

# Deciphering Federal Reserve Communication via Text Analysis of Alternative FOMC Statements

---

Taeyoung Doh, Dongho Song, and Shu-Kuei Yang

October 2020

RWP 20-14

<http://doi.org/10.18651/RWP2020-14>

FEDERAL RESERVE BANK *of* KANSAS CITY



# Deciphering Federal Reserve Communication via Text Analysis of Alternative FOMC Statements <sup>\*</sup>

Taeyoung Doh<sup>†</sup>      Dongho Song<sup>‡</sup>      Shu-Kuei Yang<sup>§</sup>

October 6, 2020

## Abstract

We apply a natural language processing algorithm to FOMC statements to construct a new measure of monetary policy stance, including the tone and novelty of a policy statement. We exploit cross-sectional variations across alternative FOMC statements to identify the tone (for example, dovish or hawkish) and contrast the current and previous FOMC statements released after Committee meetings to identify the novelty of the announcement. We then use high-frequency bond prices to compute the surprise component of the monetary policy stance. Our text-based estimates of monetary policy surprises are not sensitive to the choice of bond maturities used in estimation, are highly correlated with forward guidance shocks in the literature, and are associated with lower stock returns after unexpected policy tightening. The key advantage of our approach is that we are able to conduct a counterfactual policy evaluation by replacing the released statement with an alternative statement, allowing us to perform a more detailed investigation at the sentence and paragraph level.

JEL Classification: E30, E40, E50, G12.

Keywords: Alternative FOMC statements, counterfactual policy evaluation, monetary policy stance, text analysis, natural language processing.

---

<sup>\*</sup>We thank John Rogers for sharing his dataset on monetary policy shock estimates with us. We are grateful for Kyungmin Kim and Suyong Song as well as seminar participants at the Federal Reserve Bank of Kansas City, the interfed Brown Bag, and the KAEA webinar who provided helpful comments. All the remaining errors are ours. The opinions expressed herein are those of the authors and do not reflect the views of the Federal Reserve Bank of Kansas City or Federal Reserve System.

<sup>†</sup>Federal Reserve Bank of Kansas City; taeyoung.doh@kc.frb.org

<sup>‡</sup>Carey Business School, Johns Hopkins University; dongho.song@jhu.edu

<sup>§</sup>Federal Reserve Bank of Kansas City; shu-kuei.x.yang@kc.frb.org

# 1 Introduction

Central banks have increasingly relied on public communications to provide guidance regarding future policy actions, e.g., [Woodford \(2005\)](#) and [Blinder et al. \(2008\)](#). The practice became more prevalent when monetary policy is constrained by the effective lower bound, see [Bernanke \(2010\)](#). In this regard, both quantitative decisions (e.g., interest rates or asset purchases) made by central banks and their qualitative descriptions of the economic factors that lead to the decisions serve as important information variables for understanding monetary policy.

While the profession has moved toward treating policy statements and speeches by central bank officials as data to be analyzed, important limitations exist in the parsing of textual content. First, quantifying the tone (between a dovish stance and a hawkish one) automatically from only the publicly released text is difficult because the tone is often relative to what could have been released. Second, even when using asset prices as instruments, identifying which part of a communication is perceived as most crucial by the markets is difficult to do because texts like statements and speeches are multi-dimensional objects. In addition, assumptions are required on the maturity structure of bond prices to translate bond price movements into particular parts of the statement. Third, it is hard to evaluate the (counterfactual) impact of alternative language in the statement on the markets within the commonly used text analysis methods largely based on word counting when the alternative language mainly changes the contextual meaning of words rather than the frequency pattern of words.

The purpose of this paper is to contribute in all three dimensions to enhance our understanding of the transmission of monetary policy to the financial markets, which is important for both policy makers and market participants. We work with the Federal Open Market Committee's (FOMC) post-meeting statements in this paper. We differ from the existing literature on two fronts in achieving our objectives. First, we refine the information in the FOMC statements using a novel natural language processing algorithm known as the Universal Sentence Encoding (USE). In contrast to the word-counting methods that ignore the local context (e.g., Term Frequency-Inverse Document Frequency (TF-IDF) and Latent Semantic Analysis (LSA)), the USE provides context-aware representation of words in the document. Second, and more importantly, we consider alternative policy statements created

by the staff of the Federal Reserve Board as well as the statement released right after the meeting in parsing of policy statements. Alternative statements are available for each FOMC meeting since March 2004 and contain a more dovish (Alt A) or a more hawkish (Alt C or Alt D) statement than the benchmark one (Alt B). The different views of economic outlook and policy prescription contained in the alternative statements provide important anchoring points for interpreting the tone of the policy statement released after the meeting.<sup>1</sup>

We provide a novel measure of the surprise component of monetary policy announcements based on our text analysis. We do this in two steps. First, we characterize the “monetary policy stance” communicated by each FOMC statement. For this, we identify the “tone” of monetary policy announcements by computing the similarities in terms of the USE representation between the released statement and the alternative statements.<sup>2</sup> We define the “novelty” of monetary policy announcements by computing the distance in terms of the USE representation between the current statement and the previous statement. By taking the product of the tone and novelty of monetary policy announcements, we obtain the monetary policy stance. The first step only relies on the text analysis of the FOMC statements. In the second step, we introduce (high-frequency) bond prices to compute the surprise component of the monetary policy stance. The expected component of the monetary policy stance is defined as a weighted average of the hawkish and dovish policy stances of the alternative statements. We back out the weight that maximizes the rank correlation of the (high-frequency) bond prices and the surprise component of monetary policy announcements (aka, monetary policy shocks), namely, the difference between the monetary policy stance and its expected component.

We then use (high-frequency) stock prices to show that a tightening policy surprise according to our measure generates a negative stock price reaction. Specifically, a positive one-standard-deviation surprise leads to a 20 basis points (bps) drop in stock prices on average. Also, we verify that our measure of monetary policy shocks is highly correlated (about 50%) with forward guidance shocks identified in the existing literature (that relies

---

<sup>1</sup>Like the FOMC transcripts, alternative statements are made public five years after they were created. Although published with a significant lag, dovish and hawkish alternative statements generally incorporate information on market expectations regarding the upcoming policy statement because they are written to surprise the market in respective directions. See [FRB \(2004\)](#) for the detailed description of statement language.

<sup>2</sup>The released statement is typically very similar to the benchmark statement created before each FOMC meeting but may not be exactly same because some phrases can change during the FOMC deliberation.

on high-frequency bond prices). Both serve as external validation of our measure.

The key advantage of our approach is that we are able to conduct a counterfactual policy evaluation by replacing the released statement with either one of the alternative statements. Specifically, we work with FOMC statements released in September 2007 and December 2010 as two illustrative examples.<sup>3</sup> The post-meeting statement in September 2007 lowered the federal funds rate target by 50 bps but it noted that “the Committee judges that some inflation risks remain, and it will continue to monitor inflation developments carefully.” We consider the counterfactual statement that would lower the federal funds rate target by 25 bps but mention “the downside risks to economic growth outweigh the upside risks to inflation.” In spite of a smaller rate cut, we find that this counterfactual statement might be a more dovish statement, leading to an 1.94% increase of stock prices instead of the 0.1% increase after the release of the post-meeting statement.<sup>4</sup> Another example is the December 2010 FOMC meeting in which the counterfactual statement suggests a more explicit time-dependent forward guidance on the future path of the federal funds rate by replacing “for an extended period” by “at least through the mid-2012” to indicate how long the interest rate would stay low. It also announces additional asset purchases. The actual released statement omits such an explicit forward guidance and the expansion of asset purchases. We find that adopting the counterfactual in December 2010 would lead to about 0.77% increase in the stock market return instead of 0.05% increase after the release of the post-meeting statement. These two episodes highlight the importance of narrative information in the Fed communication, which is hard to measure without using text-analysis tools.

**Related Literature.** Our paper is related to multiple lines of research. First, our work draws on papers that identify monetary policy shocks using high-frequency bond data, (e.g., [Gürkaynak et al. \(2005\)](#), [Swanson \(2017\)](#), [Nakamura and Steinsson \(2018\)](#), [Bu et al. \(2020\)](#), [Bauer and Swanson \(2020\)](#), [Hoesch et al. \(2020\)](#)). Our empirical finding on the negative stock market response to an unexpected monetary policy tightening is consistent with [Bu et](#)

---

<sup>3</sup>See the appendix for the detailed description of alternative statements.

<sup>4</sup>We regress stock returns on our measure of monetary policy surprises and obtain the respective OLS coefficient estimates. We replace the monetary policy stance with the counterfactual one and subtract the bond price-implied expected monetary policy stance to compute the counterfactual monetary policy surprise component. We multiply the counterfactual monetary policy surprise component to the OLS slope coefficient to assess the counterfactual impact of alternative policy prescription. It is important to note that we are only replacing one data point (that corresponds to the September 2007 FOMC statement release date) while keeping all else equal in this exercise.

al. (2020), Bauer and Swanson (2020), Hoesch et al. (2020) which imply that the FOMC’s communication was largely effective in inducing the intended asset market responses during our sample period (from March 2004 to December 2014). Hoesch et al. (2020) and Lunsford (2020) find that the stock market response coefficient was unstable over time but became largely negative to policy tightening since 2004. Since this timing coincides with the creation of alternative statements (March 2004), we suspect that the intense monetary policy deliberation focused on effective communication might account for the timing of the break.<sup>5</sup> However, our paper is distinguished from all these papers in our ability to evaluate the counterfactual implications of alternative policy prescriptions.

Second, our work is also related to the increasingly popular application of text analysis in economics and finance, (e.g., Gentzkow et al. (2019), Ke et al. (2019), Hansen et al. (2017), Schonhardt-Bailey (2013), Shapiro and Wilson (2019), Jegadeesh and Wu (2017), Meade and Acosta (2015), Giavazzi et al. (2020), Lucca and Trebbi (2009)). In particular, Ke et al. (2019) propose a method to combine new information and sentiment scores from news articles to predict stock returns. We adapt their method to our measure of monetary policy stance. Nonetheless our methodology is differentiated from most of these papers in the literature because we do not apply methods based on word-counting.<sup>6</sup> We combine a context-aware representation of the text with alternative FOMC statements to better identify the tone of the released FOMC statement. In this paper, we highlight the importance of capturing the contextual meaning using the alternative FOMC statements prepared for the October 2013 meeting as an example. We also demonstrate that our text-based policy surprises are not sensitive to the choice of target maturities used in the statistical factor analysis popular in the existing literature and highly correlated with different estimates (e.g., Nakamura and Steinsson (2018) and Bu et al. (2020)).

Outline of the structure of the paper is as follows. Section 2 describes our natural language processing technique and identification scheme of monetary policy surprises using alternative statements. Section 3 discusses empirical results and policy implications. Section 4

---

<sup>5</sup>Lunsford (2020) argues that the FOMC started to issue a more explicit policy inclination from the August 2003 meeting statement and that it was a break point in the stock market response to FOMC announcements. While we cannot disentangle our story from his interpretation because no alternative statement exists before the March 2004, our story is not inconsistent with his finding.

<sup>6</sup>The exception is Giavazzi et al. (2020). who use the Doc2vec algorithm. The USE is distinct from the Doc2vec in the heavy use of the “self-attention” channel that is known to capture richer patterns of the contextual meaning of the text.

concludes.

## 2 Text-based Identification of Monetary Policy Stance

The recent development in the natural language processing provides tools to better capture the contextual meaning of a word in a text. We rely on the USE to quantify information in texts. The encoding algorithm was pre-trained based on the Stanford Natural Language Inference (SNLI) dataset. Under the USE, the entire statement is encoded as a large-dimensional vector capturing various features of the whole text including the context between words or sentences. We highlight its features in this section relative to the commonly used word counting method but leave a more complete description of technical details to the appendix. We then provide the details about the FOMC statements and explain how we identify the monetary policy stance from the alternative statements using the USE. Finally, we explain how we leverage high-frequency asset prices to measure the surprise component of monetary policy stance.

### 2.1 Universal sentence encoding versus word-counting methods

[Cer et al. \(2018\)](#) describes two versions of the USE; 1) deep averaging of word embeddings, 2) transformer-based approach using the self-attention channel. We apply the transformer-based version of the USE to calculate the similarity between texts because the self-attention channel is powerful in capturing the context-dependent meaning of sentences as we see in the following example. The USE is able to capture the dependencies between even distant words by training deep neural networks that can recognize complex dependencies of different words based on large corpora. Hence, it can score the similarity between texts in a more sensible way. For example, imagine that there are two sentences consisting of  $n_1$  and  $n_2$  words respectively:

$$\left\{ S_1 = (w_{1,1}, \dots, w_{1,n_1}) \quad , \quad S_2 = (w_{2,1}, \dots, w_{2,n_2}) \right\} \quad (1)$$

$$\downarrow$$

$$\left\{ U_1 = (U_{1,1}, \dots, U_{1,512}) \quad , \quad U_2 = (U_{2,1}, \dots, U_{2,512}) \right\}.$$

USE will find out numerical representations of  $S_1$  and  $S_2$  by two 512 dimensional vectors ( $U_1$  and  $U_2$ ) using a deep neural network architecture. The embedding representation is trained to perform a variety of tasks such as text classification similar to humans and predicting some part of the text based on the rest of it. USE is available through **Google TensorFlow**. We calculate the similarity between the two texts based on the cosine similarity between two embedding vectors:

$$\text{Sim}_{\text{USE}}(\text{Text}_1, \text{Text}_2) = \text{cosine}(U_1, U_2) = \frac{U_1' U_2}{\sqrt{U_1' U_1} \sqrt{U_2' U_2}}. \quad (2)$$

Notice that we are not restricting on pre-fixed features (e.g., the frequency of overlapping words) of the text to calculate sentence embeddings and similarity scores. The training process of embedding representations capture rich features of the text not necessarily confined to the frequency of words. This is the main difference of the USE from methods relying on word-counting that we describe below.

**Term Frequency-Inverse Document Frequency.** We provide a comparison with the Term Frequency-Inverse Document Frequency (TF-IDF) method. Here, similarities between multiple ( $N$ ) documents are determined by the frequency of words that show up in all of these documents. Specifically,<sup>7</sup>

$$W_{i,j} = \frac{n_{i,j}}{\sqrt{\sum_k n_{k,j}^2}} \ln\left(\frac{N+1}{df_j+1} + 1\right), \quad (3)$$

$$\text{Sim}_{\text{TF-IDF}}(\text{Text}_1, \text{Text}_2) = \text{cosine}(W_{\cdot,1}, W_{\cdot,2}).$$

where  $n_{i,j}$  is the count of the  $j$ -th word in the  $i$ -th document and  $df_j$  is the number of documents that contain the  $j$ -th word. The main problem of this method is that word counting does not consider the semantic similarity between different words and the algorithm cannot be trained to incorporate the contextual meaning.

**Latent Semantic Analysis.** A more sophisticated word counting method is available known as the Latent Semantic Analysis (LSA). LSA considers the high co-frequency of words in calculating the similarity score between texts. Instead of  $W_{\cdot,1}$  and  $W_{\cdot,2}$ , LSA uses low-dimensional objects obtained by the singular value decomposition of  $W = [W_{\cdot,1}, W_{\cdot,2}]$  to

---

<sup>7</sup>We follow the default method used in Python.



Table 1: Sentence similarity

	TF-IDF	LSA	USE
$\text{Sim}(S_1, S_2)$	0	0	0.91
$\text{Sim}(S_1, S_3)$	0.78	0.78	0.28

calculate the similarity between texts.<sup>8</sup> Specifically,

$$W = U\Sigma V', \quad (4)$$

$$\text{Sim}_{\text{LSA}}(\text{Text}_1, \text{Text}_2) = \text{cosine}(X_{\cdot,1}, X_{\cdot,2}), \quad X = U_{\cdot,1:2}\Sigma_{1:2,1:2}.$$

By rotating term frequency vectors to maximize the co-frequency of words across multiple documents, the LSA extracts representations that highlight the co-frequency of words used in different documents. For this reason, it is widely used in identifying a few key topics from a large number of texts. However, like the TF-IDF, it does not take into account complex dependencies between different words beyond the co-frequency, which is important for understanding semantic similarity.

**A simple illustration.** We illustrate the advantage of the USE in capturing the contextual meaning by comparing the similarity between the following sentences. We repeat the same exercise with TF-IDF and LSA for comparison.

$(S_1)$  How old are you?

$(S_2)$  What is your age?

$(S_3)$  How are you?

It is obvious that  $S_1$  and  $S_2$  are asking the same question, whereas  $S_3$  is not. Hence, the ideal classifier should recognize that  $S_1$  is more similar to  $S_2$  than  $S_3$ . However, the similarity score under the TF-IDF or LSA provides an opposite ranking whereas the USE provides a more sensible similarity score. We highlight the results in Table 1.

---

<sup>8</sup>The dimension reduction is especially powerful when we try to capture the common theme from large text corpora although it does not matter in the simple example consisting of two sentences.

Table 2: Alternative FOMC statement similarity: October 2013 FOMC Meeting

	TF-IDF	USE
$\text{Sim}(FOMC_{A,t}, FOMC_t)$	0.975	0.895
$\text{Sim}(FOMC_{C,t}, FOMC_t)$	0.972	0.990

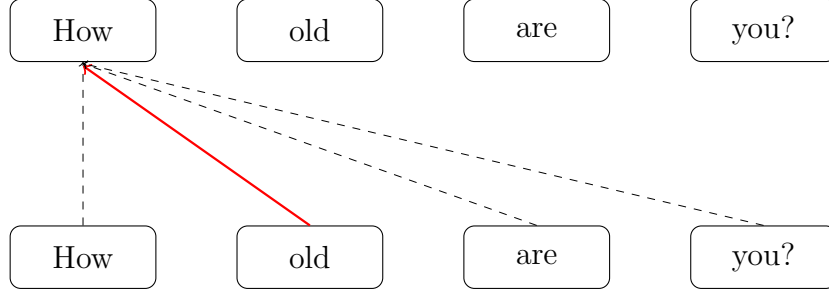
The mechanism that makes the USE capture the contextual similarity of different texts is the self-attention channel behind the deep neural network architecture. “How” ( $w_{1,1}$ ) is contextually connected with “old” ( $w_{1,2}$ ) in  $S_1$  while “How” ( $w_{3,1}$ ) is related to “are” in  $S_3$ . The contextual representation of “How” should be able to reflect this difference, which is not possible in TF-IDF and LSA because they represent words as an item in the dictionary without encoding contextual linkages in embeddings. For example, if the dictionary contains  $|V|$  words,  $w_{i,j}$  is represented by a  $|V|$ -dimensional vector in which the  $j$ th element is one and all the other elements are zeros under these approaches. This representation is known as a one-hot vector encoding. The USE does not restrict word embedding to one-hot vector encoding, allowing multiple non-zero elements. Furthermore, it transforms the given word embedding by applying the self-attention channel that can capture the contextual dependence between a group of words in the text. The self attention channel in the USE generates the contextual representation of any word in the text by taking a weighted average of all the word embeddings in the text.

Arrows in Figure 1 illustrates how attention weights link a particular word in  $S_1$  with all the other words.<sup>9</sup> Unlike TF-IDF and LSA, the USE does not use one hot vector encodings and any element in  $w_{i,j}$  can be non-zero. The attention-weighted average transforms word embeddings to perform tasks in the training stage better such as text classification and word prediction by making contextually linked words have close embeddings. Notice that unlike one-hot vector encoding, elements in word embeddings are parameters set to minimize the loss function in the training stage.<sup>10</sup>

The superiority of USE over word-counting methods in capturing the local context can be further illustrated by considering alternative FOMC statements prepared for the October

<sup>9</sup>This way of illustrating the self-attention channel follows Vaswani et al. (2017).

<sup>10</sup>One of the training dataset contains web-based question and answer texts, which facilitate the USE to detect the contextual differences in the meaning of “How” in  $S_1$  and  $S_3$  better.

Figure 1: Illustration of self-attention:  $S_1$ 

*Notes:* The red arrow highlights the contextual link between “How” and “old”.

2013 FOMC meeting. As shown in the appendix, the first paragraph of Alt A starts by acknowledging the challenge in interpreting economic data released during the intermeeting period due to the temporary shutdown of the federal government whereas Alt C does not mention this challenge. By pointing the near-term uncertainty, the reference to the government shutdown reveals that policymakers are not so sure about the recent improvement in the data. Otherwise, the description of the current outlook is fairly similar between Alt A and Alt C. The actual released statement dropped the reference to the temporary shutdown of the federal government like Alt C although it was rather close to Alt A otherwise. The textual similarity results in Table 2 suggest that the USE captures the large impact of dropping the reference to the government shutdown while TF-IDF does not.<sup>11</sup>

## 2.2 Implementation of text analysis

We characterize the “monetary policy stance” communicated by each FOMC statement. For this, we identify the “tone” of monetary policy announcements by computing the similarities between the released statement and the alternative statements. We define the “novelty” of monetary policy announcements by computing the difference between the current statement and the previous statement released after respective FOMC meetings. By taking the product of the tone and novelty of monetary policy announcements, following Ke et al. (2019), we obtain the monetary policy stance.

<sup>11</sup>When we calculate the paragraph-by-paragraph similarity, only the first paragraph makes a large difference between USE and TF-IDF. Other paragraphs do not create significant differences between the two methods.

Specifically, we define the (benchmark) monetary policy stance as

$$\text{MP stance } (t) = \underbrace{(1 - \text{Sim}(FOMC_t, FOMC_{t-1}))}_{\text{Novelty}} \underbrace{\left( \frac{\text{Sim}(FOMC_t, FOMC_{C,t}) - \text{Sim}(FOMC_t, FOMC_{A,t})}{1 - \text{Sim}(FOMC_{A,t}, FOMC_{C,t})} \right)}_{\text{Tone}} \quad (5)$$

and the (alternative) dovish and hawkish monetary policy stance as<sup>12</sup>

$$\begin{aligned} \text{Dovish MP stance } (t) &= -|1 - \text{Sim}(FOMC_t, FOMC_{t-1})|, \\ \text{Hawkish MP stance } (t) &= |1 - \text{Sim}(FOMC_t, FOMC_{t-1})|, \end{aligned} \quad (6)$$

respectively. Novelty in the current benchmark FOMC statement relative to the previous one quantifies the change in the FOMC's intended policy stance. Note the monotonicity

$$\text{tone}(\text{Dovish MP stance}) \leq \text{tone}(\text{MP stance}) \leq \text{tone}(\text{Hawkish MP stance}). \quad (7)$$

As conventional, we sign a positive tone as a hawkish stance and a negative tone as a dovish stance. We further normalize the tone measure between -1 and 1.

The above analysis is based on the USE representation at the statement level. But statements typically consist of multiple paragraphs and we may be interested in isolating the relevance of a particular paragraph. To do this, we compute the USE representation of the  $j$ -th paragraph in the  $i$ -type FOMC statement where  $i$  denotes the different versions of the statement. Let this be  $P_{j,t}^i$ . By comparing  $P_{j,t}^i$  with  $P_{k,t'}^l$ , we can decompose which paragraph contributes mostly to the change in the similarity score between statements.<sup>13</sup>

To identify monetary policy surprises around FOMC announcements, we have to proxy the market expectations for the MP stance right before the FOMC meeting. We do this by using a weighted average of the dovish MP stance and the hawkish MP stance based on the assumption that alternative statements mimic expectations of market participants with

---

<sup>12</sup>Here, we assume alternative stances differ from the benchmark stance only in terms of the tone but not novelty.

<sup>13</sup>Note that the statement level USE representation is not an equal average of paragraph-level USE representations. When we approximate the statement level USE representation by a weighted average of paragraph-level USE representations, the first and second paragraphs take most weights. While this is not the exact replication of the statement-level USE representation, the finding suggest that the ordering language may matter because it sets the context in which the subsequent words are interpreted.

more extreme views.<sup>14</sup>

$$E_{t-\Delta} \text{MP stance}(t) = (1 - p_t) \times \text{Hawkish MP stance}(t) + p_t \times \text{Dovish MP stance}(t). \quad (8)$$

Note that the weight,  $p_t$ , can vary over time.

## 2.3 Measuring the market expectations from bond prices

At the time of the FOMC announcement, the reaction of the high-frequency asset  $i$ 's prices can be captured by

$$r_{t-\Delta_l, t+\Delta_h}^i \equiv \ln \left( \frac{P_{i, t+\Delta_h}}{P_{i, t-\Delta_l}} \right) = \alpha_i + \beta_i \text{MPS}(p_t; t - \Delta) + \varepsilon_{i, t}, \quad \varepsilon_{i, t} \sim (0, \sigma_i^2). \quad (9)$$

The surprise component of the FOMC announcement measured at  $t - \Delta$  is

$$\text{MPS}(p_t; t - \Delta) = \text{MP stance}(t) - E_{t-\Delta} \text{MP stance}(t). \quad (10)$$

It is important to understand that  $\text{MPS}(p_t; t - \Delta) > 0$  corresponds to tightening monetary policy.<sup>15</sup>

---

<sup>14</sup>Lucca and Trebbi (2009) constructs a measure of monetary policy stance based on the systematic co-occurrence of words with different sentiments in FOMC statements. However, they equate market expectations to the previous meeting's monetary policy stance, ignoring the market reaction to developments during the inter-meeting period. Their alternative measure using newspaper discussions of FOMC announcements before and after the meeting do not face this issue but it is hard to identify the impact of changing the language in the statement under this approach because newspaper discussions instead of statements themselves are used to identify the policy stance.

<sup>15</sup>The underlying assumption of (8) is that the market is aware of the two alternative stances and they provide bounds when evaluating the benchmark statements, which may sound too strong given the fact that alternative statements are available only with a long time lag. One practical justification is that if we look at the bluebook in 2004, they rationalize alternative statements by intentionally beating market expectations in the hawkish or the dovish direction. Hence, alternative statements describe the Board staff's best guess for two extreme market expectations. Roughly speaking, the hawkish stance represents the most hawkish person in the financial market while the dovish stance represents the most dovish person. We are capturing the marginal investor's expectation as a weighted average of these two extreme expectations. As long as the marginal investor's expectation is within the bound set by the survey of market participants and we do not know exactly who would be the marginal investor, this assumption seems to be plausible.

## 2.4 Constructing the surprise component of monetary policy stance.

We calibrate the weight  $\{p_t\}_{t=1}^T$  that maximizes the rank correlation of the high-frequency bond returns (left-side of (9)) and the surprise component of monetary policy announcements  $MPS(p_t; t - \Delta)$ , aka, monetary policy shocks. When  $p_t$  is not time-varying, ( $p_t = \bar{p}$ ) our estimate is identical to the maximum rank correlation estimator, see Han (1987) and Sherman (1993). Here, we assume that a dovish surprise should lead to a positive bond return because bond prices move inversely with bond yields.<sup>16</sup> Specifically, we maximize the following rank correlation function with respect to  $p_t$ :

$$(p_{\tau_i})_{i=1}^T = \operatorname{argmax}_{t \neq t'} \sum \mathbf{1}(r_{\tau_t - \Delta_l, \tau_t - \Delta_h}^b > r_{\tau_{t'} - \Delta_l, \tau_{t'} - \Delta_h}^b) \mathbf{1}(MPS(p_{\tau_t}) < MPS(p_{\tau_{t'}})). \quad (11)$$

This (negative) rank correlation is maximized by calibrating  $p_t$  based on the sorted bond return. Specifically, the time series of bond returns  $\{r_{t - \Delta_l, t + \Delta_h}^b\}_{t=1}^T$  are sorted from most negative to most positive. Let the ordering of the sorted-returns be indicated with new time subscripts  $\{\tau_1, \dots, \tau_T\}$ :

$$\begin{aligned} r_{\tau_1 - \Delta_l, \tau_1 + \Delta_h}^b &= \min\{r_{t - \Delta_l, t + \Delta_h}^b\}_{t=1}^T \\ r_{\tau_T - \Delta_l, \tau_T + \Delta_h}^b &= \max\{r_{t - \Delta_l, t + \Delta_h}^b\}_{t=1}^T. \end{aligned} \quad (12)$$

For a strictly negative value of  $\beta_b$  in (9), we have that

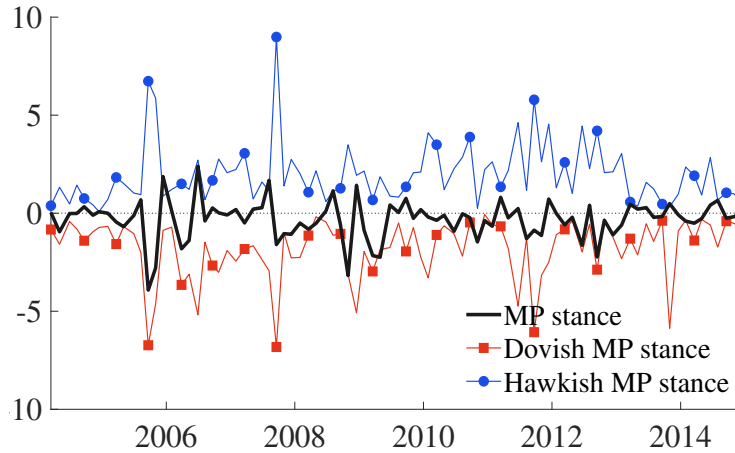
$$\begin{aligned} MPS(p_{\tau_T}) &\leq \dots \leq MPS(p_{\tau_t}) \leq \dots \leq MPS(p_{\tau_1}) \\ r_{\tau_1 - \Delta_l, \tau_1 + \Delta_h}^b &\leq \dots \leq r_{\tau_t - \Delta_l, \tau_t + \Delta_h}^b \leq \dots \leq r_{\tau_T - \Delta_l, \tau_T + \Delta_h}^b \end{aligned} \quad (13)$$

where  $\tau_t \in \{\tau_1, \dots, \tau_T\}$ . Because it is possible that there are (potentially) multiple realizations of  $\{p_{\tau_1}, \dots, p_{\tau_T}\}$  that satisfy (13), we pick the one that achieves the largest negative correlation between  $\{MPS(p_{\tau_1}), \dots, MPS(p_{\tau_T})\}$  and  $\{r_{\tau_1 - \Delta_l, \tau_1 + \Delta_h}^b, \dots, r_{\tau_T - \Delta_l, \tau_T + \Delta_h}^b\}$ . This can be done via grid search (with respect to  $p_{\tau_t}$ ). Once we select  $\{p_{\tau_1}, \dots, p_{\tau_T}\}$ , we can sort them back to match the original time subscript  $\{p_1, \dots, p_T\}$  and construct the corresponding  $MPS(p_t)$  for each  $p_t$ ,  $t \in \{1, \dots, T\}$ .

---

<sup>16</sup>The sign of the correlation is negative because  $p_t$  corresponds to a dovish probability, which contributes to a negative surprise in monetary policy stance.

Figure 2: Monetary policy stance



Notes: We normalize these measures to have a unit variance.

### 3 Empirical Results

#### 3.1 Data for alternative FOMC statements

The Federal Reserve Board staff started to prepare alternative FOMC statements from the March 2004 FOMC meeting. The latest available statement is the one prepared for the December 2014 FOMC meeting. We have 87 FOMC statements (March 2004 to December 2014) excluding two inter-meeting announcements (Aug 2007, Jan 2008). When multiple versions of hawkish or dovish alternative statements are available (e.g., Alt A1 or Alt D), we use the most extreme one to identify the tone of the released statement.<sup>17</sup>

#### 3.2 Monetary policy stance and surprises

Figure 2 provides the time series of (5) and (6) constructed based on the USE. Our measure captures the change in the policy stance including both the current action (e.g., change in the federal funds rate target) and the expected future action (e.g., forward guidance about

<sup>17</sup>For four meetings (September 2008, June 2009, June 2011, August 2013), we drop the first paragraph of each statement to calculate the textual similarity because the original version including the first paragraph has not shown enough dissimilarity between the dovish alternative statement and the hawkish statement which is crucial for our identification of the tone.

the future interest rates). As in (7), we observe the monotonicity across the three measures. While the sign of the monetary policy stance is determined by the tone, its magnitude is largely governed by the dissimilarity from the previous statement.

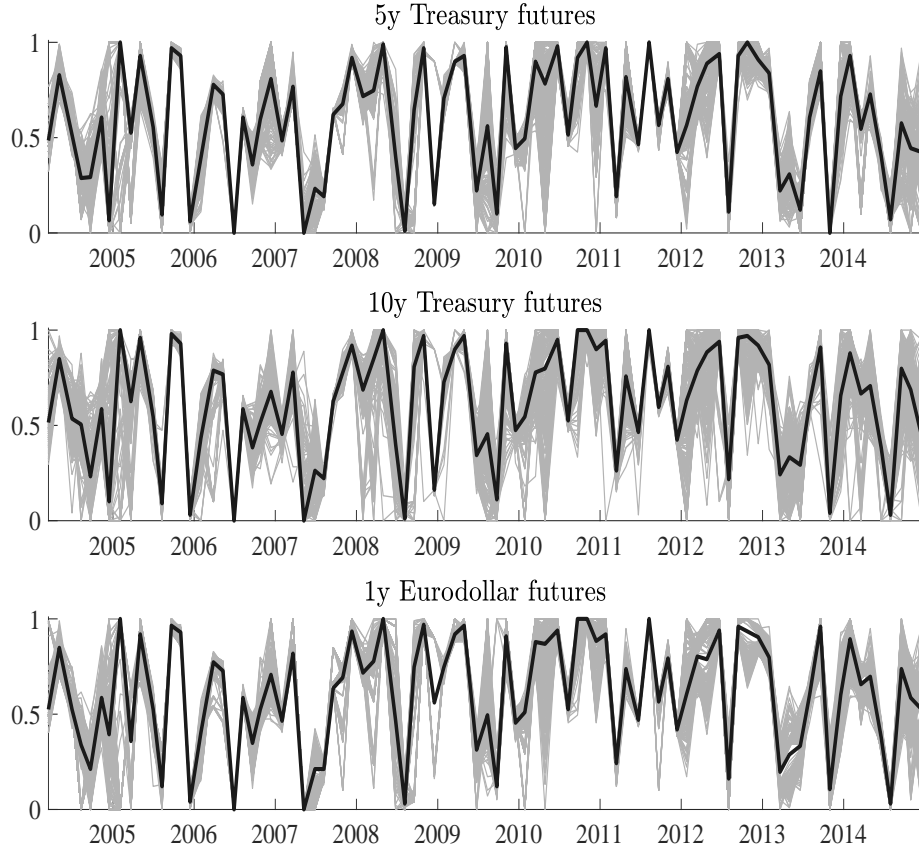
To extract monetary policy surprises, we construct the market expectation of monetary policy stance which is the weighted average of the dovish and hawkish monetary policy stance. The market-based probability that the dovish and hawkish alternative statement would be released are parameterized by  $p_t$  and  $1 - p_t$ , respectively. To calibrate the market-based probability  $p_t$ , we use high-frequency bond market return data around FOMC announcements. One virtue of our maximum rank correlation approach is that it sidesteps the burden of estimating  $\alpha_i, \beta_i, \sigma_i^2$  when identifying  $p_{1:T}$  in (9). To provide robustness to our claim, we rely on bond futures returns of various combinations of  $\Delta_l, \Delta_h \in \{10, \dots, 120\}$ min to obtain  $p_{1:T}$  and the corresponding MPS( $p_{1:T}$ ). We provide the results in Figure 3. The median values of  $\hat{p}_{1:T}$  are highly correlated with each other, e.g., 0.96 or higher. This finding implies that the dovish probabilities extracted from bond returns are fairly robust to different window intervals or instruments. The robustness of our result is different from Bu et al. (2020) who find large differences in monetary policy shock estimates depending on the maturity of the bond data. However, when both use the five-year bond return, our measures are highly correlated with each other. We interpret this as demonstrating the value of our text-based analysis in isolating the common component from the response of an individual bond return to a policy announcement without using the statistical factor analysis of bond returns of multiple maturities.

### 3.3 Stock market responses to monetary policy surprises.

Since the Federal Reserve intervenes in Treasury markets to influence the interest rate, it is not surprising that bond returns react to monetary policy surprises. But the ultimate goal of the monetary policy is to achieve the dual mandate of maximum employment and price stability by affecting the real economy. For this, understanding the links between monetary policy and asset prices above and beyond bond returns is important as highlighted by Bernanke and Kuttner (2005) who find that an unanticipated 25 bps cut in the federal funds rate leads to about 1 percent increase in the stock market return. We turn to the



Figure 3: Dovish probability comparison



*Notes:* We rely on the 5-year and 10-year Treasury bond futures returns and the 1-year Eurodollar futures returns. Returns are defined with the following interval  $\Delta_l, \Delta_h \in \{10, \dots, 120\}$ min. The median values are indicated with solid lines.

stock market reaction to our measure of monetary policy surprise to check if our text-based measure captures similar stock market responses.

For the benchmark case, we select the 5-year Treasury bond futures returns with window intervals  $\Delta_l = \Delta_h = 10$ min to back out the probability weights and construct  $\text{MPS}(\hat{p}_t)$ . Conditional on this output, we conduct the regression analysis using stock returns as an external validation check. Specifically, we regress stock returns  $r_{t-\Delta_l, t+\Delta_h}^s$  on the bond market-implied  $\text{MPS}(\hat{p}_t)$ . In essence, we are estimating (9) using an OLS with stock returns. The estimation results summarized in Table 3 imply that the bond market-implied  $\text{MPS}(\hat{p}_t)$  significantly predict stock returns measured at various window intervals. Because we normalized  $\text{MPS}(\hat{p}_t)$  to have a unit variance, we can directly interpret the magnitude of  $\beta$  coefficient in assessing the

Table 3: Stock returns: Regression results

$[\Delta_l$	$\Delta_h]$	$\alpha$	$\beta$	$t$ -stat ( $\alpha$ )	$t$ -stat ( $\beta$ )	$R^2$
$[-10$	$10]$	0.05	-0.23	1.08 [1.07]	-4.75 [-4.43]	0.19 [0.19]
$[-20$	$20]$	0.04	-0.20	0.75 [0.74]	-4.78 [-4.24]	0.12 [0.12]
$[-30$	$30]$	0.10	-0.18	1.49 [1.48]	-4.45 [-3.81]	0.08 [0.08]
$[-40$	$40]$	0.16	-0.19	2.25 [2.24]	-3.33 [-2.92]	0.07 [0.07]
$[-50$	$50]$	0.16	-0.18	2.21 [2.13]	-3.20 [-2.77]	0.07 [0.07]
$[-60$	$60]$	0.20	-0.22	2.56 [2.59]	-3.35 [-2.96]	0.08 [0.08]
$[-90$	$90]$	0.19	-0.21	2.25 [2.21]	-2.43 [-2.18]	0.06 [0.06]
$[-120$	$120]$	0.17	-0.21	1.72 [1.69]	-1.85 [-1.64]	0.05 [0.05]

*Notes:* Based on the median value of  $\hat{p}_{1:T}$ , we regress stock returns (defined at various window intervals) on  $MPS(\hat{p}_{1:T})$ . Numbers in square brackets are alternative  $t$  statistics and  $R^2$  based on bootstrap standard errors to account for the fact that our policy surprise measure is a generated regressor.

economic significance of  $MPS(\hat{p}_t)$ . On average, we find that a positive one-standard-deviation surprise leads to a 20 bps drop in stock prices. The  $R^2$  values are higher for returns defined with shorter window intervals.

Nakamura and Steinsson (2018) argue that unexpected policy easing or tightening by the FOMC identified by five different high-frequency interest rate futures data (current month and next month federal funds futures, 2, 3, 4 quarter ahead Eurodollar futures) around policy announcements often did not move the private sector's expectation of economic growth in the intended direction. For instance, they show that the Bluechip forecast of real GDP growth declined after unexpected policy tightening in many cases because the dovish announcement might have revealed the Federal Reserve's private information on the gloomy economic outlook. However, Bauer and Swanson (2020) provide evidence that the Federal Reserve does not have an information advantage over the private sector and suggest that the finding in Nakamura and Steinsson (2018) can be explained by the relatively low-frequency nature of the private sector forecast data. In addition, they use the daily-frequency stock market return data. Bauer and Swanson (2020) show that the unexpected policy tightening leads to a decline in the high-frequency stock market return even though they use the same monetary policy shock measure as Nakamura and Steinsson (2018). Our finding is consistent with Bauer and Swanson (2020) in terms of the sign of the stock market response to a monetary policy surprise. As in Bauer and Swanson (2020), our finding is not driven by

the difference between our measure and monetary policy shock estimates of [Nakamura and Steinsson \(2018\)](#) because both measures are highly correlated as shown in Table 4.

Our  $MPS(\hat{p}_t)$  is also highly correlated with other measures of monetary policy shocks based on the high-frequency asset market data around FOMC announcements. [Bu et al. \(2020\)](#) construct a monetary policy shock using the idea that the variance of the daily bond return is higher on FOMC days relative to non-FOMC days due to the monetary policy announcement. In addition to near-term maturities, they use information from the entire yield curve (up to the maturity of thirty years). They argue that the information channel effect highlighted by [Nakamura and Steinsson \(2018\)](#) is present mainly because [Nakamura and Steinsson \(2018\)](#) consider only the near-term interest rate data in constructing a measure of monetary policy shock. However, we do not find any significant difference in our sample in terms of the correlation of our measure with the two measures provided by [Nakamura and Steinsson \(2018\)](#) and [Bu et al. \(2020\)](#). Also, given the robustness of our measure to the maturity of the bond return used to back out policy surprises, we do not think that the maturity of the bond return is critical in accounting for the lack of the information channel effect in our study. More plausibly, we suspect that more efforts made by the Federal Reserve staff to fine tune statement language since 2004 might have increased the effectiveness of monetary policy communications on the asset markets. This hypothesis is consistent with the finding in [Lunsford \(2020\)](#) who shows that the information channel effect was present before August 2003 but disappeared for the later sample as the FOMC provides a more explicit policy inclination in statements.<sup>18</sup>

[Swanson \(2017\)](#) identifies multiple dimensions of monetary policy shocks using eight different asset prices consisting of three Treasury bond yields (maturities of 2, 5, 10 years) on top of the five interest rate futures used in [Nakamura and Steinsson \(2018\)](#). He computes the three principal components that account for common variations in these eight different asset prices around FOMC announcements. He rotates three components to get 1) federal funds rate (FFR) factor that affects the current month federal funds rate futures, 2) forward guidance (FG) factor that is orthogonal to the change in the current month federal funds rate futures, and 3) large-scale asset purchase (LSAP) factor that is also orthogonal to the change in the current month federal funds rate futures and plays a minimum role in explaining the data before the federal funds rate reached the effective lower bound in December 2008. In

---

<sup>18</sup>A similar observation was made by [Hoesch et al. \(2020\)](#).

practice, these factors are largely distinguished by their different loadings on the maturity spectrum of the underlying interest rate data. The FFR factor has a large non-zero loading on the current month federal funds futures while the other factors have zero loadings. In addition, the loadings of FG factor are concentrated in the one-to-five year maturity spectrum while the LSAP factor has the largest loading on the ten-year Treasury yield. Our measure is particularly highly correlated with the FG factor, indicating that the communication strategy by different wording can be mostly effective in moving the medium-term interest rates.<sup>19</sup>

Bauer and Swanson (2020) emphasize that it is important to control information from economic indicators released during the inter-meeting period because the FOMC responds to it. Our results in Table 4 showing a big discrepancy in the correlation with the forward guidance factor in Swanson (2017) between policy surprises and policy stance is consistent with their finding because controlling the market expectation before the FOMC meeting is most important for the forward guidance factor that is overweighting the short-to-medium term maturity bond returns. On the other hand, the LSAP factor which is loaded onto the slope of the yield curve and overweighting the long-term bond return response is less sensitive to controlling changes in market expectations during the inter-meeting period as the correlation with policy surprise is largely same as that with policy stance.

Table 4: Comparison with other measures

	MP stance: surprise	MP stance: level
Bu et al. (2020)	0.50	0.16
Nakamura and Steinsson (2018)	0.50	-0.10
Swanson (2017) (FFR+FG+LSAP)	0.50	0.01
FFR	0.20	0.03
FG	0.52	-0.16
LSAP	-0.12	-0.17

*Notes:* Based on the median value of  $\hat{p}_{1:T}$ , we construct  $MPS(\hat{p}_{1:T})$  and compute correlation with other existing measures of monetary policy factors. The last three factors are from Swanson (2017): 1) federal funds rate (FFR) factor; 2) forward guidance (FG) factor; and 3) large-scale asset purchase (LSAP) factor.

<sup>19</sup>The negative correlation with the LSAP factor is due to the fact that Swanson (2017) normalized a positive innovation to the LSAP as larger asset purchases than expected, resulting in policy easing.

### 3.4 Counterfactual policy evaluation.

The key advantage of our approach is that we are able to conduct a counterfactual policy evaluation by replacing the released statement with either one of the alternative statements. We also look at the paragraph-level USE representation instead of the statement-level one in order to further investigate the impact of changing a particular part of the statement on the monetary policy shock.

We consider two episodes in our sample period. First, we look at the FOMC statement released in September 2007.<sup>20</sup> The policy decision suggested in Alt B was to lower its target for the federal funds rate 50 bps whereas that in Alt C was to lower it by 25 bps. Thus, the market could view Alt C more hawkish than Alt B. On the other hand, in the assessment of risk, Alt C stated that the downside risks to economic growth outweigh the upside risks to inflation (we refer to it as “balance of risk statement”), which was omitted in Alt B. Our similarity measures suggest that the statement about the balance of risk can be perceived to be more accommodative than the 50 bps cut in the target rate.<sup>21</sup> If Alt C were the only available hawkish alternative statement, this might have posed a risk to our identification of the tone. But the Board staff made a more extreme hawkish version of the alternative statement (Alt D) for this meeting perhaps because they recognized the possibility that Alt C may sound dovish rather than hawkish due to the “balance of risk” language. The released statement was similar to Alt B by lowering the federal funds rate 50 bps but without stressing downside risk. In this section, we consider the counterfactual adoption of Alt C and quantify its impact on stock prices.

Conditional on  $\hat{\beta} = -0.23$  in (the first row of) Table 3, we multiply the counterfactual monetary policy surprise component to asset the impact on the stock returns (defined in the 10-minute interval). For this, we replace the monetary policy stance with the counterfactual one and subtract the bond price-implied expected monetary policy stance to compute the counterfactual monetary policy surprise component. It is important to note that we are only replacing one data point (that corresponds to the September 2007 FOMC statement release

---

<sup>20</sup>The detailed description of alternative statements is provided in the appendix.

<sup>21</sup>We confirm this by looking paragraph level USE representations of statements. See the appendix for further discussion. To some extent, the Board staff recognized this issue and provided a more extreme hawkish version of alternative statement (Alt D) at this FOMC cycle. Alt D does have the balance of risk statement like Alt C. Our identification of the tone of the monetary policy stance compares the released statement with Alt A and Alt D, sidestepping this ambiguity in the tone of Alt C.

date) while keeping all else equal in this exercise. We find that the counterfactual monetary policy stance turns out to be much more dovish leading to an 1.94% increase of stock prices relative to the 0.1% increase without that statement. Based on the estimate of [Bernanke and Kuttner \(2005\)](#), this might be equivalent to an additional 50 bps cut in the federal funds rate.<sup>22</sup> While suggestive, caution is needed in interpreting this result because the counterfactual policy surprise is more than three standard deviation away from the original estimate of policy surprise, making our assumption on the stability of the private sector’s expectations somewhat fragile.<sup>23</sup> If the counterfactual shock is much bigger than the realized one, the linearity of our stock market return regression may not be a good assumption for this.

We perform another counterfactual exercise for the December 2010 FOMC meeting that can be a modest intervention.<sup>24</sup> In this case, the Alt A replaces a qualitative forward guidance on the future path of the federal funds rate like “for an extended period” by a more explicit time-dependent guidance such as “at least through the mid-2012”. It also expands asset purchases by \$200 billion. The actual statement released after the December 2010 meeting was close to Alt B and did not include languages on forward guidance and asset purchases used in Alt A. We find that the stock market return might have increased by about 0.7% based on our regression if the Alt A were released.<sup>25</sup> Since the counterfactual policy surprise is less than one standard deviation away from the estimate of the original policy surprise, we regard this result as a modest policy change that is less likely to materially affect the expectations of the private sector.<sup>26</sup>

---

<sup>22</sup>Since Alt C actually contained a 25 bps cut in the federal funds rate target rather than a 50 bps cut in the released statement, our analysis suggest that the change in risk assessment actually might have had an impact equivalent to a 75 bps cut in the federal funds rate target.

<sup>23</sup>This is an example of well known [Lucas \(1976\)](#) critique. While the critique is valid for any reduced-form model that does not model the private sector’s expectations formation process explicitly, the practical relevance may be small if the counterfactual surprise is relatively modest.

<sup>24</sup>The detailed description of alternative statements is available in the appendix.

<sup>25</sup>Our paragraph level USE representations suggest that changing the language on asset purchased might have made a slightly bigger difference than changing the language on forward guidance.

<sup>26</sup>Eventually, the FOMC adopted a more explicit time-dependent forward guidance on the future path of the federal funds rate in the August 2011 meeting and [Lunsford \(2020\)](#) highlights the effectiveness of including such a language in influencing financial market conditions.

## 4 Conclusion

The central bank's public communications about current and future policy actions have increasingly received attention as a policy tool. Since March 2004, the FOMC has deliberated on alternative policy statements prepared by the Federal Reserve staff before each FOMC meeting. Two alternative statements capture the hawkish or dovish deviation from the central tendency of the market expectation right before the meeting, providing cross-sectional variations around the released statement. We apply a novel natural language processing algorithm based on a deep learning architecture to alternative FOMC statements in order to identify the tone of the released statement. This USE algorithm detects the contextual meaning of words in the statement and quantifies the information provided by language in alternative statements. We construct a new measure of monetary policy surprises by combining the high-frequency bond returns around FOMC announcements with the text analysis of alternative statements by the USE. Our text-based measure is able to capture the common factor in an individual bond return response to a policy announcement without using the statistical factor analysis common in the existing literature, which can be sensitive to the choice of target maturities. We find that an unexpected policy tightening leads to a decline in the stock market return on average. The finding vindicates that the FOMC's communication achieved mostly its intended effect on the stock market return at least since 2004, which is consistent with the recent empirical findings. Two suggestive counterfactual exercises show 1) involving alternative statements implies that changing the language describing the risk assessment might have had a much bigger impact on the stock market return than changing the federal funds rate target by 25 bps and 2) a more explicit forward guidance on the future path of the federal funds rate could have been effective in easing financial market conditions.

## References

- Bauer, Michael and Eric T Swanson**, “The Fed’s Response to Economic News Explains the,” 2020.
- Bernanke, Ben S.**, “Central Bank Independence, Transparency, and Accountability,” in “Speech of Chairman Ben S. Bernanke at the Institute for Monetary and Economic Studies International Conference” 2010.
- Bernanke, Ben S and Kenneth N Kuttner**, “What explains the stock market’s reaction to Federal Reserve policy?,” *The Journal of finance*, 2005, 60 (3), 1221–1257.
- Blinder, Alan S., Michael Ehrmann, Marcel Fratzscher, Jakob De Haan, and David-Jan Jansen**, “Central Bank Communication and Monetary Policy: A Survey of Theory and Evidence,” *Journal of Economic Literature*, 2008, 46 (4), 910–945.
- Bu, Chunya, John Rogers, and Wenbin Wu**, “Forward-Looking Monetary Policy and the Transmission of Conventional Monetary Policy Shocks,” 2020.
- Cer, Daniel, Yinfei Yang, Shengyi Kong, Nan Hua, Nicole Limtiaco, Rhomni St. John, Noah Constant, Mario Guajardo-Cespedes, Steve Yuan, Chris Tar, Yun-Hsuan Sung, Brian Strope, and Ray Kurzweil**, “Universal Sentence Encoder,” 2018. Manuscript.
- FRB**, “Some General Issues Concerning Statement Language,” in “Bluebook, 2004 March FOMC Meeting” 2004.
- Gentzkow, Matthew, Bryan Kelly, and Matt Taddy**, “Text as Data,” *Journal of Economic Literature*, 2019, 57 (3), 535–574.
- Giavazzi, Francesco, Felix Iglhaut, Giacomo Lemoli, and Gaia Rubera**, “Terrorist Attacks, Cultural Incidents and the Vote for Radical Parties: Analyzing Text from Twitter,” Technical Report, National Bureau of Economic Research 2020.
- Gürkaynak, Refet S, Brian Sack, and Eric Swanson**, “Do Actions Speak Louder Than Words? The Response of Asset Prices to Monetary Policy Actions and Statements,” *International Journal of Central Banking*, 2005, 1, 55–93.



- Han, Aaron K.**, “Non-parametric Analysis of a Generalized Regression Model,” *Journal of Econometrics*, 1987, *35*, 303–316.
- Hansen, Stephen, Michael McMahon, and Andrea Prat**, “Transparency and Deliberation Within the FOMC: A Computational Linguistics Approach,” *Quarterly Journal of Economics*, 2017, *133* (2), 801–870.
- Hoesch, Lukas, Barbara Rossi, and Tatevik Sekhposyan**, “Has the information channel of monetary policy disappeared? Revisiting the empirical evidence,” 2020.
- Jegadeesh, Narasimhan and Di Andrew Wu**, “Deciphering FedSpeak: The information content of FOMC meetings,” *Available at SSRN 2939937*, 2017.
- Ke, Zheng Tracy, Bryan T. Kelly, and Dacheng Xiu**, “Predicting Returns with Text Data,” 2019. Manuscript.
- Lucas, Robert E J.**, “Econometric policy evaluation: A critique,” in “Carnegie-Rochester Conference Series on Public Policy,” Vol. 1 Elsevier 1976, pp. 19–46.
- Lucca, David O and Francesco Trebbi**, “Measuring central bank communication: an automated approach with application to FOMC statements,” Technical Report, National Bureau of Economic Research 2009.
- Lunsford, Kurt G**, “Policy Language and Information Effects in the Early Days of Federal Reserve Forward Guidance,” *American Economic Review*, 2020.
- Meade, Ellen E. and Miguel Acosta**, “Hanging on Every Word: Semantic Analysis of the FOMC’s Postmeeting Statement,” FEDS Notes, Board of Governors of the Federal Reserve System (U.S.) 2015.
- Nakamura, Emi and Jón Steinsson**, “High-frequency identification of monetary non-neutrality: the information effect,” *The Quarterly Journal of Economics*, 2018, *133* (3), 1283–1330.
- Schonhardt-Bailey, Cheryl**, *Deliberating American Monetary Policy: A Textual Analysis*, MIT Press, 2013.

- Shapiro, Adam and Daniel J Wilson**, “The Evolution of the FOMC’s Explicit Inflation Target,” *Federal Reserve Bank of San Francisco Working Paper 2019-02*, 2019, 2019, 12.
- Sherman, Robert P.**, “The Limiting Distribution of the Maximum Rank Correlation Estimator,” *Econometrica*, 1993, 61 (1), 1123–137.
- Swanson, Eric T**, “Measuring the effects of Federal Reserve forward guidance and asset purchases on financial markets,” Technical Report, National Bureau of Economic Research 2017.
- Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin**, “Attention is all you need,” in “Advances in neural information processing systems” 2017, pp. 5998–6008.
- Woodford, Michael**, “Central Bank Communication and Policy Effectiveness,” *Proceedings of the Federal Reserve Bank of Kansas City Economic Symposium at Jackson Hole*, 2005, pp. 399–474.

## Appendix

### A Illustration

Figure A-1: FOMC statement in November 2005

Class I FOMC - Restricted Controlled (FR) Page 13 of 36

Table 1: Alternative Language for the November FOMC Announcement				
	September FOMC	Alternative A	Alternative B	Alternative C
<b>Policy Decision</b>	1. The Federal Open Market Committee decided today to raise its target for the federal funds rate by 25 basis points to 3¼ percent.	The Federal Open Market Committee decided today to leave its target for the federal funds rate unchanged.	The Federal Open Market Committee decided today to raise its target for the federal funds rate by 25 basis points to 4 percent.	The Federal Open Market Committee decided today to raise its target for the federal funds rate by 25 basis points to 4 percent.
<b>Rationale</b>	2. Output appeared poised to continue growing at a good pace before the tragic toll of Hurricane Katrina. The widespread devastation in the Gulf region, the associated dislocation of economic activity, and the boost to energy prices imply that spending, production, and employment will be set back in the near term. In addition to elevating premiums for some energy products, the disruption to the production and refining infrastructure may add to energy price volatility.  While these unfortunate developments have increased uncertainty about near-term economic performance, it is the Committee's view that they do not pose a more persistent threat. Rather, monetary policy accommodation, coupled with robust underlying growth in productivity, is providing ongoing support to economic activity.	Elevated energy prices and hurricane-related disruptions in economic activity seem to have slowed the growth of spending, set back employment, and weakened consumer and business confidence. The persistence of such effects is uncertain, but robust underlying growth of productivity and monetary policy accommodation are providing support to economic activity.	Elevated energy prices and hurricane-related disruptions in economic activity seem to have temporarily slowed the growth of spending and set back employment. However, monetary policy accommodation, coupled with robust underlying growth in productivity, is providing ongoing support to economic activity. Spending will also be boosted by rebuilding efforts in hurricane-affected areas.	The disruptive effects of recent hurricanes seem likely to be temporary, especially in light of increased spending associated with rebuilding efforts. Economic growth continues to be supported by robust underlying growth in productivity.
	3. Higher energy and other costs have the potential to add to inflation pressures. However, core inflation has been relatively low in recent months, and longer-term inflation expectations remain contained.	High energy and other costs have added to inflation pressures. However, core inflation has been relatively low in recent months, and longer-term inflation expectations remain contained.	The cumulative rise in energy and other costs has added to inflation pressures. However, core inflation has been relatively low in recent months, and longer-term inflation expectations remain contained.	Core inflation and longer-term inflation expectations remain contained. However, high energy and other costs have boosted near-term inflation expectations and price pressures, likely making further policy firming necessary.
<b>Assessment of Risk</b>	4. The Committee perceives that, with appropriate monetary policy action, the upside and downside risks to the attainment of both sustainable growth and price stability should be kept roughly equal.	[no change]	[no change]	[none]
	5. With underlying inflation expected to be contained, the Committee believes that policy accommodation can be removed at a pace that is likely to be measured. Nonetheless, the Committee will respond to changes in economic prospects as needed to fulfill its obligation to maintain price stability.	With underlying inflation expected to be contained, the Committee believes that remaining policy accommodation can be removed at a pace that is likely to be measured. Nonetheless, the Committee will respond to changes in economic prospects as needed to fulfill its obligation to maintain price stability.	[no change]	[none]

Figure A-1 provides the official FOMC statement released in November 2005. The largest discrepancy between Alt A and Alt C under USE

$$\text{sim}(FOMC_t, FOMC_{A,t}) = 0.984, \quad \text{sim}(FOMC_t, FOMC_{C,t}) = 0.858. \quad (\text{A-1})$$

Two key changes in Alt C: (i) drops the “measured pace” language for the first time; (ii) eliminates the balance of risk in the previous statement.

Figure A-2 provides the official FOMC statement released in September 2007. Alt C was supposed to be more hawkish than Alt B but the balance of risk statement overwhelmed a

Figure A-2: FOMC statement in September 2007

Class I FOMC - Restricted Controlled (FR)

Page 22 of 49

Table 1: Alternative Language for the September 2007 FOMC Announcement				
	Alternative A	Alternative B	Alternative C	Alternative D
<b>Policy Decision</b>	1. The Federal Open Market Committee decided today to <b>lower</b> its target for the federal funds rate <b>50 basis points to 4¼ percent</b> .	The Federal Open Market Committee decided today to <b>lower</b> its target for the federal funds rate <b>50 basis points to 4¼ percent</b> .	The Federal Open Market Committee decided today to <b>lower</b> its target for the federal funds rate <b>25 basis points to 5 percent</b> .	The Federal Open Market Committee decided today to keep its target for the federal funds rate at 5¼ percent.
<b>Rationale</b>	2. <b>Tighter credit conditions and the intensification of the housing correction appear likely to exert appreciable restraint on economic growth. Moreover, the potential for significant spillovers from credit market disruptions to business and household spending poses a risk to the outlook. Today's action is intended to help mitigate the adverse effects on the broader economy arising from the disruptions in financial markets and to promote moderate growth over time.</b>	Economic growth was moderate during the first half of the year, <b>but the tightening of credit conditions has the potential to intensify the housing correction and to restrain economic growth more generally. Today's action is intended to help forestall some of the adverse effects on the broader economy that might otherwise arise from the disruptions in financial markets and to promote moderate growth over time.</b>	Economic growth was moderate during the first half of the year, <b>but the tightening of credit conditions has the potential to intensify the housing correction and to restrain economic growth more generally. Today's action is intended to help forestall some of the adverse effects on the broader economy that might otherwise arise from the disruptions in financial markets and to promote moderate growth over time.</b>	Economic growth was moderate during the first half of the year. <b>Financial market conditions have deteriorated in recent weeks, leading to tighter credit and an intensification of the housing correction. These developments have the potential to restrain growth in economic activity. Nonetheless, the economy seems likely to continue to expand at a moderate pace over coming quarters, supported by solid growth outside the housing sector and a robust global economy.</b>
	3. Readings on core inflation have improved modestly <b>this year. However, the Committee judges that some inflation risks remain, and it will continue to monitor inflation developments carefully.</b>	Readings on core inflation have improved modestly <b>this year. However, the Committee judges that some inflation risks remain, and it will continue to monitor inflation developments carefully.</b>	Readings on core inflation have improved modestly <b>this year. However, the Committee judges that some inflation risks remain, and it will continue to monitor inflation developments carefully.</b>	Readings on core inflation have improved modestly <b>this year. However, a sustained moderation in inflation pressures has yet to be convincingly demonstrated. Moreover, the high level of resource utilization has the potential to sustain those pressures.</b>
<b>Assessment of Risk</b>	4. <b>Even after today's action, the Committee judges that the downside risks to economic growth outweigh the upside risks to inflation. Future policy adjustments will depend on the outlook for both inflation and economic growth, as implied by incoming information.</b>	The Committee will continue to closely follow timely indicators of economic prospects and will act as needed to foster price stability and sustainable economic growth.	Even after today's action, the Committee judges that the downside risks to economic growth outweigh the upside risks to inflation. Future policy adjustments will depend on the outlook for both inflation and economic growth, as implied by incoming information.	<b>In the current circumstances, the Committee judges that the downside risks to economic growth are now roughly balanced by the upside risks to inflation. Future policy adjustments will depend on the outlook for both inflation and economic growth, as implied by incoming information.</b>

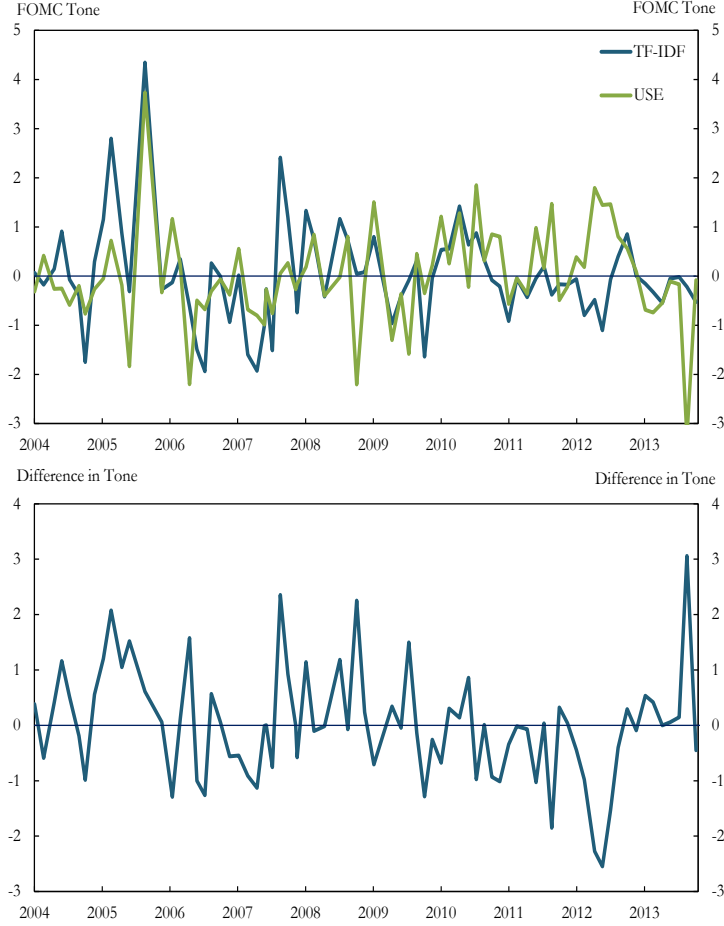
smaller rate cut. The released statement does not include a change in the risk assessment while the cutting the rate by 50 basis points as in Alt B.

$$\begin{aligned} \text{sim}(FOMC_{A,t}, FOMC_{C,t}) &= 0.99 > 0.968 = \text{sim}(FOMC_{A,t}, FOMC_t) \quad (\text{A-2}) \\ \text{sim}(FOMC_{C,t}, FOMC_{D,t}) &= 0.897 < 0.968 = \text{sim}(FOMC_t, FOMC_{D,t}). \end{aligned}$$

The similarity calculation indicates that not just the size of the rate cut but also the statement of the balance of risk matters. Our measures suggest that the balance of risk statement could have provided more accommodation than a 25 bps rate cut.

Figure A-4 compares alternative statements for the December 2010 FOMC meeting. Alt A

Figure A-3: FOMC tone identification: TF-IDF versus USE (normalized)



provides a more explicit time-dependent forward guidance on the interest rate by mentioning that rates would be low “at least through mid-2012” instead of “for an extended period.” The change in forward guidance together with the expanded asset purchases made the tone of Alt A clearly much more dovish than the released statement which was close to Alt B.

Figure A-3 compares the approach of TF-IDF with USE. Note that TF-IDF finds a small difference between Alt A and Alt C while USE detects a large difference.

$$\text{USE: } \text{sim}(FOMC_t, FOMC_{A,t}) = 0.895, \quad \text{sim}(FOMC_t, FOMC_{C,t}) = 0.99 \quad (\text{A-3})$$

$$\text{TF-IDF: } \text{sim}(FOMC_t, FOMC_{A,t}) = 0.975, \quad \text{sim}(FOMC_t, FOMC_{C,t}) = 0.972 \quad (\text{A-4})$$

Figure A-4: Alternative A FOMC statement in December 2010

Class I FOMC - Restricted Controlled (FR)

December 9, 2010

Table 1: Overview of Alternatives for the December 14 FOMC Statement

Key Components	November Statement	December Alternatives			
		A	B	C	D
Economic Activity					
Recent Developments	pace of recovery continues to be slow	pace of recovery continues to be slow		recovery is continuing	economic recovery is proceeding
Labor Market	pace of recovery continues to be slow; high unemployment; employers remain reluctant to add to payrolls; unemployment rate is elevated	pace of recovery continues to be slow		recovery is continuing	n.a.
		high unemployment		n.a.	
		employers remain reluctant to add to payrolls; unemployment rate is elevated			
Outlook	gradual return to higher resource utilization with price stability	gradual return to higher resource utilization with price stability			
	progress has been disappointingly slow	progress remains disappointingly slow	progress has been disappointingly slow	progress has been slow	n.a.
Inflation					
Recent Developments	expectations stable, but underlying inflation has trended lower; measures are somewhat low	expectations stable, but underlying inflation has continued to trend downward; measures are low	expectations stable, but underlying inflation has continued to trend downward; measures are somewhat low	expectations stable, but underlying inflation has trended lower; measures are somewhat low	although underlying inflation has trended lower, expectations have remained stable
Outlook	same as "Economic Activity" outlook above	same as "Economic Activity" outlook above			
Target Federal Funds Rate					
Intermeeting Period	0 to ¼ percent	0 to ¼ percent			
Forward Guidance	exceptionally low levels for an extended period	exceptionally low levels at least through mid-2012	exceptionally low levels for an extended period		low levels for some time
SOMA Portfolio Policy					
Approach	\$600 billion of Treasuries by end of 2011:Q2, \$75 billion per month	\$800 billion of Treasuries (\$200b more than Nov.), \$75 billion per month, through 2011:Q3	\$600 billion of Treasuries by end of 2011:Q2, \$75 billion per month	\$400 billion of Treasuries (\$200b less than Nov.), \$50 billion per month, through 2011:Q2	discontinue program announced in November
	maintain reinvestment policy	maintain reinvestment policy			for the time being, maintain existing reinvestment policy
Future Policy Action					
Approach	will adjust program as needed; will employ policy tools as necessary to support the recovery and to help ensure that inflation, over time, is at levels consistent with its mandate	will adjust program as needed; will employ policy tools as necessary to support the recovery and to help ensure that inflation, over time, is at levels consistent with its mandate			will employ policy tools as necessary to promote maximum employment and price stability

Alternatives

A change in the description of the interpretation of the incoming data seems to play a bigger role in USE than TF-IDF. We can see that in Figure A-5 - Figure A-7.

Figure A-5: Alternative A FOMC statement in October 2013

Class I FOMC – Restricted Controlled (FR)	Authorized for Public Release	October 24, 2013
<b>FOMC STATEMENT—OCTOBER 2013 ALTERNATIVE A</b>		
<ol style="list-style-type: none"> <li>1. <u>The effects of the temporary shutdown of the federal government [ , including delays in releases of some key data, ] have made the evolution of economic conditions during the intermeeting period somewhat more difficult to assess. However,</u> information received since the Federal Open Market Committee met in July <u>September generally</u> suggests that economic activity has been expanding at a moderate <u>modest</u> pace. <del>Some</del> Indicators of labor market conditions have shown <u>some</u> further improvement <del>in recent months</del>, but the unemployment rate remains elevated. <u>Available data suggest that</u> household spending and business fixed investment advanced, <del>and but that the recovery in</del> the housing sector <del>has been</del> strengthening, but mortgage rates have risen further <u>has slowed in response to higher mortgage rates</u>, and Fiscal policy is restraining economic growth. Apart from fluctuations due to changes in energy prices, inflation has been running below the Committee's longer-run objective, <del>but even though</del> longer-term inflation expectations have remained stable.</li> <li>2. Consistent with its statutory mandate, the Committee seeks to foster maximum employment and price stability. The Committee expects that, with appropriate policy accommodation, economic growth will pick up from its recent pace and the unemployment rate will gradually decline toward levels the Committee judges consistent with its dual mandate. The Committee sees the downside risks to the outlook for the economy and the labor market as having diminished, on net, since last fall, but the tightening of financial conditions observed <del>in recent months</del> <u>since the spring</u>, if sustained, could slow the pace of improvement in the economy and labor market. The Committee recognizes that inflation persistently below its 2 percent objective could pose risks to economic performance, but it anticipates that inflation will move back toward its objective over the medium term.</li> <li>3. Taking into account the extent of federal fiscal retrenchment <u>over the past year</u>, the Committee sees the improvement in economic activity and labor market conditions since it began its asset purchase program <del>a year ago</del> as consistent with growing underlying strength in the broader economy. However, the Committee <del>decided to await more evidence that progress will be sustained before adjusting</del> <u>judges that progress toward its objectives for the labor market and inflation is not yet sufficient to warrant reducing</u> the pace of its purchases. Accordingly, the Committee decided to continue purchasing additional agency mortgage-backed securities at a pace of \$40 billion per month and longer-term Treasury securities at a pace of \$45 billion per month. The Committee is maintaining its existing policy of reinvesting principal payments from its holdings of agency debt and agency mortgage-backed securities in agency mortgage-backed securities and of rolling over maturing Treasury securities at auction. Taken together, these actions should maintain downward pressure on longer-term interest rates, support mortgage markets, and help to make broader financial conditions more accommodative, which in turn should promote a stronger economic recovery and help to ensure that inflation, over time, is at the rate most consistent with the Committee's dual mandate.</li> <li>4. The Committee will closely monitor incoming information on economic and financial developments in coming months and will continue its purchases of Treasury and</li> </ol>		

Figure A-6: Released FOMC statement in October 2013

## Press Release

## FEDERAL RESERVE press release



*Release Date: October 30, 2013*

**For immediate release**

Information received since the Federal Open Market Committee met in September generally suggests that economic activity has continued to expand at a moderate pace. Indicators of labor market conditions have shown some further improvement, but the unemployment rate remains elevated. Available data suggest that household spending and business fixed investment advanced, while the recovery in the housing sector slowed somewhat in recent months. Fiscal policy is restraining economic growth. Apart from fluctuations due to changes in energy prices, inflation has been running below the Committee's longer-run objective, but longer-term inflation expectations have remained stable.

Consistent with its statutory mandate, the Committee seeks to foster maximum employment and price stability. The Committee expects that, with appropriate policy accommodation, economic growth will pick up from its recent pace and the unemployment rate will gradually decline toward levels the Committee judges consistent with its dual mandate. The Committee sees the downside risks to the outlook for the economy and the labor market as having diminished, on net, since last fall. The Committee recognizes that inflation persistently below its 2 percent objective could pose risks to economic performance, but it anticipates that inflation will move back toward its objective over the medium term.

Taking into account the extent of federal fiscal retrenchment over the past year, the Committee sees the improvement in economic activity and labor market conditions since it began its asset purchase program as consistent with growing underlying strength in the broader economy. However, the Committee decided to await more evidence that progress will be sustained before adjusting the pace of its purchases. Accordingly, the Committee decided to continue purchasing additional agency mortgage-backed securities at a pace of \$40 billion per month and longer-term Treasury securities at a pace of \$45 billion per month. The Committee is maintaining its existing policy of reinvesting principal payments from its holdings of agency debt and agency mortgage-backed securities in agency mortgage-backed securities and of rolling over maturing Treasury securities at auction. Taken together, these actions should maintain downward pressure on longer-term interest rates, support mortgage markets, and help to make broader financial conditions more accommodative, which in turn should promote a stronger economic recovery and help to ensure that inflation, over time, is at the rate most consistent with the Committee's dual mandate.

The Committee will closely monitor incoming information on economic and financial developments in coming months and will continue its purchases of Treasury and agency mortgage-backed securities, and employ its other policy tools as appropriate, until the outlook for the labor market has improved substantially in a context of price stability. In judging when to moderate the pace of asset purchases, the Committee will, at its coming meetings, assess whether incoming information continues to support the Committee's expectation of ongoing improvement in labor market conditions and inflation moving back toward its longer-run objective. Asset purchases are not on a preset course, and the Committee's decisions about their pace will remain contingent on the Committee's economic outlook as well as its assessment of the likely efficacy and costs of such

## B Technical Details of Universal Sentence Encoding Architecture

The USE architecture in this paper is based on six neural network layers, each of which has two sublayers with a self-attention channel. We first describe the original architecture and



Figure A-7: Alternative C FOMC statement in October 2013

Class I FOMC – Restricted Controlled (FR)	Authorized for Public Release	October 24, 2013
<b>FOMC STATEMENT—OCTOBER 2013 ALTERNATIVE C</b>		
<ol style="list-style-type: none"> <li>1. Information received since the Federal Open Market Committee met in July <del>September</del> suggests that economic activity <del>has been expanding</del> <u>continues to expand</u> at a moderate pace. <del>Some</del> Indicators of labor market conditions have shown <del>some</del> further improvement <del>in recent months; in particular</del>, but the unemployment rate, <del>remains</del> <u>though still</u> elevated, <del>has continued to decline</del>. Household spending and business fixed investment advanced, and the housing sector has <del>been strengthening</del>, but <u>continued to strengthen, even though</u> mortgage rates have risen further <del>on balance in recent months</del> and fiscal policy is restraining economic growth. Apart from fluctuations due to changes in energy prices, inflation has been running <del>somewhat</del> below the Committee's longer-run objective, but longer-term inflation expectations have remained stable.</li> <li>2. Consistent with its statutory mandate, the Committee seeks to foster maximum employment and price stability. The Committee expects that, with appropriate policy accommodation, economic growth will pick up from its recent pace and the unemployment rate will gradually decline toward levels the Committee judges consistent with its dual mandate. The Committee sees the downside risks to the outlook for the economy and the labor market as having diminished, on net, since last fall. <del>but the tightening of financial conditions observed in recent months, if sustained, could slow the pace of improvement in the economy and labor market. The Committee recognizes that inflation persistently below its 2 percent objective could pose risks to economic performance, but it anticipates</del> <u>The Committee has become more confident that labor market conditions will continue to improve and</u> that inflation will move back toward its <u>2 percent</u> objective over the medium term.</li> <li>3. Taking into account the extent of federal fiscal retrenchment <del>over the past year</del>, the Committee sees the improvement in economic activity and labor market conditions since it began its asset purchase program <del>a year ago</del> as consistent with growing underlying strength in the broader economy. <del>However, the Committee decided to await more evidence that progress will be sustained before adjusting the pace of its purchases. Accordingly, the Committee decided to continue purchasing additional agency mortgage-backed securities at a pace of \$40 billion per month and longer-term Treasury securities at a pace of \$45 billion per month. In light of the cumulative progress toward maximum employment and the improvement in the outlook for labor market conditions, the Committee decided to make modest downward adjustments in the pace of its of asset purchases. Beginning in November, the Committee will add to its holdings of agency mortgage-backed securities at a pace of [ \$30 ] billion per month rather than \$40 billion per month, and will add to its holdings of longer-term Treasury securities at a pace of [ \$35 ] billion per month rather than \$45 billion per month.</del> The Committee is maintaining its existing policy of reinvesting principal payments from its holdings of agency debt and agency mortgage-backed securities in agency mortgage-backed securities and of rolling over maturing Treasury securities at auction. <del>Taken together, these actions</del> <u>The Committee's sizable and still-increasing holdings of longer-term securities</u> should maintain downward pressure on longer-term interest rates, support mortgage markets, and help to make broader financial conditions more accommodative, which</li> </ol>	Alternatives	

then explain how to fine tune it to obtain the paragraph level decomposition of similarity scoring across statements.

## B.1 Deep Neural Network Layers in the USE

The first neural network in the USE is built by linking two sublayers as shown in Figure A-8 after taking a group of word embeddings that represents the source sentence as input. The first layer generates the sentence embedding vector  $(h_1^1, \dots, h_M^1)$  as output and feeds this as input for the second layer.<sup>27</sup>

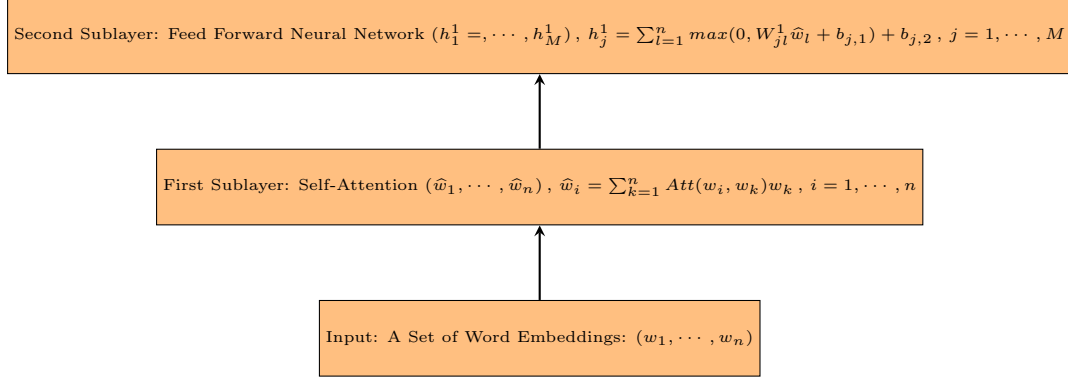


Figure A-8: First Neural Network Layer

Here, attention weights are determined by the distance between different word embeddings as follows:

$$\hat{w}_{i,j} = \sum_{k=1}^{n_i} Att(w_{i,j}, w_{i,k})w_{i,k}, \quad Att(w_{i,j}, w_{i,k}) = \frac{e^{w'_{i,j}w_{i,k}}}{\sum_{l=1}^{n_i} e^{w'_{i,j}w_{i,l}}}. \quad (\text{A-5})$$

The entire USE algorithm works by vertically stacking six neural network layers which take the sentence embedding output in the previous layer as input and generate another sentence embedding as output. Figure A-9 describes the entire process.

To train parameters in the neural network architecture, we need to define the loss function that compares outcomes based on sentence embeddings from the USE with those based on human judgement. For example, if we define the relation between two texts as one of 3

<sup>27</sup>The actual USE architecture is slightly more complicated than presented below. It involves 1)sub-word (character) level embedding, 2)positional embedding in which the order of any given word is also mapped into the embedding of that word, 3)residual connection in which input bypasses attention and feed-forward neural network channels with a certain probability known as the dropout rate, 4)output from the layer is normalized to have mean zero and standard deviation of one, and 5)8 multihead attention channels are applied in the attention sublayer.

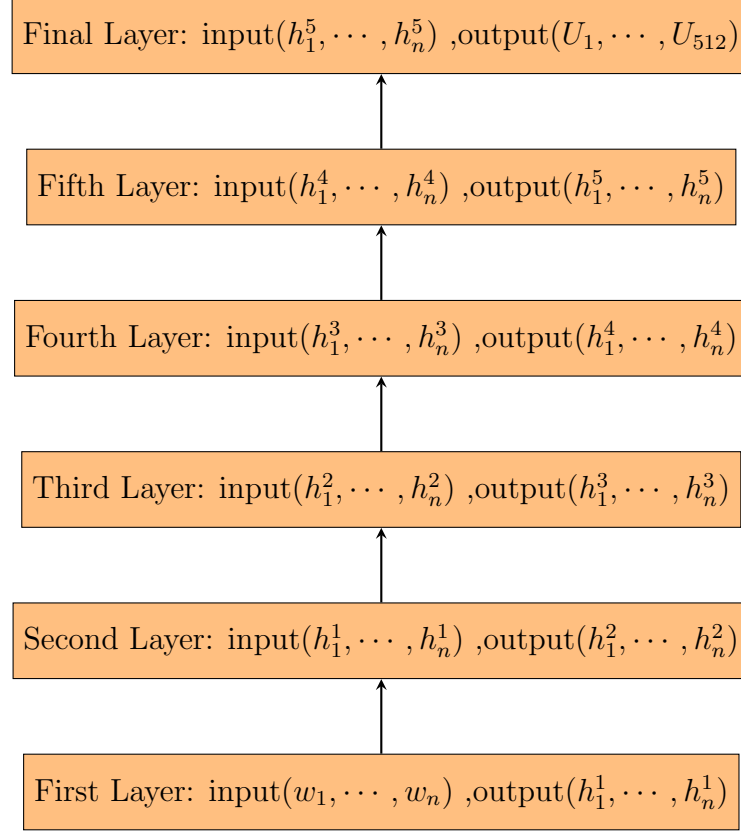


Figure A-9: Neural Network Architecture

classes (entail, contradict, neutral), we can apply the softmax classifier ( $f$ ) to the difference between two embeddings. In this case, we can choose parameters in the neural network architecture to minimize the loss function that measures the distance between the machine-classified outcome ( $f(U^i, U^j)$ ) and the one judged by humans ( $f^{\text{human}}(\text{Text}_i, \text{Text}_j)$ ) where  $U^i$  is the 512-dimensional USE representation of  $\text{Text}_i$ . In addition, two other natural language processing tasks are run to train the model.

- Skip-thought task: conditional on the center sentence, predict neighboring sentences (previous and next). The training dataset is from wikipedia articles.
- Question-answer prediction: predict the correct response for a given question among a list of correct answers and other randomly sampled answers. The training dataset is from web question-answer pages and discussion forums.

- Natural language inference: given a premise sentence and a hypothesis sentence, extract the relation between them. Let  $U_p$  and  $U_h$  be the sentence embeddings of the premise and the hypothesis, respectively. A fully-connect layer and a 3-way softmax classifier are applied for the concatenated input of  $(U_p, U_h, |U_p - U_h|, \text{ and } U_p - U_h)$ . The three-way classifier predicts if the premise entails, contradicts, or is neutral to the hypothesis. The training dataset is the SNLI corpus.

## B.2 Paragraph level USE Representations

In some cases, paragraph-by-paragraph comparison may provide more interpretable results. For instance, we may be interested in which paragraph drives the similarity score between different statements. For this, we obtain paragraph level USE representations and approximate the statement level USE representation by a weighted average of paragraph level USE representations.

Denote the USE representation of the released FOMC statement at time  $t$  by  $S_t^R$ . Similarly,  $S_t^i$ , ( $i = A, B, C, D$ ) denotes the USE representation of alternative statements. The USE representation of the  $j$ -th paragraph of the FOMC statement at time  $t$  is  $P_{j,t}^i$ . To calculate  $P_{j,t}^i$ , we run the USE algorithm for each paragraph  $j$ . The idea is to construct  $\sum_k w_k P_{k,t}^i$  that can mimic  $S_t^i$  best in terms of minimizing the squared difference between two representations of the FOMC statement at time  $t$ .

- **Step 1: Paragraph Padding** Some statements are longer than others, meaning that the corpus of FOMC statements has an unequal length depending on the statement. An easy way to fix this is pad a shorter statement with empty paragraph encodings. Suppose that  $n_{max}$  is the maximum number of paragraph of any given FOMC statement from the entire corpus of our dataset including both released statements and alternative statements. Then, we can extract the following array of the paragraph USE representation of the FOMC statement.

$$P_t^R = [P_{1,t}^R, \dots, P_{n_{max},t}^R]. \quad (\text{A-6})$$

If the number of paragraphs in the statement at time  $t$  ( $n_{R,t}$ ) is smaller than  $n_{max}$ , we add  $(n_{max} - n_{R,t})$  zero vectors of 512 dimensions. The purpose of this operation is to

make the USE representation of any FOMC statement have the same number of the USE representations at the paragraph level.

- **Step 2: Approximate the Statement Level USE Representation by a Weighted Average of Paragraph Level USE Representations**

The goal is to select weights  $(w_j, j = (1, \dots, n_{max}))$  that can mimic this statement-level USE representation using paragraph-level USE representations. We consider the following squared loss:

$$\sum_{i \in R, A, B, C, D} \sum_t (S_t^i - \sum_j w_j P_{j,t}^i)^T (S_t^i - \sum_j w_j P_{j,t}^i). \quad (\text{A-7})$$

We can put the non-negativity and unit-sum constraints on  $w_j$  such that  $w_j \geq 0, \sum_j w_j = 1$ . Once we find the solution for weights, we can mimic  $P_t^i$  by  $\sum_j w_j P_{j,t}^i$ . But the numerical optimization routine might be non-convex when you put the constraints directly. So we may consider the following transformation of  $w_j$  to make the problem an unconstrained minimization problem:

$$w_j = \frac{e^{\alpha_j}}{\sum_{k=1}^{n_{max}} e^{\alpha_k}}, \quad (\text{A-8})$$

where  $\alpha_j$  is an unconstrained parameter. Notice that  $w_j$  still satisfies the constraints but we are minimizing the loss function with respect to  $(\alpha_1, \dots, \alpha_{n_{max}})$ .

- **Step 3: Decomposing the Similarity Scoring**

For the unit-vector, the cosine similarity is simply the inner product. So we can renormalize the USE representation to have a unit length. In that case, we have the following nice decomposition of the similarity scoring between texts.

$$Sim(P_t^i, P_t^j) \propto Sim\left(\sum_{k=1}^{n_{max}} w_k P_{k,t}^i, \sum_{k=1}^{n_{max}} w_k P_{k,t}^j\right) = \sum_k \sum_{k'} w_k w_{k'} Sim(P_{k,t}^i, P_{k',t}^j). \quad (\text{A-9})$$