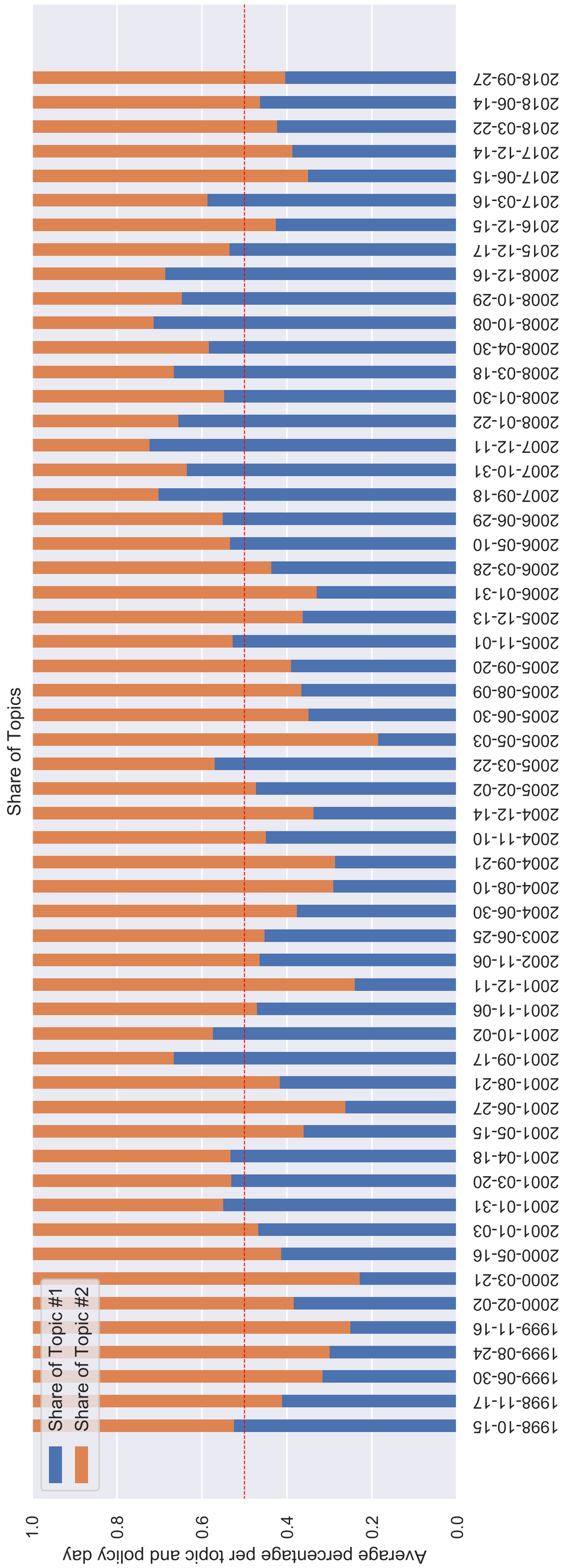



**Figure C.7** – Policy day classification according to PLSA with $\lambda_B = 0.9$.

**Figure C.8** – Document classification according to PLSA with $\lambda_B = 0.9$ - Part I.

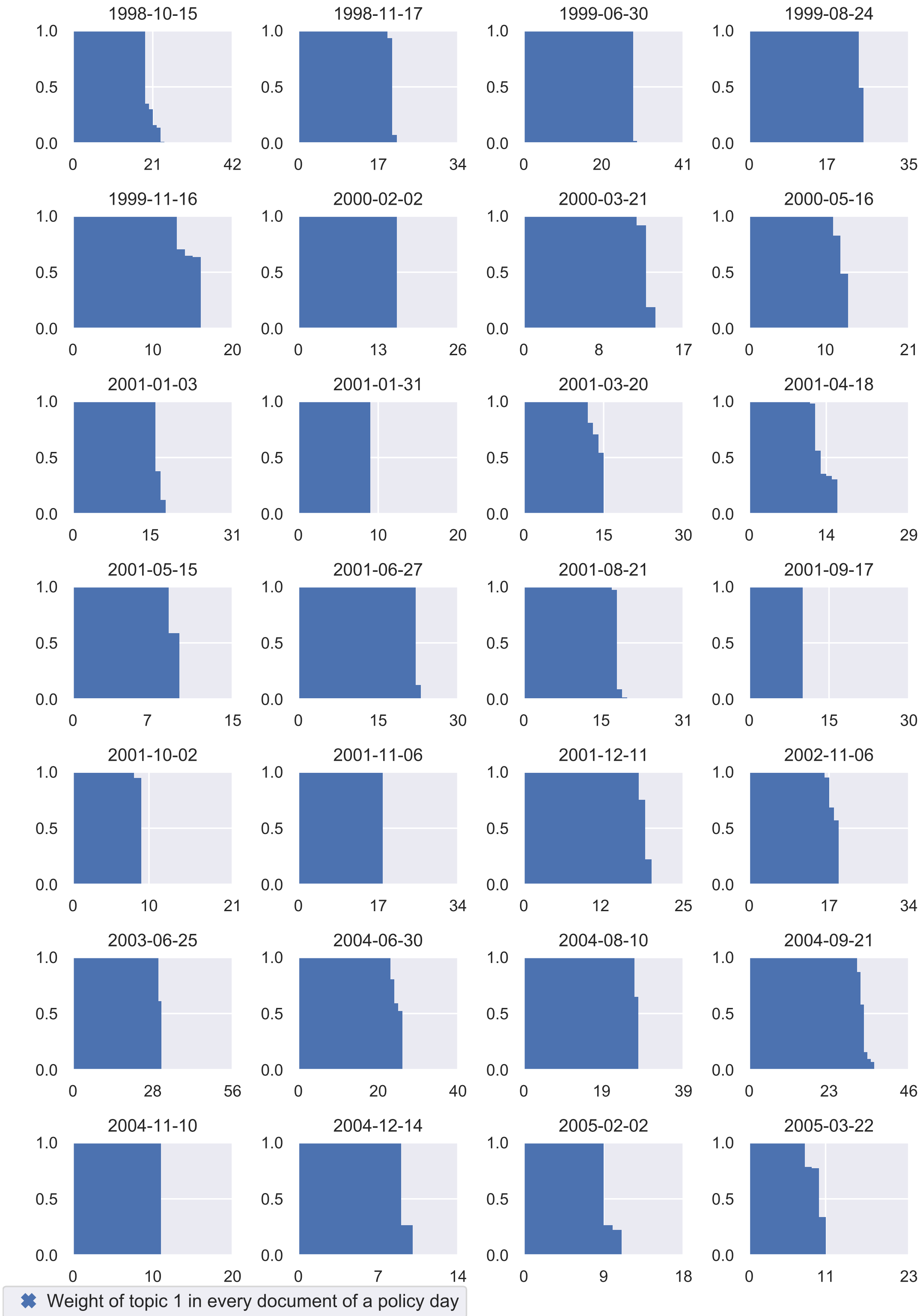
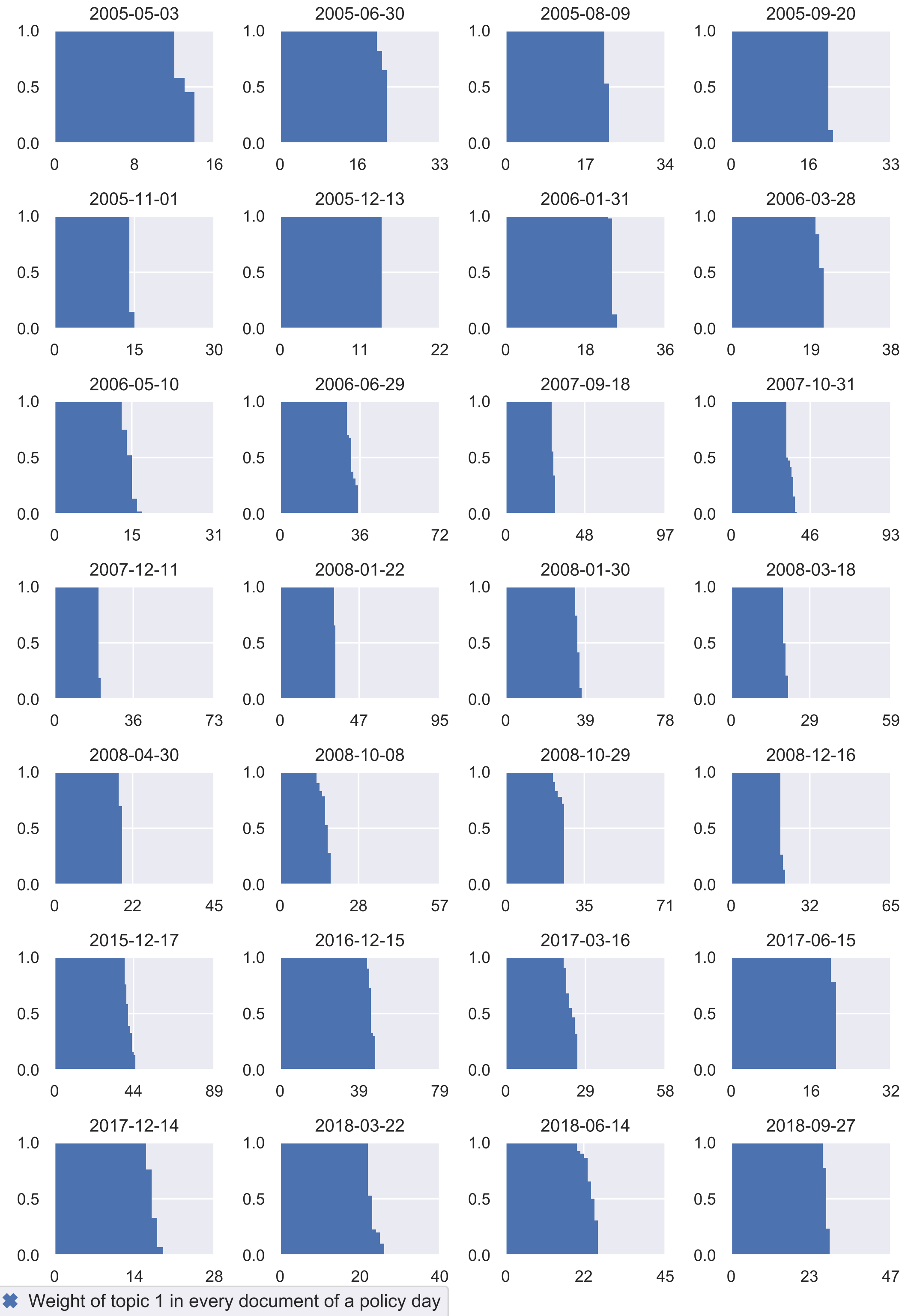


**Figure C.9** – Document classification according to PLSA with $\lambda_B = 0.9$ - Part II.



Appendix D

# Classification and Regression Results

## D.1 Classification Output

Figures D.1–D.4 show the reaction of the 10-year rate to a change in the 3-months rate, with the policy days split up according to their classification by narratives one and two given $\lambda_B = 0.1$ and $\lambda_B = 0.9$, respectively. The figures allow an eyeball assessment of the success of the classification.

Figures D.5–D.7 show the classification of the policy days that have been classified as either endogenous or exogenous by Ellingsen et al. (2003, p. 11). There are only 18 data points available for a comparison. The classification according to narratives fails to separate the policy days in a similar fashion to Ellingsen et al. Either both groups, endogenous and exogenous policy days, are classified by narrative one or, as $\lambda_B$ increases, by narrative two.

## D.2 Regression Output

Figures D.8 and D.9 show the relationship of the slope coefficients for $\lambda_B = 0.1$ and $\lambda_B = 0.9$, respectively, including the 95%-confidence interval for each $\beta_n$. Similarly, Tables D.1 and D.2 show the regression result for policy day classifications with $\lambda_B = 0.1$ and $\lambda_B = 0.9$.

**Figure D.1** – Response of the 10-year interest rate to a change in the 3-month rate on policy days influenced by Narrative one ($\lambda_B = 0.1$).

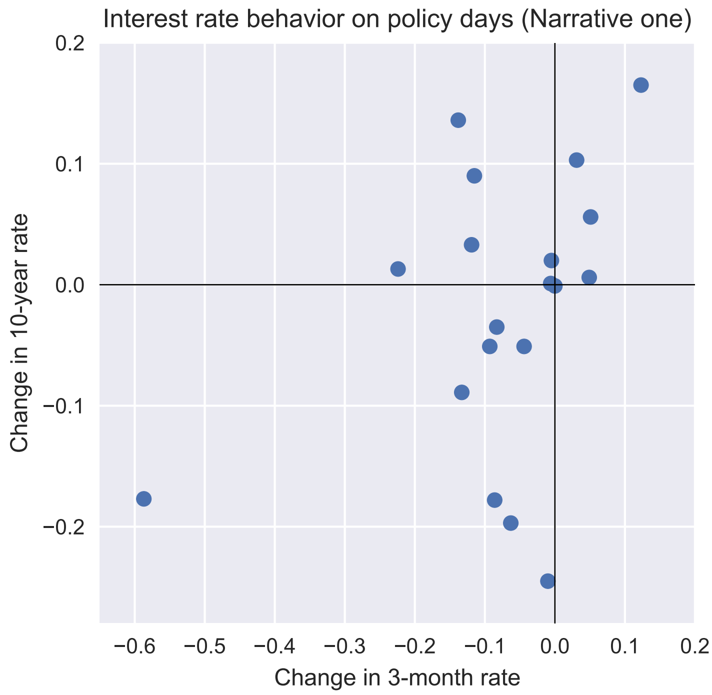



**Figure D.2** – Response of the 10-year interest rate to a change in the 3-month rate on policy days influenced by Narrative two ($\lambda_B = 0.1$).

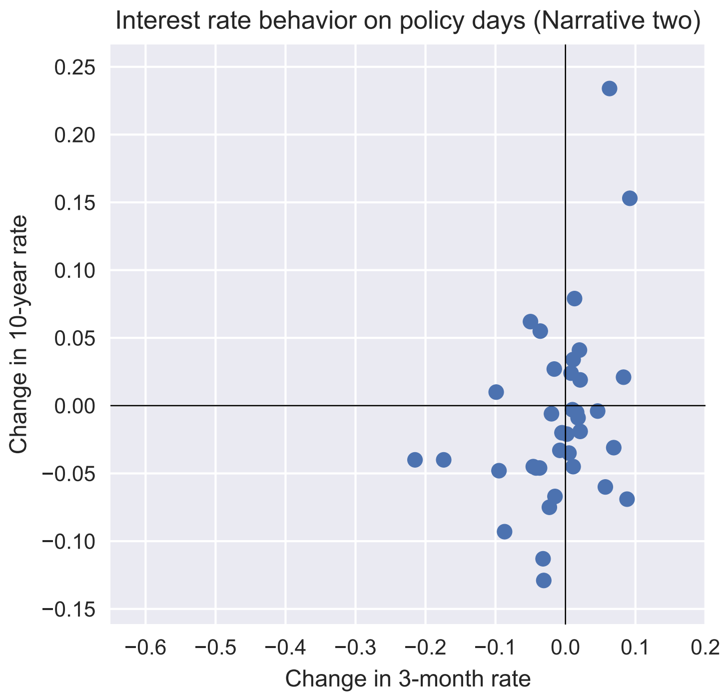


**Figure D.3** – Response of the 10-year interest rate to a change in the 3-month rate on policy days influenced by Narrative one ($\lambda_B = 0.9$).

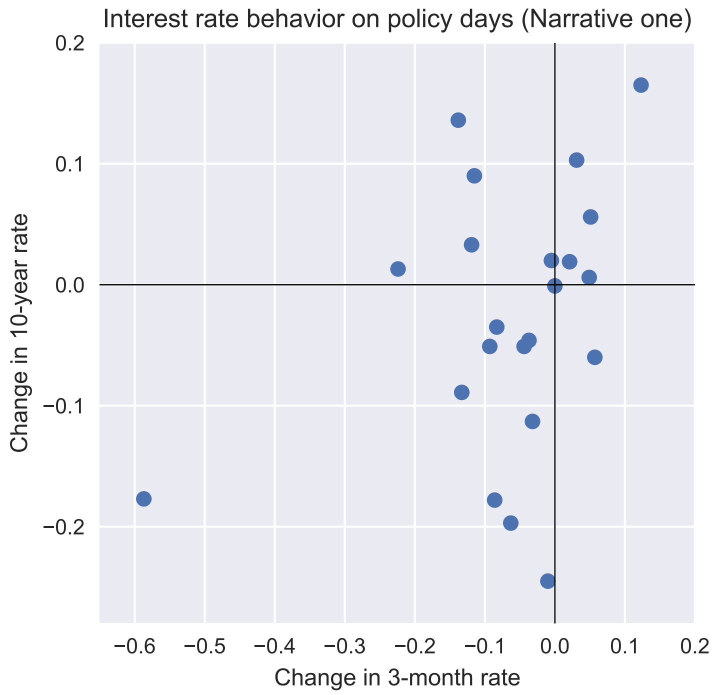



**Figure D.4** – Response of the 10-year interest rate to a change in the 3-month rate on policy days influenced by Narrative two ($\lambda_B = 0.9$).

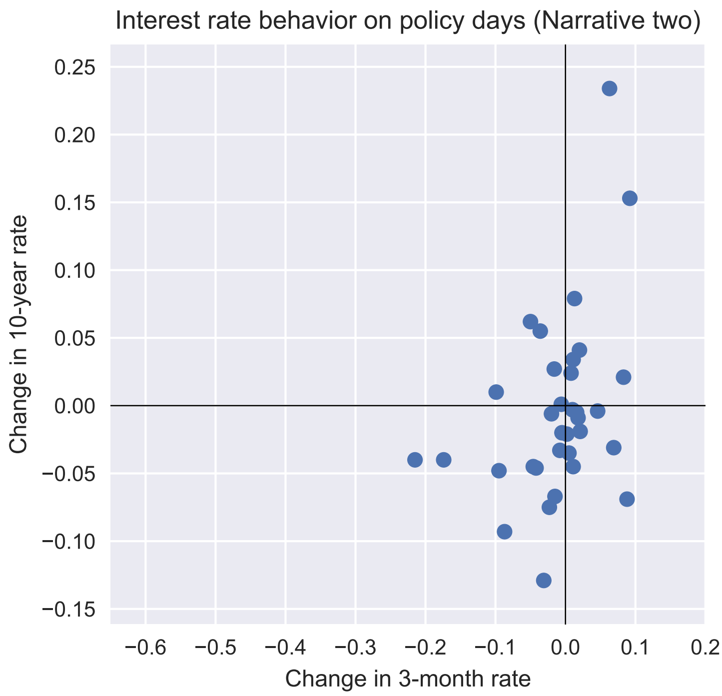


**Figure D.5** – Boxplot showing the weight of narrative one (without background topic) on policy days classified as endogenous or exogenous by Ellingsen et al. (2003, p. 11).

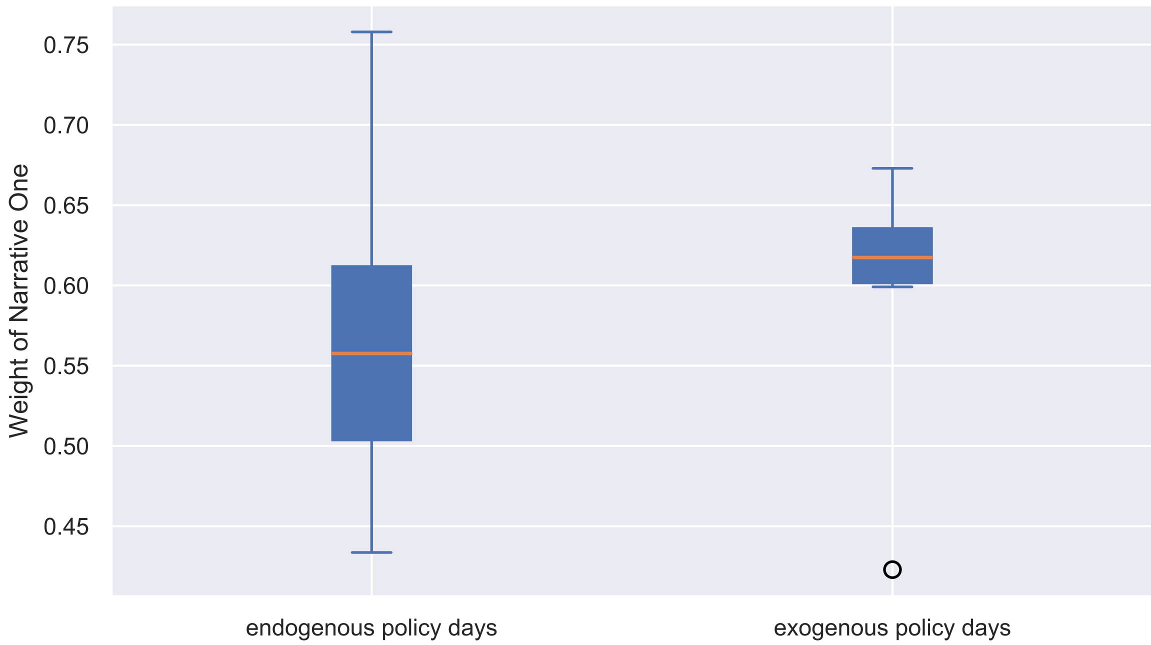



**Figure D.6** – Boxplot showing the weight of narrative one (with $\lambda_B = 0.1$) on policy days classified as endogenous or exogenous by Ellingsen et al. (2003, p. 11).

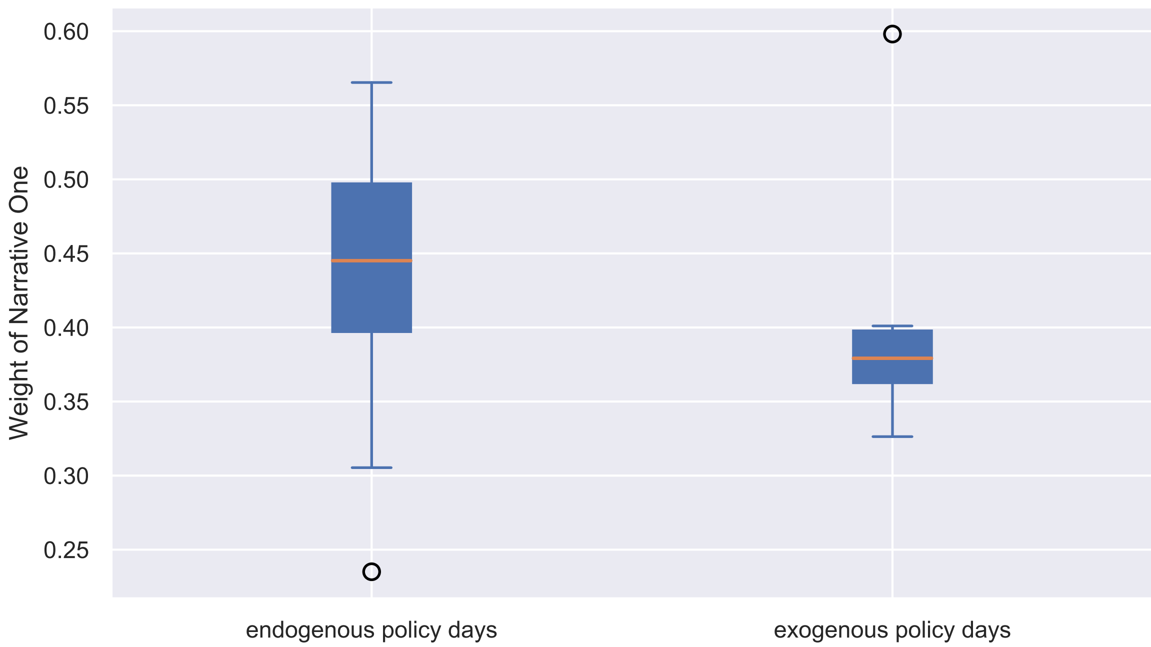


**Figure D.7** – Boxplot showing the weight of narrative one (with $\lambda_B = 0.9$) on policy days classified as endogenous or exogenous by Ellingsen et al. (2003, p. 11).

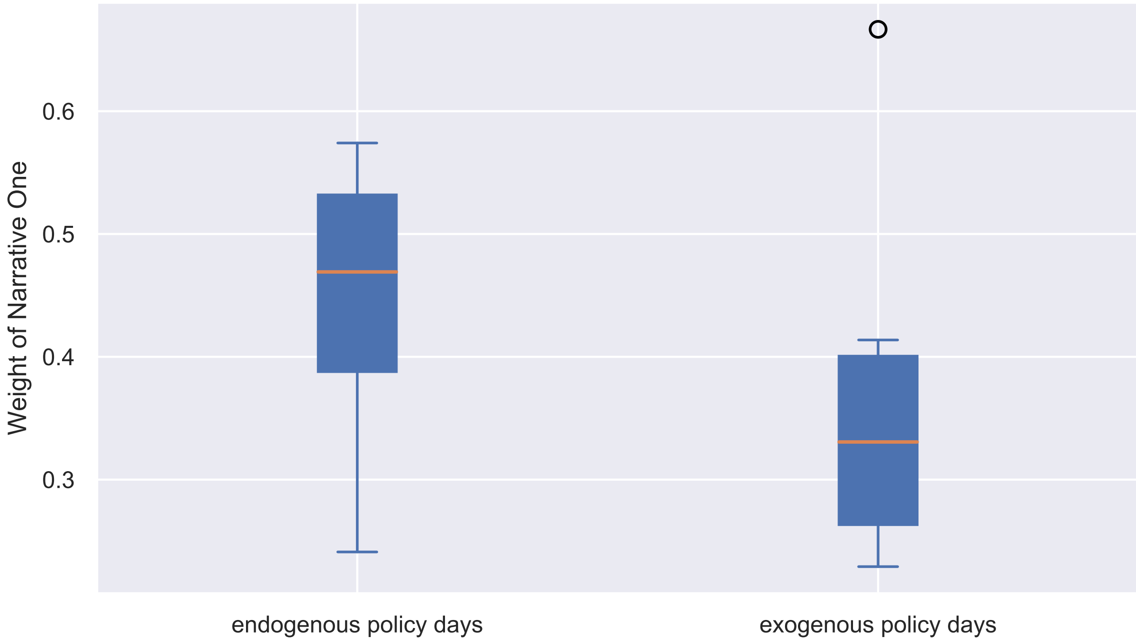



**Figure D.8** – Regression coefficients for narrative one and two with $\lambda_B = 0.1$.

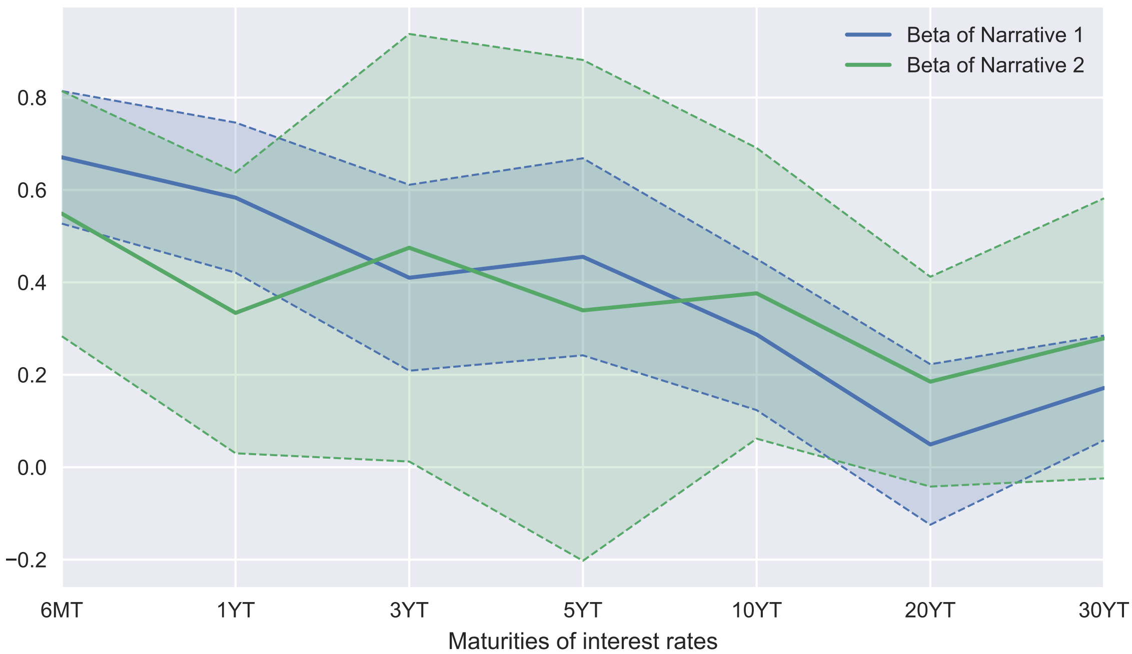
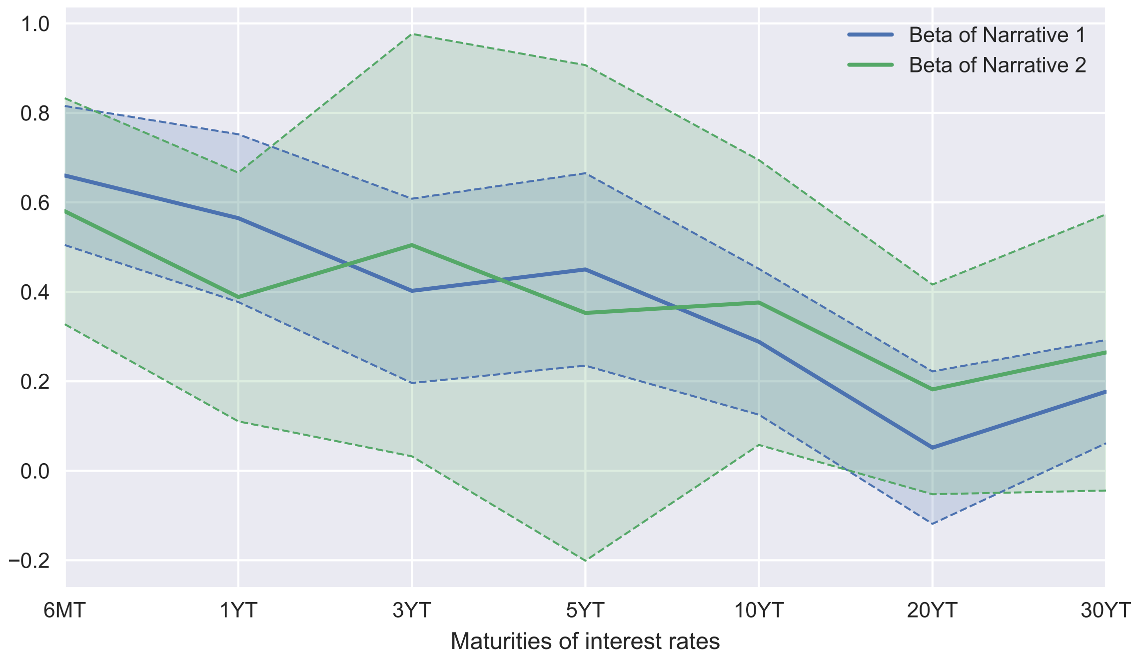


**Figure D.9** – Regression coefficients for narrative one and two with $\lambda_B = 0.9$.




| | 6m | 1y | 3y | 5y | 10y | 20y | 30y |
|---|---|---|---|---|---|---|---|
| $\alpha_n$ | -0.00 | -0.00 | -0.00 | -0.00 | -0.00 | -0.00 | -0.00 |
| | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) |
| $\beta_n^{NP}$ | 0.36** | 0.30** | 0.28** | 0.25** | 0.18** | 0.13** | 0.12** |
| | (0.05) | (0.04) | (0.04) | (0.03) | (0.03) | (0.02) | (0.02) |
| $\beta_n^{N1}$ | 0.67** | 0.58** | 0.41** | 0.46** | 0.29** | 0.05 | 0.17** |
| | (0.07) | (0.08) | (0.10) | (0.11) | (0.08) | (0.09) | (0.06) |
| $\beta_n^{N2}$ | 0.55** | 0.33* | 0.47* | 0.34 | 0.38* | 0.19 | 0.28 |
| | (0.14) | (0.15) | (0.24) | (0.28) | (0.16) | (0.12) | (0.15) |
| $R^2$ | 0.30 | 0.20 | 0.08 | 0.05 | 0.03 | 0.02 | 0.02 |
| $\beta_n^{N1} = \beta_n^{N2}$ | 0.12 | 0.25 | -0.06 | 0.12 | -0.09 | -0.14 | -0.11 |

**Table D.1** – Regression results (classification with $\lambda_B = 0.1$).

| | 6m | 1y | 3y | 5y | 10y | 20y | 30y |
|---|---|---|---|---|---|---|---|
| $\alpha_n$ | -0.00 | -0.00 | -0.00 | -0.00 | -0.00 | -0.00 | -0.00 |
| | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) |
| $\beta_n^{NP}$ | 0.36** | 0.30** | 0.28** | 0.25** | 0.18** | 0.13** | 0.12** |
| | (0.05) | (0.04) | (0.04) | (0.03) | (0.03) | (0.02) | (0.02) |
| $\beta_n^{N1}$ | 0.66** | 0.56** | 0.40** | 0.45** | 0.29** | 0.05 | 0.18** |
| | (0.08) | (0.10) | (0.11) | (0.11) | (0.08) | (0.09) | (0.06) |
| $\beta_n^{N2}$ | 0.58** | 0.39** | 0.50* | 0.35 | 0.38* | 0.18 | 0.26 |
| | (0.13) | (0.14) | (0.24) | (0.28) | (0.16) | (0.12) | (0.16) |
| $R^2$ | 0.30 | 0.20 | 0.08 | 0.05 | 0.03 | 0.02 | 0.02 |
| $\beta_n^{N1} = \beta_n^{N2}$ | 0.08 | 0.18 | -0.10 | 0.10 | -0.09 | -0.13 | -0.09 |

**Table D.2** – Regression results (classification with $\lambda_B = 0.9$).

# Declaration of Authorship

"I hereby declare

- that I have written this thesis without any help from others and without the use of documents and aids other than those stated above;
- that I have mentioned all the sources used and that I have cited them correctly according to established academic citation rules;
- that I have acquired any immaterial rights to materials I may have used such as images or graphs, or that I have produced such materials myself;
- that the topic or parts of it are not already the object of any work or examination of another course unless this has been explicitly agreed on with the faculty member in advance and is referred to in the thesis;
- that I will not pass on copies of this work to third parties or publish them without the University's written consent if a direct connection can be established with the University of St.Gallen or its faculty members;
- that I am aware that my work can be electronically checked for plagiarism and that I hereby grant the University of St.Gallen copyright in accordance with the Examination Regulations in so far as this is required for administrative action;
- that I am aware that the University will prosecute any infringement of this declaration of authorship and, in particular, the employment of a ghostwriter, and that any such

infringement may result in disciplinary and criminal consequences which may result in my expulsion from the University or my being stripped of my degree."

| ........................................ | ........................................ |
| --- | --- |
| Location, Date | Signature |

By submitting this academic term paper, I confirm through my conclusive action that I am submitting the Declaration of Authorship, that I have read and understood it, and that it is true.