Shifts in ECB Communication: A Textual Analysis of the Press Conferences*

Justyna Klejdysz[a] and Robin L. Lumsdaine[b]
[a]Ludwig Maximilian University of Munich; Ifo Institute
[b]Kogod School of Business, American University; Econometric Institute, Erasmus University Rotterdam; National Bureau of Economic Research; Center for Financial Stability; Tinbergen Institute

This paper investigates how European Central Bank (ECB) communication, made during the press conference, affects stock market volatility. First, the ECB press conferences are dissected into topics using Latent Dirichlet Allocation (LDA). Then turning points in ECB communication are captured using the estimated topic probabilities. The proposed approach does not rely on subjective interpretation of topical content. The paper finds that (i) the topics reveal communication patterns that match the ECB monetary policy stance, (ii) the content of the ECB press conference is informative for the market, and (iii) market uncertainty increases when the ECB switches to a different communication regime.

JEL Codes: E58, E65, C54, C58, G10.

*This paper builds on thesis work of the first author conducted under the supervision of the second author in partial fulfillment of the degree requirements at Erasmus University Rotterdam. The authors thank participants at the Canadian Econometric Study Group (CESG) 2019 meeting, the Society of Financial Econometrics 2019 Annual Conference, the Warsaw International Economic Meeting 2019, and the seminars at American University, the Banque de France, Imperial College London, Justus Liebig University Giessen, and the University of Hawaii for comments. We are grateful to the editor Sharon Kozicki, two anonymous referees, Peter Tillmann, and Michel van der Wel for insightful suggestions. Klejdysz acknowledges the Society of Financial Econometrics travel grant and support from the Polish Ministry of Finance. Lumsdaine is grateful for the generous hospitality during various visits to the University of Portsmouth in contributing to the completion of this project. Any remaining errors are the responsibility of the authors alone. Views and opinions expressed are those of the authors only and do not necessarily represent the views of any institutions with which they are affiliated.

1. Introduction

This paper considers the problem of quantifying communication of the European Central Bank (ECB) during the press conferences on the Governing Council meeting days. Communication became a key tool for central banks to maintain transparency, manage market expectations, and achieve policy goals in a zero lower bound environment, where the room for maneuvering interest rates was limited (Blinder et al. 2008). Statements explaining monetary policy decisions are scrutinized by financial market participants; however, there is a great deal of subjectivity for a human reader trying to glean information by spotting patterns in multiple long text documents.

The ECB uses various channels to communicate its monetary policy stance: press conferences, monetary policy accounts, monthly bulletins, speeches, and interviews. The press conference that takes place on the same day as the Governing Council decision announcement is the primary communication device. It provides explanations for the monetary policy decision, the core assessment of the economic and monetary situation, and the forward guidance. Two main parts of a typical press conference are (i) an introductory statement, which is agreed upon by the members of the Governing Council, and (ii) a question-and-answer (Q&A) session, when journalists have the opportunity to ask clarification questions. This structure makes the ECB press conference a case study of both prepared and extemporaneous remarks.

A growing body of economic literature applies tools from computational linguistics to analyze central bank communication. The focus of this paper is to use these tools to study how the dynamics of topical composition of the ECB press conference affects stock market volatility on the Governing Council meeting days. The analysis proceeds in two stages. The first stage provides a low-dimensional representation of the transcripts by dissecting the ECB press conferences into topics. The second stage constructs a topic-based measure that captures any changes in the ECB communication regime.

To identify topics, we apply Latent Dirichlet Allocation (LDA) (Blei, Ng, and Jordan 2003), a generative model for text that enables extraction of multiple themes that are not specified in advance. In the analysis, the transcripts are represented by a document-term matrix, with each row representing a single document and each column corresponding to a unique word. The idea is to decompose the document-word relationships into topic probabilities in each document and word probabilities in each topic. Topics are interpreted as latent dimensions underlying the text.

The second part of the analysis is motivated by the communication patterns discovered with LDA. The model can identify phases when a single topic dominates in ECB communication and when a variety of topics is discussed. A novel aspect of our research is the construction of a score based on variations in the probability of the most dominant topic on a given conference day to capture substantial textual changes in the press conferences. The score is derived separately for the decision summary, the economic analysis, the monetary analysis, and the answers provided in the Q&A session during the tenures of Jean Claude Trichet and Mario Draghi. We examine the performance of the measure in explaining stock market volatility with event-based regressions.

The key findings are as follows. First, content exploration with LDA shows clustering of similar topics in each section of the press conference over time. This is expected, as the ECB strives to send a consistent message over time and similar speeches are easier to interpret. Of primary interest, therefore, are fundamental updates to the ECB wording, i.e., periods when one topic dies out and is replaced with a different topic. Comparison of the topic proportions over time with ECB monetary policy decisions shows that the changes in different sections of the introductory statement reflect the changes in the monetary policy regime. Analyzing the Q&A section, LDA identifies a discontinuity in topic probabilities, corresponding to the first press conference held by Mario Draghi.

Second, market volatility is higher on the conference days that the ECB introduces major revisions to the monetary analysis section, as compared with those conference days when the ECB sends a relatively homogeneous message.[1] This suggests that major revisions to the content of the introductory statement are more difficult to digest for the market, even when they occur in conjunction with changes in the monetary policy stance.

[1]Throughout this paper, we use the term "homogeneous message" to refer to statements that are primarily focused on a single topic.

This paper makes three distinct contributions to the field of analyzing central bank communication with computational linguistics tools. First, to our knowledge ours was the first study to apply LDA to the ECB press conferences, although the framework was successfully employed to analyze the statements, minutes, and transcripts of the Federal Open Market Committee (FOMC) (Hansen and McMahon 2016; Fligstein, Brundage, and Schultz 2017; Hansen, McMahon, and Prat 2017; Jegadeesh and Wu 2017).[2] Common alternatives to quantify text in economic literature are hand-coding (Jansen and De Haan 2005; Rosa and Verga 2007) or automated methods that rely on keyword counting (Tetlock 2007; Loughran and McDonald 2011). These approaches are deductive, as they typically capture meaning along a single, predefined dimension, like expansion-contraction or hawkish-dovish. LDA offers several advantages in that it satisfies the following conditions (DiMaggio, Nag, and Blei 2013): (i) it is reproducible; (ii) it is automated, so that it is easily updated when new documents arrive; (iii) it is inductive, to enable content discovery without imposing prior beliefs about what to look for in the text; and (iv) and it recognizes that terms may have different meanings in different contexts.

Second, this paper proposes a new content measure that is derived from LDA output but does not rely on subjective labeling of topics. A persistent challenge when using textual analysis is how to exploit the output to extract information that is relevant for financial market participants or improves the understanding of central bank decision makers. Current applications of LDA to central bank communication often rely on assigning substantive interpretations to topics based on the top most probable words in a topic (Hansen and McMahon 2016; Jegadeesh and Wu 2017). In contrast, our proposed measure captures the degree of discussion homogeneity, circumventing the need for assigning subjective topic labels. To facilitate content exploration, we employ automated measures of topic interpretability in the model selection procedure. By providing a summary of the whole document collection, the model

[2] We subsequently became aware of a small but growing literature that applied topic models to ECB press conferences in other contexts; see, e.g., Diessner and Lisi (2019), Dybowski and Kempa (2019), and Vo (2019). Since the initial writing of our paper, the use of topic models to analyze ECB communication has become quite popular.

not only enables study of the extent to which consecutive speeches are similar but also (i) what wording makes the speeches similar, (ii) whether the topics are recurring, and (iii) how long the transition period to a new topic is.

The third contribution is methodological. LDA is a hierarchical Bayesian model, where the hyperparameters that index prior distributions on a set of latent variables are found to substantially influence model inference (Asuncion et al. 2009; Wallach, Mimno, and McCallum 2009; George and Doss 2018). We adopt a fully Bayesian approach to formally infer the values of hyperparameters. In contrast, to date, textual analyses in economics commonly has chosen the values of the hyperparameters in an ad hoc manner (Griffiths and Steyvers 2004).

The structure of the paper is as follows. Section 2 reviews strategies to quantify text in economic research, and Section 3 presents the methodology of LDA. Section 4 describes the data and text preprocessing steps. Section 5 investigates the estimated topics and the shifts in ECB communication. Section 6 concludes. An extensive set of appendices, containing estimation details and documenting other decisions made in conducting the analysis, then follows.

2. Related Literature

Our work lies in the intersection of two strands of literature: the impact of central bank communication on the financial market, and natural language processing (NLP), in particular topic modeling. This section provides an overview of the methods for mapping words to meaningful quantities within economic literature, with a focus on central bank communication.

The literature on central bank communication uses three approaches to gauge the effect of communication: an indirect approach, manual coding, and automated textual analysis. The automated methods are most relevant for this paper. The indirect approach does not quantify verbal information. Instead, using high-frequency data, it measures financial market movements within a narrow window of the decision announcement and surrounding communication. A stylized fact following from indirect analyses is that the market reaction to central bank communication is more pronounced than the reaction to monetary policy decisions (Gürkaynak, Sack, and Swanson 2005; Ehrmann and Fratzscher 2009; Brand, Buncic, and Turunen 2010). Furthermore, for the ECB the market reaction to the press conference is stronger for less anticipated decisions, indicating that the introductory statement provides relevant clarifications (Ehrmann and Fratzscher 2009). The reasoning behind this result is that in times of high uncertainty (when the surprise component in a policy decision is large), the reaction to the actual decision is muted, as the market expects a subsequent explanation and instead responds to that.

A step further is to identify pieces of information that move the markets. The information can come either in the form of topics or tone. To extract the information content, one can follow a manual or an automated approach. The manual approach involves hand-coding the statements on an ordinal scale or classifying verbal expressions to predefined categories. For example, Ehrmann and Fratzscher (2009) manually classify real-time newswire reports during the ECB press conference via the following content categories: economic outlook, inflation, second-round effects, money growth, and interest rates. Statements on inflation and interest rates turn out to be the most important market movers. By hand-coding each ECB introductory statement on a scale ranging from -2 (very dovish) to 2 (very hawkish), Rosa and Verga (2007) find that ECB words are complementary to data on macroeconomic variables in predicting the moves in the key ECB interest rate, showing that market expectations react to the unexpected component of the press conference content. The main criticism of the manual approach is high subjectivity and low reproducibility. Furthermore, because the manual coding is done ex post, coders might unintentionally mitigate the unexpected component in the statement and fail to capture how the financial markets understood the message at the release time (Blinder et al. 2008).

To overcome these issues, a strand of literature has turned to automated approaches to ensure that the analysis is transparent and scalable. Overall, within the automated methods one can either define a priori dimensions to look for in the text, or apply an algorithm to discover dimensions. In the former case, the most intuitive and relatively simple technique is a dictionary method, where a researcher predefines a list of keywords describing meanings of interest. Documents are then summarized by the number of occurrences of each word in the keyword list. In principle, by defining word lists that separate multiple categories, it is possible to capture multiple dimensions in text (Tetlock 2007); however, typically only two opposing concepts are considered. For example, the word counts can be converted to a single communication measure of incremental changes in hawkish and dovish monetary policy inclinations (Apel and Grimaldi 2012), positive and negative tone (Tetlock, Saar-Tsechansky, and Macskassy 2008; Jegadeesh and Wu 2013; Born, Ehrmann, and Fratzscher 2014) or periods of greater or less uncertainty (Jegadeesh and Wu 2017).

One of the main difficulties with the dictionary approach is developing a word list that accurately captures the meaning for a specific application. Since words often carry different sentiment or meaning in different contexts, dictionaries developed in one domain of study can lead to word misclassification when used in other disciplines (Loughran and McDonald 2011). This necessitates development of methods that are customized to central bank communication. Previous literature has gone beyond the generic dictionaries by capturing contextual information with (i) field-specific dictionaries (ii) sentence-level scores, and (iii) intensity of specific themes. Looking at the ECB, Picault and Renault (2017) manually develop a field dictionary based on the introductory statements and (similarly to our paper) investigate the European stock market reaction to the press conference. They find that market volatility increases (decreases) when the statements about monetary policy are hawkish (dovish) and the tone about the economic outlook is negative (positive). Instead of considering word occurrences in isolation, Lucca and Trebbi (2009) devise a sentence-based score and show that discourse orientation in the FOMC statements explains a large portion of the federal funds rate variation. Finally, analysis of thematic content highlights the importance of shifts in discourse reflected by changing topic intensity over time—for example, the increasing role of financial stability in monetary policy considerations (Peek, Rosengren, and Tootell 2016; van Dieijen and Lumsdaine 2018).

In this paper we employ a probabilistic topic model, specifically LDA, to identify the most important dimensions in the ECB press conferences. The central application of topic models is summarizing a large collection of documents and discovering patterns in textual data. However, the topics themselves are rarely the final objective of the analysis. Although there are examples where topic models mainly augment descriptive analysis (Quinn et al. 2010; Fligstein, Brundage, and Schultz 2017), recent applications to central bank communication attempt to derive communication measures using estimated topics, often in combination with dictionary methods in order to understand how the information in central bank communication affects market returns, volatility, and interest rate expectations (Hansen and McMahon 2016; Jegadeesh and Wu 2017). A closely related work to our paper is Jegadeesh and Wu (2017) who use LDA to investigate how the U.S. stock market reacts to proportions of discussion on, and tone adopted in, different topics in the FOMC minutes. They find that the Federal Reserve's discussion of its policy stance and inflation is most informative for the market, whereas topics like trade and consumption are not informative. Unlike the above implementations, we avoid deriving conclusions that depend on subjective interpretations of topics, instead focusing on the properties of the estimated document-topic probabilities, similar to Hansen, McMahon, and Prat (2017).

3. Methodology

3.1 Latent Dirichlet Allocation

Latent Dirichlet Allocation (LDA) introduced by Blei, Ng, and Jordan (2003) is a mixed membership model for text. The basic idea is that observations (words) are grouped into documents and each of these groups (documents) is modeled with a mixture of distributions. The components of the mixture are topics, which are multinomial probability distributions over a fixed vocabulary. The topics are shared across all documents (each document is built from the same components), but the proportions of topics in documents vary.

LDA ignores both the document order and the word order within the documents. A document is represented as a "bag of words." The inference is based on the notion of word co-occurrence. Words that often appear together across documents are likely to belong to the same topic. Intuitively, LDA trades off two conflicting goals in finding a good topical representation for a collection of documents, that of assigning words in each document to few topics versus assigning a high probability in each topic to few words (DiMaggio, Nag, and Blei 2013).

The central inferential problem in LDA is to determine the posterior distribution of topic proportions in documents ($\boldsymbol{\Theta}$), word proportions in topics ($\boldsymbol{\Phi}$), and word-topic assignments ($\mathcal{Z}$), given the hyperparameters ($\boldsymbol{\alpha}$, $\boldsymbol{\beta}$) and corpus of documents, $\mathcal{W}$. The formal statement of this idea, details on the notation used, and derivations are contained in Appendix A. A corpus is a collection of documents, where D is the number of documents, N_d is the number of words in document d, and V is the number of distinct words in the collection of documents. It is assumed that each document is composed of K topics in different proportions. The posterior distribution is proportional to the complete data likelihood function times the prior:

$$p(\boldsymbol{\Phi}, \boldsymbol{\Theta}, \mathcal{Z}|\mathcal{W}, \boldsymbol{\alpha}, \boldsymbol{\beta}) \propto \prod_{d=1}^{D} \underbrace{p(\boldsymbol{\theta_d}|\boldsymbol{\alpha})}_{Dirichlet} \prod_{k=1}^{K} \underbrace{p(\boldsymbol{\phi_k}|\boldsymbol{\beta})}_{Dirichlet} \left(\prod_{d=1}^{D} \prod_{i=1}^{N_d} \underbrace{p(w_i^{(d)}|z_i^{(d)}, \boldsymbol{\Phi})}_{Multinomial} \underbrace{p(z_i^{(d)}|\boldsymbol{\theta_d})}_{Multinomial} \right). \tag{1}$$

Implementation of LDA involves important model specification and selection decisions. In particular, the estimation results vary according to the number of topics (K) and hyperparameter settings ($\boldsymbol{\alpha}$, $\boldsymbol{\beta}$). The goal is to obtain $p(\boldsymbol{\Phi}|\mathcal{W}, \boldsymbol{\alpha}, \boldsymbol{\beta}), p(\boldsymbol{\Theta}|\mathcal{W}, \boldsymbol{\alpha}, \boldsymbol{\beta})$ and $p(\mathcal{Z}|\mathcal{W}, \boldsymbol{\alpha}, \boldsymbol{\beta})$, that is, the word probabilities, topic probabilities, and word-topic assignments. These distributions cannot be computed in closed form. As estimation is straightforward, we omit details here and note that a thorough discussion of the estimation method and how it builds on more common strategies for approaching LDA is contained in Appendix A.

3.2 Model Evaluation

Choosing the number of latent topics and assessing their quality is a long-studied problem in unsupervised topic modeling. Typically, there is a trade-off between predictive accuracy of the model and topic interpretability (Chang et al. 2009).

Metrics of predictive performance, such as held-out likelihood or perplexity, are conventionally used to assess model quality (Blei, Ng, and Jordan 2003; Wallach et al. 2009).[3] However, the predictive metrics have limitations. Usually fine-grained, highly specific topics yield the best model fit, but they are not easy to interpret or to generalize (Boyd-Graber, Mimno, and Newman 2014). Furthermore, predicting the content of the preprocessed text is rarely the objective of research in political, economic, or social science applications.

Roberts et al. (2014) argue that a semantically interpretable topic has two qualities: (i) it is coherent—the highest probability words for the topic tend to co-occur within documents, and (ii) it is exclusive—the words that have high probability under one topic have low probabilities under other topics.

Our model selection procedure prioritizes interpretation over prediction. We first discard solutions with the lowest coherence or exclusivity, akin to Roberts et al. (2014), and then select the solution with the lowest perplexity among the remaining models.

3.2.1 Coherence and Exclusivity

Automated measures of coherence usually assume that co-occurrence frequency of terms within documents is informative about semantical relatedness of the terms and are based on averaging some measure of pairwise association between the most probable words in a topic (Newman et al. 2010).

The models estimated on the corpus of the ECB press conferences are evaluated with a semantic coherence score of Mimno et al. (2011). The score is shown to match well with human judgments and is defined as

$$\text{Coherence}_k = \sum_{j=2}^{N} \sum_{i=1}^{j-1} \log \frac{D(w_i^{(k)}, w_j^{(k)}) + 1}{D(w_i^{(k)})}, \tag{2}$$

where $D(\cdot)$ is a function that returns the number of documents containing all of the words provided as arguments, and $w_i^{(k)}$ denotes a word from the list of top N words with the highest probability in topic k. Intuitively, the measure is related to the conditional probability of observing a word given another higher-ranked word. The

[3]Perplexity is defined as the inverse of the geometric mean per-word likelihood of the test data; see Appendix A for discussion.

semantic coherence of Mimno et al. (2011) relies on the word frequencies in documents being modeled, hence it is more intrinsic in nature.

Coherence measures inform inference about internal consistency of topic representation, but they do not penalize topics that are similar (Roberts et al. 2014). A counterpoint to semantic coherence is topic exclusivity that captures inter-topic similarity by comparing the usage rate of words with high probability in a topic relative to other topics. Exclusivity of term v in topic k is defined as (Bischof and Airoldi 2012; Airoldi and Bischof 2016):

$$\text{Exclusivity}_{v,k} = \frac{\phi_{k,v}}{\sum_{i=1}^{K} \phi_{i,v}}. \tag{3}$$

In other words, the exclusivity score is the probability of a word in a topic divided by the sum of probabilities of this word in all topics. Exclusivity of topic k is computed as an average of the scores for the top N words in that topic. A high exclusivity score indicates that the most common words in a particular topic are not common to other topics.

3.2.2 Topic Cardinality and Word Ranking

Topic-based measures of coherence and exclusivity operate on a ranking of the top N words with the highest probability. The standard practice is to select N arbitrarily (usually $N = 10$). To achieve more stable evaluation, we compute semantic coherence (2) and exclusivity (3) for different topic cardinalities: $N = 5, 10, 15, 20$ and average them (Lau and Baldwin 2016).

The word ranking based on term probability in a topic favors terms with high frequency in a corpus, but the most common words might not carry any semantically useful information, and can be used similarly in every topic. We use a FREX (frequency-exclusivity) score of Bischof and Airoldi (2012) to represent a topic with words that are both frequent and exclusive. The score combines these two dimensions via the harmonic mean of frequency and exclusivity:

$$\text{FREX}_{v,k} = \left(\frac{\omega}{\text{ECDF}(\text{Exclusivity}_{v,k})} + \frac{1-\omega}{\text{ECDF}(\phi_{k,v})} \right)^{-1}, \tag{4}$$

where ECDF is the empirical cumulative distribution function and ω is a weight given to exclusivity (set to 0.5).

3.3 Measuring Tone

We additionally compare the estimated topics with lexicon-based tone-measures.[4] Following Shapiro and Wilson (2021), we calculate the net negativity score as the difference between the fraction of negative and positive words after text preprocessing, based on the list of positive and negative words provided in the Loughran and McDonald (2011) dictionary (LM dictionary). The computations ignore negations. The score can be computed on different levels: across sections or for the whole statement.

4. Data

This section introduces the ECB press conference and describes the steps to convert text to numerical data. It also presents the financial data used to measure the market reaction to the topic dynamics of the press conference.

4.1 The ECB Press Conference

The ECB's monetary policy decisions are published at 13:45 CET on the day of the Governing Council monetary policy meeting. The press conference starts at 14:30 CET on the same day. It begins with an introductory statement by the ECB President, who explains the monetary policy decision.

The press conference consists of six major sections: (i) summary of the ECB's monetary policy decision; since July 2013 this summary also includes forward guidance; (ii) economic analysis; (iii) monetary analysis; (iv) "cross-check" paragraph; (v) fiscal policy and structural reforms; (vi) question-and-answer (Q&A) session.

The economic analysis and monetary analysis sections are the two pillars by which the Governing Council evaluates the risks to price stability. The economic analysis part looks at the short- to medium-term outlook, whereas the monetary analysis part assesses

[4]We are grateful to an anonymous referee for suggesting this analysis.

medium- to long-term trends. The cross-check paragraph was introduced in 2003 and its role is to compare signals from the two pillars.[5]

The ECB held monetary policy meetings and related press conferences on a monthly basis until December 2014. From 2015 the frequency of the meetings was changed to a six-week cycle. Our analysis considers all ECB press conferences between January 2004 and April 2018, covering 91 speeches from Jean-Claude Trichet (whose eight-year term expired at the end of October 2011) and 65 speeches from Mario Draghi. The textual data have been scraped from the ECB website.[6]

4.2 Preparing the Documents

The ECB press statements have a standardized structure, with sections that are fixed over time. Each section defines a main theme that is easily captured at the preprocessing stage (i.e., via the section headings), but latent topics within the theme are more difficult to capture by the human reader and vary over time. Importantly, the sections are sufficiently long to enable us to run LDA on them separately. For our purposes, therefore, a document is defined at the section level and a separate model is estimated for each section. The main motivation for treating the sections separately is to allow us to track the topics within sections and compare the changes across sections. In addition, focusing on the sections separately provides more confidence about the context in which words should be understood, alleviating drawbacks of the "bag-of-words" representation, as well as being able to assess whether substantial updates occur to just a single part of the statement or to multiple sections simultaneously. Running the LDA on the whole introductory statement would likely lead to topics that are dominated by the general sectional themes that are already known before running the model.

[5]In May 2003 the ECB introduced the new structure of the introductory statement in which economic analysis is discussed first and monetary analysis is put second. The ECB motivated this decision by stating that "the Governing Council wishes to clarify communication on the cross-checking of information in coming to its unifed overall judgement on the risks to price stability" (European Central Bank 2003).

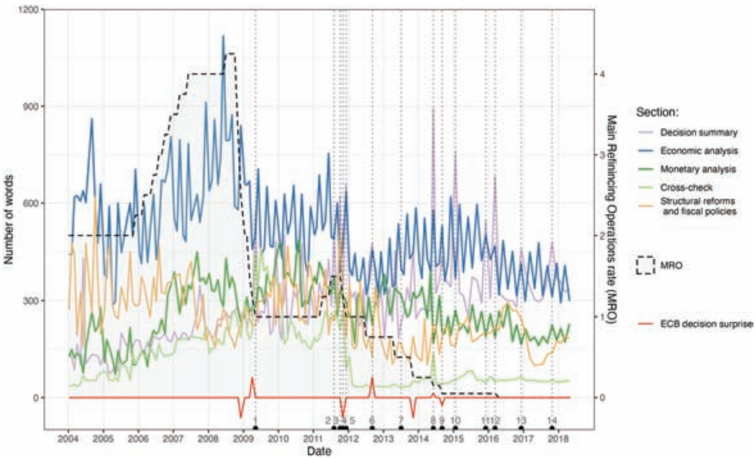
[6]See https://www.ecb.europa.eu/press/pressconf (accessed April 2018).

For each press conference we (i) break the transcript into individual paragraphs; (ii) assign each paragraph to a section; and (iii) extract answers from the Q&A session. We use keywords that are defined as bold word sequences in the HTML code of the press conference to record the section where each paragraph is located. For example, a paragraph that contains the keyword "key ECB interest rates" is identified as the first paragraph of the decision summary, and a paragraph containing the keyword "economic analysis" marks the beginning of the economic analysis section.

Figure 1 shows how the number of words per section of the introductory statement evolved over time, along with the main refinancing operations rate (MRO, dashed line), monetary policy surprise, and decisions regarding non-standard monetary policy measures. The surprise component (red line) is measured by subtracting the Bloomberg® survey median forecast from the ECB rate announcement (Bloomberg L.P. 2018). Based on the raw word counts, economic analysis is given broader coverage than monetary analysis. In addition, the ECB appears to have communicated relatively more on the economic outlook when it was raising the interest rate (until about the mid-2008) than when it cut the interest rate. We base this inference on a comparison of the relative number of words devoted to the economic analysis section versus the monetary analysis section.[7] The spikes in the number of words in the decision summary can be matched with ECB announcements about new monetary policy tools and implementation details. In addition, since Mario Draghi became the ECB President in November 2011, the coverage of the cross-check part has sharply decreased and currently contains only a single sentence that the cross-check of the monetary analysis and the economic analysis confirms the need for the undertaken monetary policy action. Because of LDA's deficiency in handling documents that are too short and the low informational value of the cross-checking over Draghi's tenure, the cross-check section is not considered in the estimation.

[7]We note, however, that this pattern may just reflect the time period associated with these episodes, with the cuts being driven by the European sovereign debt crisis rather than the overall state of the economy, and hence may not hold generally.

Figure 1. Number of Words per Section of the ECB Introductory Statement and the Main Refinancing Operations (MRO) Rate



Note: The figure shows the raw word counts in five sections of the introductory statement: decision summary, economic analysis, monetary analysis, cross-check, structural reforms, and fiscal policies. The dashed line represents the level of the main refinancing operations rate. The red line shows the policy decision surprise which is measured by subtracting the Bloomberg® survey median forecast from the ECB rate announcement (Bloomberg L.P. 2018). The timeline markers represent the following events: 1. Announcement of the first covered bond purchase program (CBPP1) and one-year longer-term refinancing operation (LTRO); 2. Announcement of six-month LTRO; 3. Announcement of CBPP2; 4. Announcement of three-year LTRO, collaterals, and reserve ratio. 5. The first introductory statement by Mario Draghi; 6. Introduction of the forward guidance; 7. Announcement of the outright monetary transactions (OMT); 8. Announcement of targeted longer-term refinancing operations (TLTROs); 9. Announcement of CBPP3 and the asset-backed securities purchase program (ABSPP); 10. Announcement of the expanded asset purchase program (APP, known as quantitative easing); 11. Announcement about extension of APP; 12. Announcement of the corporate sector purchase program (CSPP) and TLTRO2; 13. Announcement about extension of APP; 14. Announcement about unwinding of the stimulus.

4.3 Vocabulary Selection

Text preprocessing choices can substantially affect model output (Boyd-Graber, Mimno, and Newman 2014; Denny and Spirling 2018). Common text treatments are removing punctuation and numbers, lowercasing, stop word removal, term normalization (stemming or lemmatization), n-gram inclusion, and removing very common or very rare words.

First, we remove neutral sentences or parts of sentences that introduce the next section and are repeated in every speech, for example: "Ladies and gentlemen, the Vice President and I are very pleased to welcome you to our press conference", "Let me now explain our assessment in greater detail, starting with the economic analysis", "We are now at your disposal for questions". The complete list of expressions that were removed is provided in Appendix B. We also clean from the Q&A section the answers in French, because English translations of these answers (that are included in the analysis) immediately follow.

The second step is to convert all words to lowercase, remove punctuation, stop words, and month names. Stop words are common function words like "the" or "and" with no inherent useful information, and their overwhelming presence in all documents can produce spurious associations between content words (Roberts et al. 2014).[8] We also remove all words containing non-alphabetic characters, except for abbreviations for money aggregates (M1, M2, M3) and groups of countries (G3, G7, G8, etc.).

The third step is term normalization: each term is classified into its part of speech (POS) using the Stanford POS tagger (Collobert et al. 2011) and reduced to its dictionary form by lemmatization.[9] An alternative approach to reduce inflectional and derivational word forms is stemming. We opt to use a lemmatizer because it is more accurate than a stemmer and it is unlikely to over-conflate (Schofield and Mimno 2016). See Appendix B for discussion.

Finally, we identify collocations and create multiword expressions, called n-grams, which allow one to capture the broader context of a word and reduce ambiguities resulting from the "bag-of-words" assumption.[10] The list of n-grams, available from the authors on

[8]The stop word list is from http://snowball.tartarus.org/algorithms/english/stop.txt (accessed April 2018). It includes pronouns, articles, prepositions, and conjunctions.

[9]The Stanford POS-tagging algorithm is used to provide auxiliary information about the part of speech for the WordNet lemmatizer in Python.

[10]Specifically, we manually rate bigrams and trigrams occurring at least 50 times to filter out n-grams that do not add new meaning beyond the meaning of

Table 1. Data Dimensionality Reduction After Preprocessing Steps

	Raw	Stop Words Removal and Lemmatization	Creating n-grams
Total Words	776,112	365,930	326,343
Average Section Length	829	391	349
Unique Words (Vocabulary Size, Overall)	9,224	6,120	6,216
Decision Summary	1,798	1,252	1,337
Economic Analysis	1,805	1,263	1,346
Monetary Analysis	1,589	1,042	1,094
Cross-Check	901	654	699
Structural Reforms, Fiscal Policies	2,381	1,675	1,741
Q&A	8,827	5,935	6,021

Note: This table reports descriptive statistics of the vocabulary in the ECB press conferences before and after implementing the preprocessing steps: stop words removal, lemmatization, and creating n-grams. Generally, using n-grams increases the vocabulary size; for example, one might have the bigram "monetary_policy" and the unigrams "monetary" and "policy" in the vocabulary list. The last column presents the vocabulary size used in the estimation of Latent Dirichlet Allocation (LDA).

request, includes technical terms used by the ECB such as "full allotment" or "covered bond", expressions providing context for very common words, like "key ecb interest rate unchanged", as well as long-used statements specific to ECB communication, such as the premise to "never pre commit" to any future policy action.

Table 1 reports descriptive statistics of the vocabulary before and after implementing the preprocessing steps.

4.4 Financial Data

We use daily closing values of the VSTOXX index to measure investors' reaction to ECB communication patterns on press

---

the constituting words (i.e., "regard second", "question say"). We then sort the n-grams by pointwise mutual information (Bouma 2009) to select those n-grams that occur less often but where the association between words is high (i.e., "tail risk" or "banca italia"). Selected n-grams are treated as separate words in the analysis. Words that appear without collocation stay as separate words.

conference days. The VSTOXX index represents the implied volatility of the Euro Stoxx 50 index (EURO STOXX 50 real-time option prices) and is designed to reflect market expectations of near-term volatility. The index was also investigated in the context of ECB communication and monetary policy actions by Fratzscher, Lo Duca, and Straub (2016) and Picault and Renault (2017), and is often used as a proxy for uncertainty in the euro area. The daily closing values of the VSTOXX index for stock market volatility are sourced from Bloomberg®. The series is log-transformed and differenced to approximate the daily percentage change.

A number of control variables are considered in the empirical investigation: the surprise component of the ECB interest rate decision, a dummy variable for the announcements regarding non-standard monetary policy measures (the complete list of the announcements is presented in Figure 1), German two-year government bond yields, and the surprise component of the U.S. jobless claims. The data on German government bond yields, the MRO rate, and released values of the U.S. jobless claims are collected from Bloomberg®. The sample period for the financial variables is from January 2004 to April 2018. All surprise components are constructed by deducting the Bloomberg® survey median expectations of professional forecasters from the released value.

5. Results

This section describes the main findings. It starts with general remarks about model selection and properties of the estimated topic-word and document-topic distributions. Next, it investigates the changing attention to different topics over time.

5.1 Estimated Topics

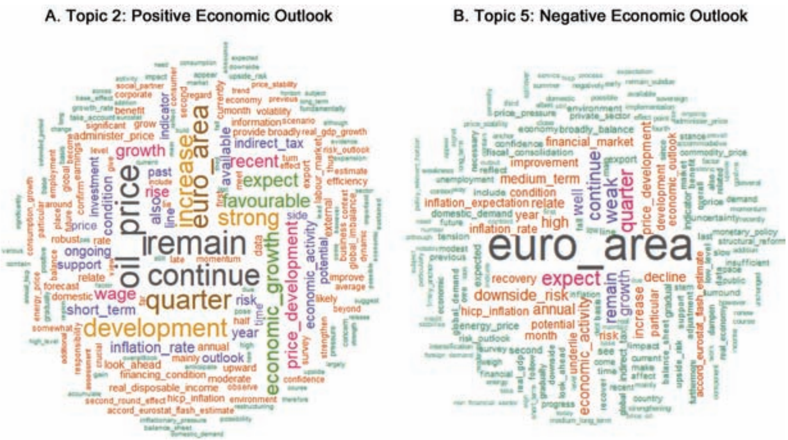
The model (1) is estimated for the four separate sections of the press conference—decision summary, economic analysis, monetary analysis, and Q&A—using a different number of topics for each. In line with the findings of Chang et al. (2009), higher model complexity (more topics) results in lower perplexity, but also in lower average coherence. Exclusivity does not seem to be related to semantic coherence, confirming that the two measures capture distinct aspects of topic interpretability. The set of solutions with the highest coherence and exclusivity is dominated by relatively parsimonious models. The selected dimensionality varies across sections, but it does not exceed 10 topics. Diagnostic plots illustrating model selection are presented in Appendix C.

We find that document-topic distributions are generally sparse in all sections, i.e., few topics comprise a document. The conclusion about sparsity of the document-topic distributions does not change if a different number of topics is specified. Furthermore, LDA groups the press conferences in time although no information about the order of documents is incorporated in the estimation procedure. The sparsity of document-topic distributions and the similarity of consecutive documents lead to identification of different phases of ECB communication. Although the sections of the press conference were considered separately in the estimation, the algorithm identifies a rise of a new topic in each section at approximately the same time.[11]

As expected, frequent words in the corpus often end up scattered across the top most likely words in many topics. The term re-ranking using the FREX score downgrades general terms and corpus-specific stop words and reveals intuitive topic interpretations based on keywords that are both frequent and exclusive to a specific topic. For example, Figure 2 presents word clouds that show relative (raw) frequency of words in the two most popular topics in the economic analysis section. The more frequent a term is in a topic, the larger it becomes in a cloud. If topics are represented solely by their most frequent terms, they would be described by non-exclusive words and many topics in the economic analysis section would appear to be similar. On the other hand, re-ranking words by the FREX score captures important differences between words in the topics, suggesting

[11] It is worth stressing that the topic sparsity in the ECB press conferences is not detected if one follows the heuristics about Dirichlet prior parameters (Griffiths and Steyvers 2004), widely applied in economic research, instead of estimating the hyperparameters. The heuristic ($\alpha = \frac{50}{K}$) imposes that the document-topic distribution is smooth for $K < 50$. In line with the heuristic regarding the Dirichlet prior parameter for topic-word distributions ($\beta = 0.1$) the estimated word-topic distributions are sparse: there is a limited number of words with relatively high probability in each topic.

Figure 2. Distributions over Terms Represented as Word Clouds



Note: The word clouds show the top 200 most frequent terms in two topics of the economic analysis section. The topics were estimated using Latent Dirichlet Allocation (LDA) algorithm. In the word clouds the size of a term is proportional to the term probability. If topics were represented in terms of their most frequent terms, they would be described by non-exclusive, high-frequency words (for example, "euro area", "continue", "remain") and many topics in this section would appear to be similar. However, the FREX score gives high ranks to keywords that are both frequent and distinctive for a specific topic. Top terms ranked by the FREX score in topic 2 (left) are *side, robust, economic growth, earnings, favourable, efficiency, lie, short term, consumption growth,* whereas top terms ranked by the FREX score in topic 5 (right) are *weak, low level, economic outlook, gradual, public, expected, modest, insufficient, global demand, slow.*

one of them be labeled as "Positive economic outlook", and the other one as "Negative economic outlook" (see Table 2).

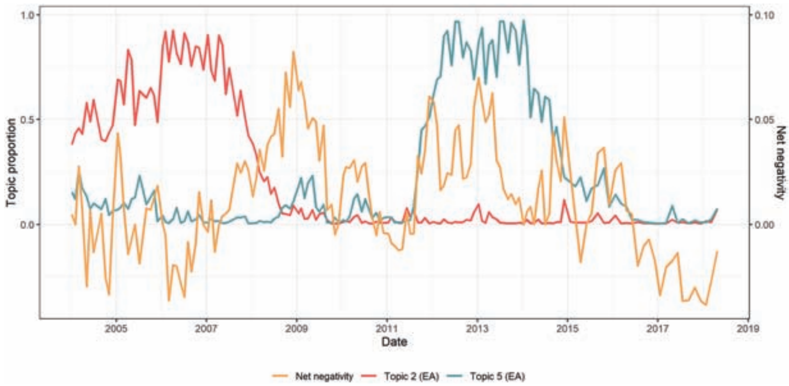
Topic labels are consistent with the tone that is associated with the occurrence of those topics. Figure 3 shows the proportion of the two topics of the economic analysis section, topic 2 (labeled as "Positive economic outlook") and topic 5 (labeled as "Negative economic outlook", along with the net negativity score (right scale) computed for the economic analysis (EA) section. During the time that topic 2 dominates, the net negativity score is lower. Conversely, the net negativity score is higher when topic 5 dominates the discussion. Importantly, these topics represent more than just

Table 2. Top 10 Terms Describing Topics of the Economic Analysis Section, Ranked by the FREX Score

1 "Projections"	**2 "Positive Economic Outlook"**	**3 "Wage-Price Spiral"**
ecb	side	scheme
eurosystem	economic_growth	avoid
range	robust	party
staff_macroeconomic_projection	earnings	food_price
projection	efficiency	sound
revise	favourable	behaviour
staff_projection	oil_price	shock
foresee	lie	constraint
upwards	short_term	power
downwards	consumption_growth	call
4 "Stimulus"	**5 "Negative Economic Outlook"**	**6 "Recovery"**
correction	weak	private_consumption
function	low_level	economic_recovery
stimulus	economic_outlook	exchange_rate
macroeconomic	gradual	measure
owing	public	structural_reform
financial_system	expected	monetary_policy
commodity_price	underlie	closely
aim	modest	geopolitical_risk
keep	insufficient	pick
protectionist_pressure	global_demand	household

Note: The FREX score gives high ranks to terms that are both frequent and exclusive.

Figure 3. Proportion of Topic 2 (labeled as "positive economic outlook") and Topic 5 (labeled as "negative economic outlook"), Along with the Net Negativity Score, for the Economic Analysis (EA) Section



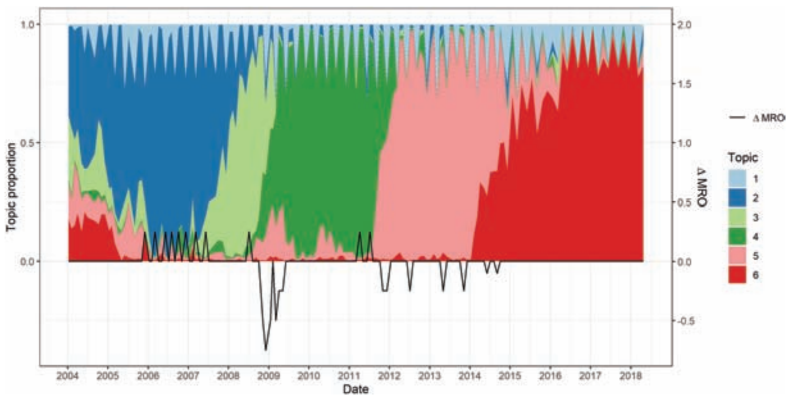
Note: This figure plots the proportion of the economic analysis section devoted to two topics along with the net negativity score for the section. The topics were estimated using the LDA algorithm (Blei, Ng, and Jordan 2003). The net negativity score is the difference between the fraction of negative and positive words in the section, based on the list of positive and negative words provided in the Loughran and McDonald (2011) dictionary.

the tone, however. Topic 2 is not consistently active in every time period that net negativity is lower, and topic 5 is not consistently active in every time period that the net negativity is higher.

5.2 Interpreting Topical Content

As external validation of the ECB communication patterns identified by LDA, we compare the attention to different topics with changes in the main refinancing operations rate in order to analyze how different communication regimes correspond to the phases of the ECB monetary policy stance. For the interpretation of textual themes, we focus on the economic analysis section and the Q&A section. The results obtained for the remaining sections are provided in Appendix D.

Figure 4. Topics in the Economic Analysis Section over Time



Note: This figure plots the proportion of the economic analysis section devoted to each topic along with the ECB MRO rate decisions. The topics were estimated using the LDA algorithm (Blei, Ng, and Jordan 2003). The sample comprises 156 transcripts of the section from the ECB press conferences between 2004 and 2018.

Figure 4 graphs topic proportions over time in the economic analysis section. The key terms of topic 1 ("staff macroeconomic projection", "range", "revise", "upwards", "downwards"; see Table 2) appear to capture a discussion about macroeconomic projections. The topic is especially active on the press conference days in March, July, September, and December when the quarterly staff macroeconomic projections are presented.

The remaining topics in the economic analysis section can be reasonably associated with various phases of the ECB monetary policy stance. Topic 2 remains strong during the tightening phase 2005–07. The topic is mostly characterized by both frequent and exclusive terms such as "robust", "favourable", and "efficiency", emphasizing a positive economic outlook. It declines shortly after the sequence of the rate hikes; its proportion falls permanently below 50 percent on the meeting in December 2007, whereas the last rate hike in the sequence occurred in June 2007.

Topic 3 is the most prominent during the first phase of policy responses to the financial turmoil that started in August 2007 (Stark 2009). In that period the ECB particularly often used the keyword "scheme" to express the concern about a wage-price spiral, but in general the fundamentals of the euro-area economy were described as “sound”.[12]

The bankruptcy of Lehman Brothers in September 2008 marks the intensification of the crisis and precedes an abrupt change in ECB communication. Topic 4 surges in November 2008, exactly on the first conference day the ECB cut its key interest rate by 50 basis points after the Lehman collapse.[13] Distinctive for this phase is a discussion about “financial system” and “stimulus”. This phase ended with two interest rate increases in April and July 2011, which turned out to be premature (Constâncio 2018).

The rise of topic 5 marks the start of the recession in the third quarter of 2011 that lasted until the first quarter of 2013, according to the Euro Area Business Cycle Network.[14] This phase is associated with the easing cycle where the language used by the ECB (“weak”, “low level”, “modest”, “insufficient”, “slow”) reflected the weakness of the economy.

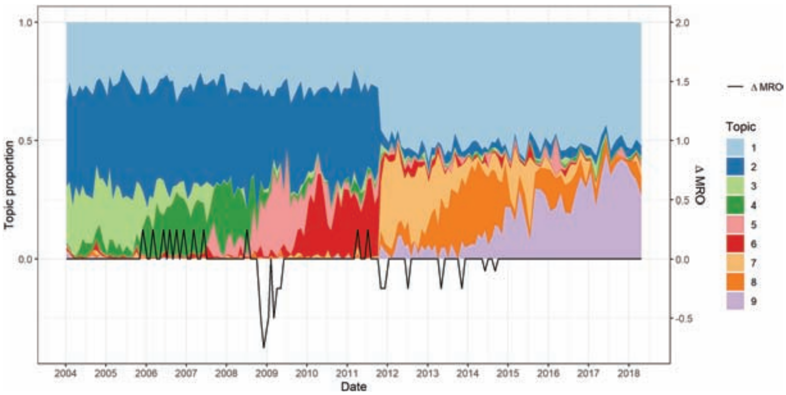
The discourse represented by topic 6 was emerging gradually, as the interest rates were approaching the zero lower bound. The timing coincides with the ECB’s introduction of its unconventional monetary policy instruments and hence predominant for topic 6 is the keyword “monetary policy measure”, but the other frequent and exclusive terms are “economic recovery”, “structural reform”, “exchange rate”, “household”, and “private consumption”. Interestingly, a reading of the statements confirms that the ECB expressed concerns about exchange rate developments, discussed the structural reforms, private consumption, and the situation of the households as a part of its economic analysis solely in the statements where topic 6 is active (2004–05 and 2013–17) and never in between. What

[12]The ECB has repeatedly used the term “scheme” and “shock” in the following context: “the Governing Council is concerned about the existence of schemes in which nominal wages are indexed to consumer prices. Such schemes involve the risk of upward shocks in inflation leading to a wage-price spiral” (press conference, July 3, 2008).

[13]The first press conference after the Lehman collapse was held on October 2, 2008 and the decision was to keep the interest rates unchanged. The first interest rate cut in response to the financial crisis was unscheduled. It took place on October 8, 2008 as a part of the coordinated action with other major central banks.

[14]See https://eabcn.org/dc/chronology-euro-area-business-cycles (accessed April 2023).

Figure 5. Topics in the Q&A Section over Time



Note: This figure plots the proportion of the Q&A section devoted to each topic along with the ECB MRO rate decisions. The topics were estimated using the LDA algorithm (Blei, Ng, and Jordan 2003). The sample comprises 156 transcripts of the section from the ECB press conferences between 2004 and 2018.

is common to these two periods is that both concern the phase of the economic recovery. The recovery discussed in 2004–05 followed the protracted period of economic slowdown experienced from mid-2001 to mid-2003 (European Central Bank 2009). This suggests that there might exist some recurring textual patterns of central bank communication, although the current sample is too short to draw definitive conclusions between communication patterns and the business cycle.

Turning to the Q&A session, recall that following the introductory statement the ECB has the opportunity to clarify its messages and emphasize its point of view about the economic outlook. Therefore, the questions may reveal ambiguities in ECB communication or indicate topics that reporters find important. In contrast to the introductory statement, which is prepared by the whole Governing Council, the answers provided by the ECB president during the Q&A session are non-prompted. Therefore, we can expect differences between the Q&A session and the other sections, as well as between the wording used by Jean-Claude Trichet and Mario Draghi.

Figure 5 shows the topical representation of answers provided during the Q&A session. Several interesting points emerge. A spontaneous section, in comparison with the prepared sections, appears to have a larger proportion of words that do not contribute to the informational value of the president's response, for example "mean", "come", "particular", or "already" (see Table 3). As expected, LDA with an asymmetric prior on the document-topic distribution was able to handle these very common words in an appropriate fashion and sequester them into topics 1 and 2.

There is a discontinuity in the topics' probabilities occurring at the first conference held by Mario Draghi in November 2011. The discontinuity in the time series of topics 1 and 2 may reflect different speaking styles of both presidents, but there is also a clear split among specialized topics discussed during the tenures of Trichet and Draghi.

Starting with the answers of Trichet, the attention to the topic "Vigilance" was dominating in advance of and during the tightening phase in 2005–07. This observation is in line with Jansen and De Haan (2007), who found that the term "vigilance"/"vigilant" was used extensively in ECB communication starting in March 2004 and continued to be mentioned after the tightening cycle, but less often. The code word "vigilance" used to be a clear signal for financial markets that the ECB would pre-announce any interest rate hike.[15] Topic 5 also has a natural label. It clusters terms related to various liquidity-injecting operations provided to the banking sector (main refinancing operations and longer-term refinancing operations (LTROs)).

The Q&A sessions held by Draghi seem to be richer in content. The focal points are the forward guidance and non-standard monetary policy measures (LTROs—long-term refinancing operations, OMTs—outright monetary transactions, the asset purchase program), the Greek crisis, and ELA (emergency liquidity assistance, on which the Greek banks have been highly dependent since being cut off from standard ECB funding options).

5.3 Shifts in ECB Communication

We exploit the feature of the ECB statements that topics emerge and disappear over time to investigate whether the transition periods in

[15] According to Reuters, June 22, 2011: "The ECB used the phrase 'strong vigilance' in March before increasing rates in April. It also used the phrase repeatedly during its 2005–07 rate hike cycle, typically one month before it raised rates, although there were exceptions to that rule."

Table 3. Top 10 Terms Describing Topics of the Q&A Section, Ranked by the FREX Score

1 "General Terms"	**2 "General Terms"**	**3 "Vigilance"**	**4**	**5 "Liquidity"**
one	particular	vigilant	correction	commercial_bank
point	already	vigilance	episode	refinance
way	line	body	labour_productivity	supply
let	present	homework	counter	decrease
time	respect	diagnosis	social_partner	refinance_operation
first	observe	banca_italia	dynamic	bold
second	anchor	erm_ii	economist	exceptional
mean	mention	financial_environment	salary	money-market
come	necessary	invite	counterpart	unlimited
also	credible	favourable	m1	main_refinance_operation
6	**7 "LTRO/OMT"**	**8 "Low Inflation"**	**9 "QE"**	
head	esm	abs	asset_purchase	
restore	ltros	weak	eurozone	
doctrine	access	low_inflation	sustained	
phase	omts	ssm	objective	
ahead	omt	weakness	npls	
peer	compact	supervisor	mostly	
governance	ltro	forward_guidance	path	
advanced_economy	risk_aversion	lending	qe	
commensurate	ela	energy_price	constâncio	
public_finance	contraction	subdue	towards	

Note: The FREX score gives high ranks to terms that are both frequent and exclusive. We do not find a clear interpretation for topics 4 and 6, therefore the topics have no labels.

ECB communication increase market volatility. The market reaction is measured with the VSTOXX index.

The aim is to derive a simple topic-based measure that captures the phases of ECB communication when the message was relatively homogeneous, that is, primarily focused on a single topic. An intuitive approach is to define a summarizing communication measure for each section as a proportion of the topic with the highest probability on a conference day:

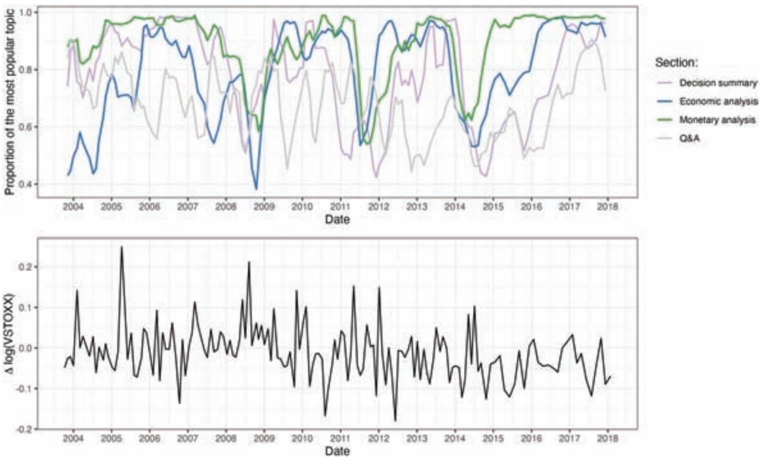
$$I_d = \max_{k \in 1,\ldots,K} (\hat{\theta}_{d,k}). \tag{5}$$

Large values (near one) imply that ECB communication is dominated by a single topic, whereas small values represent a situation where a variety of topics is discussed.

The analysis is constrained to four sections: decision summary, economic analysis, monetary analysis (the latter two referring to the two pillars of the ECB decision making), and the Q&A session (because of its unique clarification role). Because the purpose of the communication measure is to analyze how topics change over time, we ignore topics that constitute a fixed part of discussion. First, we omit the topic on macroeconomic projections, because the words in this section mainly capture the vocabulary used to describe the numerical projections but not their meaning (see Table 2). While the new rounds of quarterly macro projections may affect the ECB narrative, these narrative changes should be reflected in different topics in the other sections. Second, topics 1 and 2 in the Q&A section are omitted, because they place high probability on general terms (corpus-specific stop words). The stop words are frequent, but not content-bearing (see Table 3). In the sections mentioned above, the probability of the dominating topic from the set of remaining topics is then normalized by dividing by the sum of topic probabilities in this set.[16] Figure 6 graphs the communication measures derived from LDA document-topic distributions.

[16]For the economic analysis section the measure is $I_d^{EA} = \max_{k \in \{2,\ldots,6\}} (\hat{\theta}_{d,k}) \times \frac{1}{\Sigma_{i \in \{2,\ldots,6\}} \hat{\theta}_{d,i}}$. For the Q&A section it is $I_d^{QA} = \max_{k \in \{3,\ldots,9\}} (\hat{\theta}_{d,k}) \times \frac{1}{\Sigma_{i \in \{3,\ldots,9\}} \hat{\theta}_{d,i}}$.

Figure 6. Topic-Based Communication Measures for Four Sections of the ECB Press Conference: Decision Summary, Economic Analysis, Monetary Analysis, and Q&A

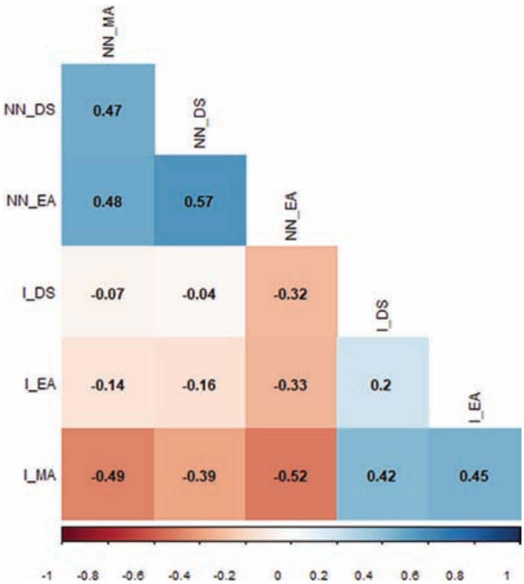


Note: This figure plots the topic-based communication measures, smoothed using a three-point moving-average filter (top panel), and the daily percentage change (close to close) of the VSTOXX on the day of the ECB press conference (bottom panel). Topic-based communication measures are constructed as the probability of the dominant topic in a specific section.

5.3.1 Co-movement of Topic-Based Communication Measures and Tone

The topic-based communication measures in Figure 6 appear to co-move, indicating that the ECB updates the different sections at approximately the same time. The rise of new topics can be linked to the intensification of financial market tensions and changes in policy stance. Updates in topics can also be accompanied by a change in tone. The association between topics and tone is illustrated in Figure 7, which shows the correlation matrix of the net negativity (NN) and topic-based communication measures (I, for sections of the introductory statement; DS—decision summary, EA—economic analysis, MA—monetary analysis). The strongest correlations of the tone-based measures are with the topics of the monetary analysis section. A lower focus on a single topic (or transition to a different topic) in each section is associated with higher net negativity

Figure 7. Correlation Matrix for the Topic-Based Communication Measures ($I\_$) and Net Negativity ($NN\_$) Scores in the Decision Summary (DS), Economic Analysis (EA), and Monetary Analysis (MA) Sections



Note: This figure shows the correlation between the topic-based communication measures and the net negativity scores for each section. The net negativity score is the difference between the fraction of negative and positive words in the section, based on the list of positive and negative words provided in the Loughran and McDonald (2011) dictionary. A negative correlation between $I\_$ and $NN\_$ indicates that a lower focus on a single topic (or transition to a different topic) in a section is associated with higher net negativity expressed in this section.

expressed in the monetary analysis. Figure 8 gives further insights into topic concentration in the monetary analysis and economic analysis sections. It shows that a change in topics (lower topic concentration) is usually accompanied by a spike in the net negativity score. This observation indicates that the change in narrative in the statements is often accompanied by a change in tone.

5.3.2 The Impact of ECB Communication

We analyze the impact of ECB communication through event-based regressions, where only statement days are considered. The empirical

Figure 8. Topic-Based Communication Measures for the Economic Analysis and the Monetary Analysis Sections, Along with the Respective Net Negativity Scores for Each Section

investigation is complicated by the fact that the ECB press conference always takes place on the same day that a monetary policy decision is announced. To control for the effects of policy actions, we include the absolute surprise component in the ECB monetary policy decision, as in Rosa and Verga (2007) and Ehrmann and Fratzscher (2009).[17]

[17]In this paper we measure the surprise component using the median expected monetary policy rate from the Bloomberg® survey. We also considered the monetary policy surprise as captured by high-frequency interest rate changes in the press release window using the Euro Area Monetary Policy Event-Study Database developed by Altavilla et al. (2019), of which we were subsequently made aware. The results are qualitatively similar and are not reported here but are available from the authors on request. We thank Peter Tillmann for the suggestion to use the alternative surprise measure.

Following Coenen et al. (2017), to account for non-standard policy measures we include a dummy variable for the days when such measures were announced and the absolute change in German two-year government bond yields, which is intended to capture the absolute surprise component in decisions about unconventional monetary policy tools. To control for other macroeconomic news, the surprise component of the U.S. jobless claims releases on Thursdays at 8:30 ET is included, as it coincides with the ECB press conference.[18] Appendix E provides descriptive statistics and correlations.

The event-based regressions are nested in the following equation:

$$\begin{aligned}\Delta V_t = \alpha &+ \beta_1|s_t^{MRO}| + \beta_2 D_t^A + \beta_3|r_t^{DE}| + \beta_4|s_t^{JC}| + \beta_5\Delta V_{t-1}\\ &+ \beta_6 I_t^{DS} + \beta_7 I_t^{EA} + \beta_8 I_t^{MA} + \beta_9 I_t^{QA}\\ &+ \beta_{10} I_t^{QA} \times D_t^{Draghi} + \beta_{11} NN_t + \varepsilon_t, \qquad (6)\end{aligned}$$

where ΔV_t denotes the daily percentage change in the VSTOXX index on the conference day t relative to the previous day; s_t^{MRO} and s_t^{JC} are surprise components of the MRO rate and the U.S. jobless claims, respectively; D_t^A is an indicator for announcements regarding non-standard monetary policy measures; r_t^{DE} is a daily change in German two-year government bond yields; and $I_t^{DS}, I_t^{EA}, I_t^{MA}, I_t^{QA}$ denote the index values that capture changes in communication by section: decision summary, economic analysis, monetary analysis, Q&A. The communication score for the Q&A section is also interacted with an indicator variable for presidency (D_t^{Draghi}). In addition, we control for the net negativity score (NN_t) calculated jointly for the economic analysis and monetary analysis sections (the difference between the fraction of negative and positive words in those sections based on LM dictionary). Table 4 presents the estimation results.[19]

[18] In the sample period there were seven press conferences that took place on Wednesday instead of Thursday. Also, four times a year the ECB/Eurosystem macroeconomic projections are published on the ECB website. Including an indicator variable to control for the timing of these macroeconomic projections does not affect our qualitative conslusions and hence the results are omitted.

[19] The model selection necessarily involved human judgment in balancing multiple criteria (exclusivity, coherence, predictive power). As a robustness check, we construct the communication variable for each section using the output of a topic model with the number of topics greater by one or less by one than in the baseline

Table 4. Regression Results

	Dependent Variable: ΔV_t			
	(1)	(2)	(3)	(4)
$\|s_t^{MRO}\|$	–0.115	–0.150	–0.137	–0.134
	[0.333]	[0.209]	[0.239]	[0.261]
ΔV_{t-1}	–0.017	0.017	0.071	0.071
	[0.876]	[0.881]	[0.518]	[0.518]
r_t^{DE}	–0.176**	–0.167**	–0.156*	–0.156*
	[0.039}	[0.049]	[0.060]	[0.060]
$\|s_t^{JC}\|$	0.001	0.0004	0.0003	0.0003
	[0.111]	[0.298]	[0.452]	[0.443]
D_t^A	–0.023	–0.023	–0.014	–0.014
	[0.240]	[0.253]	[0.483]	[0.489]
I_t^{DS}		0.030	0.020	0.020
		[0.374]	[0.538]	[0.547]
		(0.008; 0.043)	(–0.001; 0.033)	(0; 0.033)
I_t^{EA}		–0.031	0.0001	–0.0001
		[0.375]	[0.998]	[0.998]
		(–0.052; –0.015)	(–0.026; 0.014)	(–0.025; 0.014)
I_t^{MA}		–0.096*	–0.116**	–0.121**
		[0.059]	[0.021]	[0.035]
		(–0.119; –0.052)	(–0.133; –0.067)	(–0.14; –0.061)
I_t^{QA}		0.028	0.025	0.024
		[0.483]	[0.517]	[0.531]
		(0.008; 0.045)	(0.005; 0.042)	(0.004; 0.043)
$I_t^{QA} \times D_t^{Draghi}$			–0.050***	–0.051***
			[0.005]	[0.006]
			(–0.053; –0.044)	(–0.053; –0.044)
NN_t				–0.049
				[0.869]
Constant	–0.019**	0.052	0.069	0.074
	[0.020]	[0.262]	[0.133]	[0.181]
Observations	156	156	156	156
Adjusted R^2	0.034	0.058	0.103	0.097

Note: p-values are in brackets and the sampling uncertainty (5th to 95th percentile) is in parentheses. The sampling uncertainty is computed based on 400 draws from the posterior distribution. *p < 0.1; **p < 0.05; ***p < 0.01.

model. We also test for a topic model that strictly dominates other models in terms of both coherence and exclusivity. Table F.1 in Appendix F reports the estimates. The results are qualitatively similar to the baseline specification.

In the regressions, we use the values of the communication measures derived from the matrix of document-topic distributions averaged across 400 draws from a Markov chain. As a result, there is uncertainty arising from the sampling algorithm used to estimate topics. The regression analysis is repeated for each draw to obtain a distribution of the effect, similarly to Hansen, McMahon, and Prat (2017). Table 4 therefore also reports the range of the 5th to 95th percentiles of these distributions.

The major transitions in ECB communication regarding monetary analysis contain incremental information about the ECB monetary policy decisions not already incorporated in market expectations, after controlling for announcements about non-standard monetary policy measures. The uncertainty proxied by the VSTOXX index is on average lower when the ECB sends a homogeneous message by focusing primarily on a single topic than in times of transitions to a different topic.[20] The results in Table 4 show that the implied volatility on the conference day decreases by approximately 1.16 percentage points when the proportion of the most dominant topic in the monetary analysis section increases by 10 percentage points. This effect is significant at the 5 percent level. The results suggest that increased focus on a single topic in the monetary analysis section at any point in time is associated with less market volatility.

Conversely, we also see episodes where the lack of focus (as seen by a period of transition between topics) was associated with higher market volatility. Looking across the four topics in the monetary analysis section over time, the three key dates when the topics shifted permanently are October 2, 2008; October 6, 2011; and May 8, 2014 (see Figure D.2 in Appendix D). The shift in topics is reflected in substantial updates to the statements. In each case, prior to the shift, certain sentences appear repeatedly in the statements, and after the

[20]This finding is reminiscent of that of Ehrmann and Talmi (2020), who show that market volatility decreases when consecutive central bank communications are semantically similar; for the ECB, similarity has increased over time and the effect has similarly strengthened. We emphasize, however, that in this paper we focus on the within-statement coherence and do not explicitly consider cross-statement similarity. We thank an anonymous referee for drawing our attention to this paper.

shift new elements appear or replace the previous sentences and are subsequently repeated.[21]

October 2008 marked a key turning point in the global financial crisis, as the month began with great uncertainty following the September 29 stall where the U.S. House of Representatives voted down the plan put forth by the U.S. Department of Treasury to address the unfolding crisis. After five days of turmoil, the Troubled Assets Relief Program (TARP) was passed on October 3, bringing financial markets around the world back from the brink of collapse. Looking at 10 statements preceding the statement on October 2, 2008, the ECB was repeatedly pointing at prevailing upside risks to price stability at medium to longer-term horizons, the underlying strength of monetary expansion, and temporary factors which may overstate the impact of monetary expansion.[22] There was little evidence that the financial market turbulence since early August had strongly influenced the availability of bank credit in Europe. The ECB in its statements confirmed that the borrowing by non-financial corporations had remained strong. A large drop in the probability of topic 1 occurred at the meeting on October 2, just one day before the TARP passage. The ECB explained that the latest available data (from August) had not embodied the impact of the intensification of the financial market turmoil, and spent more time explaining substitution effects. From the next meeting onward (November 6, 2008) the ECB was pointing at the diminishing impact of upside risks to price stability and an identifiable impact of financial market tensions. The average change in the volatility index rose by 4.8 percentage points, from 1.6 percent to 6.4 percent, on the press conference days during the time when the concentration index fell below 0.6, compared with 10 meetings before October 2008 when the concentration index was above 0.8.

October 2011 marked the point when the European sovereign debt crisis was intensifying and threatening the banking sector. The

[21] These findings are consistent with those of Ehrmann and Talmi (2020), who find an increase in government bond yield volatility when statements change greatly relative to the previous statement and a larger effect the longer the string of similar statements that preceded it.

[22] This is evident in the list of terms under Topic 1 in Table E.2 in Appendix E.

ECB meeting on October 6, 2011 was Jean-Claude Trichet's last meeting before Mario Draghi took the helm. The common parts of the statements preceding the statement on October 6, 2011 were the dampening impact of the steepening yield curve on M3, the stable size of the balance sheets of the banks with the expanding provision of credit to the private sector, and strengthened growth of loans to non-financial corporations. Overall, the monetary expansion was described as moderate and inflationary pressures were contained. The ECB was calling on banks to take appropriate measures to further strengthen their capital. Starting with the statement from November 3, 2011 (the first under Draghi) the ECB introduced a new narrative, focusing on factors related to the intensification of financial market tensions and their negative effects on monetary developments. Prior to the October 6 meeting, the monetary analysis concentration index hovered above 0.95; it then fell below 0.6 and remained at that level until February 9, 2012. The average change in the volatility index on the press conference days during this time (October 2011–February 2012) rose by 2.6 percentage points from –3.6 percent to –1 percent. In comparison, during the 10 meetings before October 2011, the concentration index was above 0.9 (excluding the meeting on August 4, 2011, when the exceptionally high volatility resulted from overall anxiety in both Europe and the United States about deepening economic problems and the uncertainty over the ECB purchasing bonds of Italy and Spain).

May 2014 marked the last month that deposit rates were positive. In the statements preceding the statement on May 8, 2014, the ECB was acknowledging subdued monetary and credit dynamics, reflecting the state of the business cycle, heightened credit risk, and the ongoing adjustment of financial and non-financial sector balance sheets. Up to May 2014, the ECB was repeatedly expressing its concerns about the transmission of monetary policy to the financing conditions in euro-area countries, the fragmentation of euro-area credit markets, and the resilience of the banking sector. Starting in June, a comprehensive package of non-standard policy measures was gradually introduced in order to improve credit conditions. Prior to the May 8 meeting, the monetary analysis concentration index hovered above 0.9; it then fell below 0.7 and remained at that level until November 2014. The average change in the volatility index rose by

0.7 percentage points, from –2.7 percent to –2 percent, on the press conference days during the time when the concentration index was below 0.7, compared with 10 meetings before May 2014, when the concentration index was above 0.9.

The changing composition of the decision summary is not significant. This is expected, as any effect of this section should be already subsumed into the effect of announcements about the policy rate and non-standard monetary policy measures. Although stock market volatility was on average higher under the leadership of Trichet than under Draghi, the changing composition of the Q&A session is on its own not informative for the market. This result agrees with the findings of Ehrmann and Fratzscher (2009), who analyze the reaction of three-month EURIBOR (euro-area interbank offered rate) futures and find that the Q&A session does not systematically add information beyond that contained in the introductory statement, suggesting that in most cases the introductory statement provides sufficient clarity.

5.3.3 Consideration of Other Market Variables

Because we are interested in the influence of ECB communication on market uncertainty, we have to this point focused on the VSTOXX, as it is designed to reflect market participants' expectations of near-term volatility and is often used as a proxy for uncertainty in the euro area. Our interest in uncertainty is primarily motivated by the observed pattern in the press conferences, namely that at distinct points in time the ECB introduces substantial updates to the statements. Specifically, we are interested in how market participants perceive and digest these new narratives and whether these changes are abrupt and visible to market participants in a way that increases their uncertainty. New content in the statements can raise uncertainty if it is unexpected or insufficient (i.e., it does not contain sufficient explanations of the economic or monetary situation that motivated the change), compared with the content that was repeated in the statements before the major update and was familiar to market participants. It is perhaps not surprising that the monetary analysis section is where we find a significant effect on market volatility, as this section usually discusses monetary and financial conditions in the euro area. In this subsection, we consider whether there are similar effects for other market variables.

We estimate the event-based regressions using the daily percentage change in a number of other market variables, namely the Euro Stoxx Index, the Euro 50 Stoxx (on which the VSTOXX is based), the DAX, and the CAC 40 (see Table 5). As expected, the communication variables also affect the euro-zone market indices themselves, particularly the monetary analysis section index and the Q&A interaction with the indicator variable for Draghi's tenure. The positive and significant coefficients are suggestive of the idea that a higher concentration in one topic is good for stock markets.[23] Consistent with our intuition, the effect on the FTSE 100 is significant but weaker; a similar result is obtained using changes in the Nikkei, but there it is the economic analysis index rather than the monetary analysis one that appears to have a modest effect.

In addition, we verified that the communication variables do not seem to affect an unrelated volatility indicator by performing the same regression as in the paper using the Japan VIX index (VXJ), the implied volatility index of the Nikkei. As suspected, the communication indices do not affect the Japan VIX, strengthening the argument that the communication indices matter primarily for euro-area uncertainty.

6. Conclusions

In this paper we empirically search for the main communication patterns in the ECB press conferences and analyze how shifts in those patterns affect a key stock market volatility index on the Governing Council meeting days. Using a generative model for text, we decompose each section of the press conference into a set of coherent and exclusive topics. This approach has the potential to reveal previously understudied dimensions in the transcripts.

---

[23]We thank an anonymous referee for raising a series of questions that led us to the results in this subsection and for suggesting this interpretation. We are intrigued by this possibility but note that at this point it is a somewhat speculative conclusion and warrants a more comprehensive investigation.

Table 5. Regression Results Using the Euro Stoxx Index (SXXE), the Euro 50 Stoxx (SX5E), the DAX, the CAC 40, the FTSE (UKX), the Nikkei 225 (NKY), and the Japan VIX (VXJ)

	Dependent Variable						
	$\Delta SXXE$ (1)	$\Delta SX5E$ (2)	ΔDAX (3)	ΔCAC (4)	ΔUKX (5)	ΔNKY (6)	ΔVXJ (7)
$\lvert s_t^{MRO} \rvert$	0.091***	0.091***	0.105***	0.096***	0.053**	0.015	–0.117
	[0.001]	[0.001]	[0.0003]	[0.0005]	[0.019]	[0.572]	[0.188]
Lagged Dependent Var.	0.192*	0.192*	0.119	0.183*	0.132	–0.070	0.028
	[0.051]	[0.051]	[0.227]	[0.058]	[0.124]	[0.445]	[0.730]
r_t^{DE}	0.042**	0.042**	0.035*	0.045**	0.055***	0.059***	–0.087
	[0.020]	[0.020]	[0.064]	[0.018]	[0.0005]	[0.002]	[0.160]
$[s_t^{JC}]$	–0.0001	–0.0001	–0.0001	–0.0001	–0.0001	–0.0002**	0.0003
	[0.379]	[0.379]	[0.262]	[0.389]	[0.414]	[0.039]	[0.356]
D_t^A	–0.003	–0.003	–0.006	–0.004	–0.003	0.007	–0.008
	[0.449]	[0.449]	[0.212]	[0.411]	[0.392]	[0.107]	[0.594]
I_t^{DS}	–0.001	–0.001	–0.003	–0.0003	–0.002	–0.002	0.004
	[0.871]	[0.871]	[0.669]	[0.972]	[0.700]	[0.823]	[0.861]
I_t^{EA}	–0.007	–0.007	–0.006	–0.008	–0.001	0.014*	–0.017
	[0.345]	[0.345]	[0.436]	[0.332]	[0.826]	[0.073]	[0.538]
I_t^{MA}	0.022*	0.022*	0.022*	0.022*	0.015	0.012	–0.006
	[0.077]	[0.077]	[0.087]	[0.086]	[0.147]	[0.325]	[0.881]
I_t^{QA}	0.001	0.001	0.005	0.001	0.005	0.013	–0.030
	[0.950]	[0.950]	[0.601]	[0.903]	[0.494]	[0.139]	[0.306]
$I_t^{QA} \times D_t^{Draghi}$	0.007*	0.007*	0.008**	0.007*	0.004	–0.002	0.013
	[0.059]	[0.059]	[0.046]	[0.095]	[0.242]	[0.587]	[0.312]
NN_t	–0.049	–0.049	–0.034	–0.053	–0.010	0.049	0.446**
	[0.437]	[0.437]	[0.618]	[0.423]	[0.853]	[0.469]	[0.045]
Constant	–0.014	–0.014	–0.017	–0.015	–0.015	–0.027**	0.024
	[0.238]	[0.238]	[0.193]	[0.228]	[0.161]	[0.031]	[0.557]
Observations	156	156	156	156	156	156	156
Adjusted R^2	0.153	0.153	0.134	0.146	0.116	0.093	0.030

Note: *$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

The results show that similar ECB press conferences are clustered in time even though nothing in our approach imposes such clustering. The main topics surge, die out over time, and rarely reappear in the analyzed sample period, 2004–18. Market volatility increases when the ECB substantially updates its wording in the monetary analysis section, as compared with keeping it rather static relative to the previous period, controlling for the unexpected components in standard and non-standard monetary policy measures. The revisions to the ECB narrative in general accompany the changes in policy direction, but the results suggest that shifts in ECB communication introduce incremental volatility above and beyond that created by a change in policy stance. Although in our paper we do not consider similarity of consecutive statements explicitly, but rather within-statement homogeneity, our findings corroborate the results of Ehrmann and Talmi (2020), who show that market volatility increases when substantial changes occur following a sequence of similar statements.

The main contribution to the current literature that applies computational linguistics tools to analyze central bank communication is a new topic-based communication measure that does not depend on subjective interpretations of topics. Furthermore, the topic model is estimated using a fully Bayesian approach rather than making ad hoc choices about model hyperparameters. Estimating hyperparameters reveals specific features of modeled transcripts. Although we use a "bag-of-words" algorithm that does not incorporate document ordering, the results demonstrate the ability of Latent Dirichlet Allocation to identify series of documents that change the current discourse. We emphasize, however, that the model does not fully eliminate the need to read statements in order to understand what they are about or to guide modeling decisions. Nonetheless it sheds light on how a central bank introduces new policy narratives and to what extent markets are sensitive to those transitions.

Appendix A. Estimation Details

This appendix first derives the posterior distribution of latent variables given the observed words, in the context of Latent Dirichlet Allocation. It then provides a discussion of choices in model specification and an overview of two popular strategies to approximate the posterior distributions in LDA: Markov chain Monte Carlo (MCMC) methods—in particular, collapsed Gibbs sampling (Griffiths and Steyvers 2004). The Metropolis-within-Gibbs sampling approach, which extends upon collapsed Gibbs sampling, is then presented as the preferred estimation method.

A.1 Latent Dirichlet Allocation

Latent Dirichlet Allocation (LDA), introduced by Blei, Ng, and Jordan (2003), is a mixed membership model for text. The basic idea is that observations (words) are grouped into documents and each of these groups (documents) is modeled with a mixture of distributions. The components of the mixture are topics, which are multinomial probability distributions over fixed vocabulary. The topics are shared across all documents (each document is built from the same components), but the proportions of topics in documents vary.

To formalize this idea, let D be the number of documents, N_d is the number of words in document d, V is the number of distinct words (vocabulary size) in a collection of documents (a corpus), K is the number of topics. The corpus is denoted as $\mathcal{W} = \{\boldsymbol{w}^{(1)}, \ldots, \boldsymbol{w}^{(D)}\}$, where $\boldsymbol{w}^{(d)} = \{w_i^{(d)}\}_{i=1}^{N_d}$ is the collection of words in document d and $w_i^{(d)} \in \{1 : V\}$ is the i-th word in document d. Let $\mathcal{Z} = \{\boldsymbol{z}^{(1)}, \ldots, \boldsymbol{z}^{(D)}\}$ denote topic assignments, where $\boldsymbol{z}^{(d)} = \{z_i^{(d)}\}_{i=1}^{N_d}$ and $z_i^{(d)} \in \{1 : K\}$ is a topic assignment for word $w_i^{(d)}$.[24] Let $\boldsymbol{\Theta}$ be a $D \times K$ matrix of topic proportions in documents and $\boldsymbol{\Phi}$ is a $K \times V$ matrix of word probabilities. $\boldsymbol{\theta}_d$ is a K-dimensional vector of topic proportions in document d where $\theta_{d,1}, \cdots, \theta_{d,K}$ are positive random variables that sum to 1. Similarly, topic k, $\boldsymbol{\phi}_k$, is a V-dimensional vector where $\phi_{k,1}, \cdots, \phi_{k,V}$ are positive random

[24]Blei, Ng, and Jordan (2003) define $z_i^{(d)}$ and $w_i^{(d)}$ as unit vectors of length K and V, respectively, that contain a single 1 in the k-th or v-th element, respectively, and zero otherwise, $k = 1, \ldots, K$ and $v = 1, \ldots, V$. Such defined multidimensional variables have the multinomial distribution. In general, a multinomial vector contains counts that sum to n. Because in our case $n = 1$, $z_i^{(d)}$ and $w_i^{(d)}$ can be defined as unidimensional variables with $p(z_i^{(d)}|\boldsymbol{\theta}_d) = \prod_{k=1}^{K} \theta_{d,k}^{I(z_i^{(d)}=k)}$ and $p(w_i^{(d)}|\boldsymbol{\phi}_k) = \prod_{v=1}^{V} \phi_{k,v}^{I(w_i^{(d)}=v)}$.

variables that sum to 1. It is assumed that K and V are known and fixed. The generative process for text is as follows (Blei, Ng, and Jordan 2003):

(i) For document $d = 1, \ldots, D$ choose the topic proportions $\boldsymbol{\theta}_d \backsim Dirichlet(\boldsymbol{\alpha})$, where $\boldsymbol{\alpha}$ is a K-dimensional hyperparameter.

(ii) For topic $k = 1, \ldots, K$ choose the word distribution $\boldsymbol{\phi}_k \backsim Dirichlet(\boldsymbol{\beta})$, where $\boldsymbol{\beta}$ is a V-dimensional hyperparameter.

(iii) For document $d = 1, \ldots, D$: for word $i = 1, \ldots, N_d$:
  (a) choose the topic $z_i^{(d)} \backsim Multinomial(\boldsymbol{\theta}_d)$;
  (b) choose the word $w_i^{(d)} \backsim Multinomial(\boldsymbol{\phi}_{zij})$.

We only observe the set of documents, $\mathcal{W}$. The underlying topic assignments $\mathcal{Z}$, word probabilities $\boldsymbol{\Phi}$, and topic proportions in documents $\boldsymbol{\Theta}$ are latent; $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ are concentration hyperparameters that are selected in advance.

The central inferential problem in LDA is to determine the posterior distribution of topic proportions in documents ($\boldsymbol{\Theta}$), word proportions in topics ($\boldsymbol{\Phi}$), and word-topic assignments ($\mathcal{Z}$).

The joint posterior density is

$$\begin{aligned} p(\boldsymbol{\Phi}, \boldsymbol{\Theta}, \mathcal{Z}|\mathcal{W}, \boldsymbol{\alpha}, \boldsymbol{\beta}) &= \frac{p(\boldsymbol{\Phi}, \boldsymbol{\Theta}, \mathcal{Z}|\mathcal{W}, \boldsymbol{\alpha}, \boldsymbol{\beta})}{p(\mathcal{W}|\boldsymbol{\alpha}, \boldsymbol{\beta})} \\ &\propto p(\mathcal{W}, \mathcal{Z}|\boldsymbol{\Phi}, \boldsymbol{\Theta}, \boldsymbol{\alpha}, \boldsymbol{\beta}) p(\boldsymbol{\Theta}|\boldsymbol{\alpha}) p(\boldsymbol{\Phi}|\boldsymbol{\beta}). \end{aligned} \tag{A.1}$$

The following priors are assumed for model parameters $\boldsymbol{\Phi}$ and $\boldsymbol{\Theta}$:

$$p(\boldsymbol{\Theta}|\boldsymbol{\alpha}) = \prod_{d=1}^{D} p(\boldsymbol{\theta}_d|\boldsymbol{\alpha}) = \prod_{d=1}^{D} Dirichlet(\boldsymbol{\theta}_d; \boldsymbol{\alpha}), \tag{A.2}$$

$$p(\boldsymbol{\Phi}|\boldsymbol{\beta}) = \prod_{k=1}^{K} p(\boldsymbol{\phi}_k|\boldsymbol{\beta}) = \prod_{k=1}^{K} Dirichlet(\boldsymbol{\phi}_k; \boldsymbol{\beta}). \tag{A.3}$$

To derive the joint likelihood function of $\mathcal{W}$ and $\mathcal{Z}$, we first consider the density of data $\mathcal{W}$ given topic assignments $\mathcal{Z}$ and model parameters:

$$p(\mathcal{W}|\mathcal{Z}, \mathbf{\Phi}, \mathbf{\Theta}, \boldsymbol{\alpha}, \boldsymbol{\beta}) = p(\mathcal{W}|\mathcal{Z}, \mathbf{\Phi}) = \prod_{d=1}^{D} \prod_{i=1}^{N_d} p(w_i^{(d)}|z_i^{(d)}, \mathbf{\Phi}). \quad \text{(A.4)}$$

The probability $p(w_i^{(d)}|z_i^{(d)}, \mathbf{\Phi}) = \phi_{z_i^{(d)}, w_i^{(d)}}$ is an element of matrix $\mathbf{\Phi}$ located in $z_i^{(d)}$-th row and $w_i^{(d)}$-th column. The density function of $\mathcal{Z}$ is

$$p(\mathcal{Z}|\mathbf{\Phi}, \mathbf{\Theta}, \boldsymbol{\alpha}, \boldsymbol{\beta}) = p(\mathcal{Z}|\mathbf{\Theta}) = \prod_{d=1}^{D} \prod_{i=1}^{N_d} p(z_i^{(d)}|\boldsymbol{\theta}_d). \quad \text{(A.5)}$$

The probability $p(z_i^{(d)}|\mathbf{\Theta}) = \theta_{d, z_i^{(d)}}$. The joint density of data and latent variable $\mathcal{Z}$ (the complete data likelihood function) is

$$p(\mathcal{W}, \mathcal{Z}|\mathbf{\Phi}, \mathbf{\Theta}, \boldsymbol{\alpha}, \boldsymbol{\beta}) = \prod_{d=1}^{D} \prod_{i=1}^{N_d} p(w_i^{(d)}|z_i^{(d)}, \mathbf{\Phi}) p(z_i^{(d)}|\boldsymbol{\theta}_d). \quad \text{(A.6)}$$

The posterior distribution is proportional to the complete data likelihood function times the prior:

$$p(\mathbf{\Phi}, \mathbf{\Theta}, \mathcal{Z}|\mathcal{W}, \boldsymbol{\alpha}, \boldsymbol{\beta}) \propto \prod_{d=1}^{D} \underbrace{p(\boldsymbol{\theta}_d|\boldsymbol{\alpha})}_{Dirichlet} \prod_{k=1}^{K} \underbrace{p(\boldsymbol{\phi}_k|\boldsymbol{\beta})}_{Dirichlet} \left(\prod_{d=1}^{D} \prod_{i=1}^{N_d} \underbrace{p(w_i^{(d)}|z_i^{(d)}, \mathbf{\Phi})}_{Multinomial} \underbrace{p(z_i^{(d)}|\boldsymbol{\theta}_d)}_{Multinomial} \right). \quad \text{(A.7)}$$

A.2 Choices in Model Specification and Estimation

Implementation of LDA involves important model specification and selection decisions. In particular, the estimation results vary according to the number of topics (K) and hyperparameter settings ($\boldsymbol{\alpha}$, $\boldsymbol{\beta}$). This section discusses the decisions made with respect to both of these dimensions.

A.2.1 The Number of Topics

For the number of topics, there is no "right" answer to this choice; rather, the choice depends on interpretability and goals of the analysis (Grimmer and Stewart 2013; Roberts et al. 2014). DiMaggio, Nag, and Blei (2013) note that "the test of the model as a whole is its ability to identify a number of substantively meaningful and analytically useful topics, not its success in optimizing across all topics."

Typically, in choosing the number of topics, there is a trade-off between predictive accuracy of the model and topic interpretability (Chang et al. 2009). Metrics of predictive performance, such as held-out likelihood or perplexity, are conventionally used to assess model quality (Blei, Ng, and Jordan 2003; Wallach et al. 2009). Perplexity is defined as the inverse of the geometric mean per-word likelihood of the test data.

To evaluate the model fit, one can ask how well the model predicts words in a testing set. Noisy topics will fail to replicate test data, resulting in low held-out likelihood and high perplexity. However, the predictive metrics have limitations. Usually fine-grained, highly specific topics yield the best model fit, but they are not easy to interpret or to generalize (Boyd-Graber, Mimno, and Newman 2014). Furthermore, predicting the content of the preprocessed text is rarely the objective of research in political, economic, or social sciences, especially since the preprocessing steps substantially simplify the original documents (Grimmer and Stewart 2013).

One strand of literature focuses on evaluating topic quality from the perspective of interpretability using automated measures that correlate well with human ratings and thus are better able to serve real-world objectives such as discerning meaningful themes or augmenting the subsequent causal analysis with human-interpretable textual information.

Topics are usually interpreted based on top words with the highest probability (Blei, Ng, and Jordan 2003; Griffiths and Steyvers 2004). Roberts et al. (2014) argue that a semantically interpretable topic has two qualities: (i) it is coherent—the highest probability words for the topic tend to co-occur within documents, and (ii) it is exclusive—the words that have high probability under one topic have low probabilities under other topics. The metrics of coherence and exclusivity that we use for the model selection are described in the paper.

A.2.2 The Concentration Hyperparameters

The concentration hyperparameters determine the amount of smoothing or sparsity of the topic-word and the document-topic distributions. The parameter $\boldsymbol{\alpha}$ informs the model about the concentration of topics within the document. Low α_k (less than 1) means that a document is more focused (i.e., a single topic dominates); high α_k (greater than 1) means that discussion is less focused and several topics occur with similar intensity in the document. Similarly, $\boldsymbol{\beta}$ informs about the concentration of words within a topic. Low beta means a few words are characteristic of the topic. A large $\boldsymbol{\beta}$ implies more uniform topic-word probabilities and leads to similar topics.

Several studies demonstrate that selection of the hyperparameters has a strong influence on both prior and posterior distributions of $\boldsymbol{\Theta}$ and $\boldsymbol{\Phi}$ (Asuncion et al. 2009; Wallach, Mimno, and McCallum 2009; George and Doss 2018). Implementations of LDA typically assume that Dirichlet priors are symmetric ($\beta_1 = \cdots = \beta_V = \beta$ and $\alpha_1 = \cdots = \alpha_K = \alpha$). It is expected that $\beta < 1$ so that many words have low probabilities in a topic.

Following the recommendation of Wallach, Mimno, and McCallum (2009), we implement a combination of priors which is found to be superior: an asymmetric Dirichlet prior over $\boldsymbol{\Theta}$ and a symmetric Dirichlet prior over $\boldsymbol{\Phi}$. First, an asymmetric Dirichlet prior over the document-topic distributions allows some topics to be more likely. These topics may place high probability on words that appear more frequently than other words in every document. Second, asymmetry increases stability of the results as the number of topics increases: if additional topics are redundant, they will be seldom used.

Another decision point is determining the values for hyperparameters. An ad hoc specification of the hyperparameters dominates in the economic literature. Griffiths and Steyvers (2004) provide the most widely applied recommendation: $\alpha = \frac{50}{K}$, $\beta = 0.1$ (Tirunillai and Tellis 2014; Hansen and McMahon 2016; Fligstein, Brundage, and Schultz 2017; Hansen, McMahon, and Prat 2017). This choice is not based on any particular principle.

In this paper we infer the values of concentration parameters in a fully Bayesian setting by placing proper prior distributions on $\boldsymbol{\alpha}$

and $\boldsymbol{\beta}$ (Jacobs, Donkers, and Fok 2016).[25] We use Metropolis-within-Gibbs sampling, which extends upon collapsed Gibbs sampling, for estimation.

A.3 Collapsed Gibbs Sampling

The Gibbs sampler is one technique to produce a sample from the posterior distribution by sequentially drawing samples of a random variable from its conditional distribution given the current values of all other variables. Application of the standard Gibbs algorithm would consider the following sample scheme:

(i) Sample $\boldsymbol{\phi}_k | \boldsymbol{\Phi}_{-k}, \boldsymbol{\Theta}, \mathcal{Z}, \mathcal{W}, \boldsymbol{\alpha}, \boldsymbol{\beta}$ for $k = 1, \ldots, K$

(ii) Sample $\boldsymbol{\theta}_d | \boldsymbol{\Phi}, \boldsymbol{\Theta}_{-d}, \mathcal{Z}, \mathcal{W}, \boldsymbol{\alpha}, \boldsymbol{\beta}$ for $d = 1, \ldots, D$

(iii) Sample $z_i^{(d)} | z_{-i}^{(d)}, \mathcal{Z}^{(-d)}, \boldsymbol{\Theta}, \boldsymbol{\Phi}, \mathcal{W}, \boldsymbol{\alpha}, \boldsymbol{\beta}$ for $d = 1, \ldots, D$; $i = 1, \ldots, N_d$,

where the notation $-d$, $-i$, $-k$ refers to all elements except the d^{th}, i^{th}, and k^{th}, respectively. As iterations continue, the sample values converge to the target posterior distribution in Equation (1) in the paper. The Gibbs sampler is inefficient, because $\boldsymbol{\Theta}$ and $\boldsymbol{\Phi}$ strongly depend on topic assignments $\mathcal{Z}$ and the chain is highly autocorrelated. The classical procedure can be improved upon using the conjugacy of the Dirichlet distribution and the multinomial distribution. Parameters $\boldsymbol{\Theta}$ and $\boldsymbol{\Phi}$ are integrated out from the full conditional posterior distribution for $z_i^{(d)}$. The collapsed Gibbs sampler considers simulating

$$z_i^{(d)} | z_{-i}^{(d)}, \mathcal{Z}^{(-d)}, \mathcal{W}, \boldsymbol{\alpha}, \boldsymbol{\beta} \text{ for } d = 1, \ldots, D; i = 1, \ldots, N_d. \quad \text{(A.8)}$$

To derive the sampling distribution, let $c_{k,d,v} = \sum_{i=1}^{N_d} I(z_i^{(d)} = k, w_i^{(d)} = v)$ denote the number of words of type v assigned to

[25] Another principled way of setting hyperparameters is iterating between Gibbs sampling and a gradient-based optimization for hyperparameters (Wallach 2006) or finding the hyperparameters by grid search (Asuncion et al. 2009).

topic k in document d. In the following, an asterisk means that the corresponding index is summed out, that is

$$c_{k,*,v} = \sum_{d=1}^{D} c_{k,d,v}; \; c_{k,d,*} = \sum_{v=1}^{V} c_{k,d,v}; \; c_{k,*,*} = \sum_{d=1}^{D}\sum_{v=1}^{V} c_{k,d,v}. \quad \text{(A.9)}$$

As $z_i^{(d)}$ takes only K different values, the sampling distribution is multinomial with probabilities (Griffiths and Steyvers 2004):

$$p(z_i^{(d)} | z_{-i}^{(d)}, \mathcal{Z}^{(-d)}, \mathcal{W}, \boldsymbol{\alpha}, \boldsymbol{\beta} \propto \frac{\left(c^{-(d,i)}_{z_i^{(d)},d,*} + \alpha_{z_i^{(d)}}\right)}{\left(\sum_{k=1}^{K} c^{-(d,i)}_{k,d,*} + \alpha_k\right)} \times \frac{\left(c^{-(d,i)}_{z_i^{(d)},*,w_i^{(d)}} + \beta_{w_i^{(d)}}\right)}{\left(\sum_{v=1}^{V} c^{-(d,i)}_{z_i^{(d)},*,v} + \beta_v\right)}, \quad \text{(A.10)}$$

where $c^{-(d,i)}$ denotes a count that does not include word i in document d. See the next subsection for the derivation.

For a single draw we can estimate $\mathbf{\Phi}$, $\mathbf{\Theta}$ from the counts:

$$\theta_{d,k} = \frac{\alpha_k + c_{k,d,*}}{\sum_{k=1}^{K}(\alpha_k + c_{k,d,*})}; \; \phi_{k,v} = \frac{\beta_v + c_{k,*,v}}{\sum_{v=1}^{V}(\beta_v + c_{k,*,v})}. \quad \text{(A.11)}$$

Posterior mean estimates are obtained by averaging over the draws. However, the posterior inference is complicated by a label switching problem (Stephens 2000). The problem emerges because the complete data likelihood (A.6) is invariant to permutations of the topics' labels (there are $K!$ permutations), hence the posterior will inherit the invariance of the likelihood if priors are symmetric. Various relabeling algorithms can be applied to undo label switching before averaging over the draws. Many off-the-shelf solutions provide posterior estimates based on a single iteration of Gibbs sampling. For example, `R` package `lda` (Chang 2015) uses the state at the last iteration of Gibbs sampling and `R` package `topicmodels` (Hornik and Grun 2011) by default returns the sample with the highest posterior likelihood.

A.4 Full Conditional Distribution of Topic Assignments

This section presents the derivation of the full conditional distribution of topic assignments $z_i^{(d)}$ for word i in document d, which is required to construct a Gibbs sampler.

$$
\begin{aligned}
&p(z_i^{(d)}|z_{-i}^{(d)}, \mathcal{Z}^{(-d)}, \mathcal{W}, \boldsymbol{\alpha}, \boldsymbol{\beta}) \propto p(\mathcal{Z}, \mathcal{W}|\boldsymbol{\alpha}, \boldsymbol{\beta}) \\
&\quad = \int\int p(\mathcal{Z}, \mathcal{W}, \boldsymbol{\Theta}, \boldsymbol{\Phi}|\boldsymbol{\alpha}, \boldsymbol{\beta}) d\boldsymbol{\Theta} d\boldsymbol{\Phi} \\
&\quad = \int \prod_{d=1}^{D} \left(p(\boldsymbol{\theta}_d|\boldsymbol{\alpha}) \prod_{i=1}^{N_d} p(z_i^{(d)}|\boldsymbol{\theta}_d) \right) d\boldsymbol{\Theta} \\
&\qquad \times \int \prod_{k=1}^{K} p(\boldsymbol{\phi}_k|\boldsymbol{\beta}) \prod_{d=1}^{D} \prod_{i=1}^{N_d} p(w_i^{(d)}|z_i^{(d)}, \boldsymbol{\Phi}) d\boldsymbol{\Phi} \\
&\quad = \prod_{d=1}^{D} \int p(\boldsymbol{\theta}_d|\boldsymbol{\alpha}) \prod_{i=1}^{N_d} \prod_{k=1}^{K} \theta_{d_k}^{I(z_i^{(d)}=k)} d\boldsymbol{\theta}_d \\
&\quad \times \prod_{k=1}^{K} \int p(\boldsymbol{\phi}_k|\boldsymbol{\beta}) \prod_{d=1}^{D} \prod_{i=1}^{N_d} \prod_{v=1}^{V} \phi_{k,v}^{I(z_i^{(d)}=k, w_i^{(d)}=v)} d\boldsymbol{\phi}_k \\
&\quad \propto \int p(\boldsymbol{\theta}_d|\boldsymbol{\alpha}) \prod_{k=1}^{K} \theta_{d,k}^{c_{k,d,*}} d\boldsymbol{\theta}_d \times \prod_{k=1}^{K} \int p(\boldsymbol{\phi}_k|\boldsymbol{\beta}) \prod_{v=1}^{V} \phi_{k,v}^{c_{k,*,v}} d\boldsymbol{\phi}_k \\
&\quad = \int \frac{\Gamma(\sum_{k=1}^{K} \alpha_k)}{\prod_{k=1}^{K} \Gamma(\alpha_k)} \prod_{k=1}^{K} \theta_{d,k}^{c_{k,d,*}+\alpha_k-1} d\boldsymbol{\theta}_d \\
&\qquad \times \prod_{k=1}^{K} \int \frac{\Gamma(\sum_{v=1}^{V} \beta_v)}{\prod_{v=1}^{V} \Gamma(\beta_v)} \prod_{v=1}^{V} \phi_{k,v}^{c_{k,*,v}+\beta_v-1} d\boldsymbol{\phi}_k \\
&\quad \propto \frac{\prod_{k=1}^{K} \Gamma(c_{k,d,*}+\alpha_k)}{\Gamma(\sum_{k=1}^{K} c_{k,d,*}+\alpha_k)} \underbrace{\int \frac{\Gamma(\sum_{k=1}^{K} c_{k,d,*}+\alpha_k)}{\prod_{k=1}^{K} \Gamma(c_{k,d,*}+\alpha_k)} \prod_{k=1}^{K} \theta_{d,k}^{c_{k,d,*}+\alpha_k-1} d\boldsymbol{\theta}_d}_{=1\text{, integrating over the entire support of Dirichlet}} \\
&\qquad \times \prod_{k=1}^{K} \frac{\prod_{v=1}^{V} \Gamma(c_{k,*,v}+\beta_v)}{\Gamma(\sum_{v=1}^{V} c_{k,*,v}+\beta_v)} \\
&\qquad \times \underbrace{\int \frac{\Gamma(\sum_{v=1}^{V} c_{k,*,v}+\beta_v)}{\prod_{v=1}^{V} \Gamma(c_{k,*,v}+\beta_v)} \prod_{v=1}^{V} \phi_{k,v}^{c_{k,*,v}+\beta_v-1} d\boldsymbol{\phi}_k}_{=1\text{, integrating over the entire support of Dirichlet}}
\end{aligned}
$$

$$= \frac{\prod_{k=1}^{K}\Gamma(c_{k,d,*}+\alpha_k)}{\Gamma(\sum_{k=1}^{K}c_{k,d,*}+\alpha_k)} \times \prod_{k=1}^{K}\frac{\prod_{v=1}^{V}\Gamma(c_{k,*,v}+\beta_v)}{\Gamma(\sum_{v=1}^{V}c_{k,*,v}+\beta_v)}$$
$$\propto \frac{\prod_{k=1}^{K}\Gamma(c_{k,d,*}+\alpha_k)}{\Gamma(\sum_{k=1}^{K}c_{k,d,*}+\alpha_k)} \times \prod_{k=1}^{K}\frac{\Gamma(c_{k,*,w}+\beta_{w_i^{(d)}})}{\Gamma(\sum_{v=1}^{V}c_{k,*,v}+\beta_v)}. \tag{A.12}$$

The next step is to separate terms, which depend on the current term (d,i). This involves splitting the counts into the part that does not count the current position and the part counting only the current position. We also use that $\Gamma(x+1) = x\Gamma(x)$.

$$p(z_i^{(d)}|z_{-i}^{(d)}, \mathcal{Z}^{(-d)}, \mathcal{W}, \boldsymbol{\alpha}, \boldsymbol{\beta})$$
$$\propto \frac{\prod_{k=1;k\neq z_i^{(d)}}^{K}\Gamma\left(c_{k,d,*}^{-(d,i)}+\alpha_k\right)\times\Gamma\left(c_{z_i^{(d)},d,*}^{-(d,i)}+\alpha_{z_i^{(d)}}+1\right)}{\Gamma\left(1+\sum_{k=1}^{K}c_{k,d,*}^{-(d,i)}+\alpha_k\right)}$$
$$\times \prod_{k=1;k\neq z_i^{(d)}}^{K}\frac{\Gamma\left(c_{k,*,w_i^{(d)}}^{-(d,i)}+\beta_{w_i^{(d)}}\right)}{\Gamma\left(\sum_{v=1}^{V}c_{k,*,v}^{-(d,i)}+\beta_v\right)}\times\frac{\Gamma\left(1+c_{z_i^{(d)},*,w_i^{(d)}}^{-(d,i)}+\beta_{w_i^{(d)}}\right)}{\Gamma\left(1+\sum_{v=1}^{V}c_{z_i^{(d)},*,v}^{-(d,i)}+\beta_v\right)}$$
$$\propto \frac{\prod_{k=1}^{K}\Gamma\left(c_{k,d,*}^{-(d,i)}+\alpha_k\right)\times\left(c_{z_i^{(d)},d,*}^{-(d,i)}+\alpha_{z_i^{(d)}}\right)}{\Gamma\left(\sum_{k=1}^{K}c_{k,d,*}^{-(d,i)}+\alpha_k\right)\times\left(\sum_{k=1}^{K}c_{k,d,*}^{-(d,i)}+\alpha_k\right)}$$
$$\times\prod_{k=1}^{K}\frac{\Gamma\left(c_{k,*,w_i^{(d)}}^{-(d,i)}+\beta_{w_i^{(d)}}\right)}{\Gamma\left(\sum_{v=1}^{V}c_{k,*,v}^{-(d,i)}+\beta_v\right)}\times\frac{\left(c_{z_i^{(d)},*,w_i^{(d)}}^{-(d,i)}+\beta_{w_i^{(d)}}\right)}{\left(\sum_{v=1}^{V}c_{z_i^{(d)},*,v}^{-(d,i)}+\beta_v\right)}$$
$$\propto\frac{\left(c_{z_i^{(d)},d,*}^{-(d,i)}+\alpha_{z_i^{(d)}}\right)}{\left(\sum_{k=1}^{K}c_{k,d,*}^{-(d,i)}+\alpha_k\right)}\times\frac{\left(c_{z_i^{(d)},*,w_i^{(d)}}^{-(d,i)}+\beta_{w_i^{(d)}}\right)}{\left(\sum_{v=1}^{V}c_{z_i^{(d)},*,v}^{-(d,i)}+\beta_v\right)}$$
$$\propto\left(c_{z_i^{(d)},d,*}^{-(d,i)}+\alpha_{z_i^{(d)}}\right)\times\frac{\left(c_{z_i^{(d)},*,w_i^{(d)}}^{-(d,i)}+\beta_{w_i^{(d)}}\right)}{\left(\sum_{v=1}^{V}c_{z_i^{(d)},*,v}^{-(d,i)}+\beta_v\right)}. \tag{A.13}$$

A.5 Metropolis-within-Gibbs Sampling

MCMC methods have the advantage of being asymptotically exact, but collapsed Gibbs sampling requires ad hoc hyperparameter specification. The approach adopted in this paper deviates from the common strategies in order to achieve asymptotically exact results and formally infer concentration hyperparameters. The estimation is based on collapsed Gibbs sampling mixed with a Metropolis-Hastings step. In marketing research Jacobs, Donkers, and Fok (2016) implement Metropolis-within-Gibbs sampling to predict purchases with LDA, where a product purchase corresponds to a word and a customer corresponds to a document.

The basic LDA model is extended by adding one more layer to the hierarchical structure where log-normal prior distributions are imposed on the Dirichlet concentration parameters. Based on the considerations in Section 3.2, the Dirichlet prior on the topic-document distributions is asymmetric, whereas the Dirichlet prior on the topic-word distributions is symmetric. The posterior distribution (marginalized over $\mathbf{\Theta}$ and $\mathbf{\Phi}$) is rewritten as

$$p(\mathcal{Z}, \boldsymbol{\alpha}, \beta | \mathcal{W}) \propto \left(\prod_{d=1}^{D} \prod_{i=1}^{N_d} \underbrace{p(w_i^{(d)} | z_i^{(d)}, \beta)}_{Multinomial} \underbrace{p(z_i^{(d)} | \boldsymbol{\alpha})}_{Multinomial} \right) \underbrace{\pi(\beta)}_{Lognormal} \prod_{k=1}^{K} \underbrace{\pi(\alpha_k)}_{Lognormal} . \tag{A.14}$$

The choice of the parameters for the prior distributions is guided by heuristics proposed by Griffiths and Steyvers (2004) for text modeling. The mode of the prior distribution for β is set to 0.1 and the variance is such that 95 percent of the probability mass is under 1. This specification reflects a prior belief that the word-topic distributions are sparse. The mode of the prior distribution for α_k, $k = 1, \ldots, K$, is set to $\frac{50}{K}$ and the variance is chosen such that 95 percent of the probability mass is under $\frac{50}{3}$. This prior specification favors more uniformly distributed document-topic probabilities, although it remains rather uninformative.

A.5.1 Metropolis-Hastings Step

In each sampling step of the Metropolis-within-Gibbs sampling procedure the topic assignments $\mathcal{Z}$ are drawn from the collapsed full posterior distribution (A.10). However, the full conditional distributions of $\boldsymbol{\alpha}$ and β are non-standard and cannot be obtained using the Gibbs sampler.

The full conditional posterior distribution of β is

$$p(\beta|\mathcal{Z}, \mathcal{W}, \boldsymbol{\alpha}) \propto \pi(\beta) \prod_{k=1}^{K} \left(\frac{\Gamma(V\beta)}{\Gamma(V\beta + \sum_{v=1}^{V} c_{k,*,v})} \prod_{v=1}^{V} \frac{\Gamma(\beta + c_{k,*,v})}{\Gamma(\beta)} \right). \tag{A.15}$$

The full conditional posterior distribution of α_k, $k = 1, \ldots, K$ is

$$p(\alpha_k|\mathcal{Z}, \mathcal{W}, \boldsymbol{\alpha}_{-k}, \beta) \propto \pi(\alpha_k) \prod_{d=1}^{D} \left(\frac{\Gamma\left(\sum_{k=1}^{K} \alpha_k\right)}{\Gamma\left(\sum_{k=1}^{K} \alpha_k + c_{k,d,*}\right)} \times \frac{\Gamma\left(\alpha_k + c_{k,d,*}\right)}{\Gamma(\alpha_k)} \right). \tag{A.16}$$

The samples from the conditional posterior distributions (A.16) and (A.15) are obtained using the random-walk Metropolis-Hastings sampler. In general, the sampler makes use of proposal distributions of a known functional form for each Dirichlet concentration parameter. The proposal distributions specify the probability of moving to another "candidate" point in the parameter space in the next iteration, given the sample value in the current iteration. The candidate sample is then accepted or rejected, based on the acceptance probability. Specifically, the random-walk Metropolis-Hastings step is composed of the following parts:

(i) The candidate values β^* are sampled from $\log N(\beta, s_\beta^2)$, where $\beta^{(m)}$ denotes the current value of the parameter and s_β^2 is the variance of the proposal distribution, and the candidate values for α_k^* are sampled from $\log N(\alpha_k^{(m)}, s_{\alpha k}^2), k = 1, \ldots, K$.

(ii) For each univariate Metropolis-Hastings sampler, we compute the log acceptance probability (log transformation is applied to evaluate the gamma functions). For example, for the parameter β:

$$\log \delta = \min(r, 0) \tag{A.17}$$
$$r = \log(p(\beta^*|\mathcal{Z}, \mathcal{W}, \boldsymbol{\alpha})) + \log(q(\beta^{(m)}|\beta^*)$$
$$- \log(p(\beta^{(m)}|\mathcal{Z}, \mathcal{W}, \boldsymbol{\alpha})) - \log(q(\beta^*|\beta^{(m)}),$$

where $q(\beta|\beta^{(m)})$ is a proposal density.

(iii) Set $\beta^{(m+1)} = \beta^*$ with probability δ.
Set $\beta^{(m+1)} = \beta^{(m)}$ with probability $1 - \delta$.

A.5.2 Calibration of the Proposal Distribution

Variances of the proposal distributions are calibrated within the first 500 iterations of the Metropolis-within-Gibbs sampling. The procedure for calibrating the proposal distributions closely follows Jacobs, Donkers, and Fok (2016). The target acceptance rate is 50 percent. The calibration window size is 10. For each calibration window the number of accepted samples (n_A) is stored. If n_A falls outside the 95 percent confidence bounds of the binomial distribution $B(10,\ 0.5)$, then the variance is decreased by $\max\left(\sqrt{\frac{n_A}{10\times 0.5}}, \frac{1}{2}\right)$ or increased by $\min\left(\sqrt{\frac{n_A}{10\times 0.5}}, 2\right)$. See Jacobs, Donkers, and Fok (2016) for details. The initial Metropolis-Hastings standard deviations are $s_\beta = 0.9$, $s_{\alpha_k} = 0.5$.

A.5.3 Posterior Analysis

The estimation is conducted using the whole vocabulary.[26] The sampler runs for 6,000 iterations. Some portion of the initial sample must be discarded as the burn-in period, because the starting values are not sampled from the target posterior distribution. We discard the first 2,000 draws as the burn-in. Every tenth draw is stored. This results in 400 samples from the posterior distribution. We repeat the above procedure for different numbers of latent topics: $K = 3, \ldots, 20$.

Stephens's algorithm (Stephens 2000) is implemented to verify whether label switching has occurred and to perform relabeling if

[26]The multiple starts procedure shows that the chain is not sensitive to the starting values; therefore, in estimation that procedure is omitted.

necessary. The posterior mean estimates of the model parameters are obtained by averaging over the draws.

To achieve a robust evaluation, we compute the measures of semantic coherence and topic exclusivity for different topic cardinalities: $N = 5, 10, 15, 20$, where N denotes the number of words with the highest probability in a topic (Lau and Baldwin 2016). A single score for the model with K components is obtained by averaging the topic-level scores.

A.5.4 Multiple Random Starts

Standard MCMC methods, such as the Metropolis-Hastings algorithm, are known to slowly traverse the support of highly multimodal distributions (Jasra, Holmes, and Stephens 2005). Therefore it is important to investigate the influence of initialization on the solution. In topic modeling, perplexity is a standard measure to examine the model fit and the convergence of the Markov chains initialized from different starting points (Asuncion et al. 2009; Airoldi et al. 2014). Lower perplexity indicates better performance of the model in predicting out-of-sample words. To evaluate perplexity, we split words into a training set (75 percent of words per document) that is used to estimate model parameters and a testing set (25 percent of words per document). Words in the testing set serve to evaluate the generalization capability of the model and therefore are not used in the parameter estimation. Perplexity is defined as the inverse of the geometric mean per-word likelihood of the test data (Hornik and Grun 2011):

$$Perplexity = exp\left(-\frac{\sum_{d=1}^{D}\sum_{v=1}^{V} c_{*,d,v}^{test}\log(\sum_{k=1}^{K}\phi_{k,v}\theta_{d,k})}{\sum_{d=1}^{D} N_d^{test}}\right), \tag{A.18}$$

where $\phi_{k,v}$ and $\theta_{d,k}$ are estimated on the training data.

The sampler is run from five multiple random starts on the training data. For each starting point the sampler runs for 2,000 iterations. In each iteration topic assignments are simulated with a collapsed Gibbs sampling step and concentration parameters for the Dirichlet distributions are simulated with a Metropolis-Hastings step. In each run of the sampler and in each iteration, we use the parameters estimated on the training data to measure the goodness of fit for the test data. In addition, some portion of the initial sample must be discarded as the burn-in period, because the starting values are not sampled from the target posterior distribution. We therefore discard the first 1,000 draws as the burn-in portion.

It is assumed that the MCMC chain has converged to the posterior distribution when the values of perplexity across iterations stabilize.[27]

After convergence, differences in the estimated perplexities for multiple runs turned out to be marginal, indicating that the estimated results are stable across initializations (see Table A.1).

A.5.5 Implementation

We implement the estimation procedure and model diagnostics in `C++` integrated with `R` using application programming interface (API) enclosed in the `Rcpp` package (Eddelbuettel and François 2011). The full code is provided in the replication files.

[27]The absolute percentage change in perplexity score between every tenth iteration is less than 1 percent.

Table A.1. Perplexity Scores for Five Chains with Different Initializations (iterations 1,000–2,000)

Section	Mean					Standard Deviation				
	1	**2**	**3**	**4**	**5**	**1**	**2**	**3**	**4**	**5**
DS	357.87	362.74	363.39	363.65	361.44	2.26	2.32	2.40	2.44	2.29
EA	383.23	380.24	383.34	383.37	380.74	1.54	1.51	1.50	1.44	1.44
MA	293.57	297.42	293.52	293.49	293.72	1.36	1.28	1.31	1.29	1.37
QA	1,039.17	1,039.39	1,041.37	1,039.19	1,038.50	1.30	1.51	1.32	1.27	1.39

Note: Sections (number of topics in parentheses): DS – Decision summary (6); EA – Economic analysis (6); MA – Monetary analysis (4); QA – Q&A (10).

Appendix B. Vocabulary Selection

The analysis presented in the paper required many text preparation steps, such as removing punctuation and numbers, lowercasing, stop word removal, term normalization, or n-gram inclusion. Preparing documents involves preprocessing decisions such as which normalization technique to use or how to reduce vocabulary size, taking into account individual characteristics of our data set. This appendix discusses additional text preprocessing details.

We remove formulaic phrases that are often used to introduce a section of the ECB press conference. These phrases are repeated in many speeches, but have low informational value. The list of removed expressions is provided in Table B.1.

Two term normalization approaches are usually distinguished—stemming and lemmatization. In the paper we opt to use a lemmatizer, although both techniques aim to reduce inflectional and derivational word forms to a common base form.

Stemming refers to applying a set of rules to remove the affixes (for example, it reduces "increasing" to "increas," "stability" to "stabil," "financial" to "financi"). The most widely used methods are algorithmic stemmers (i.e., Porter 2001), which operate without a lexicon and thus ignore word meaning.

Table B.1. List of Expressions Removed from the Corpus of the ECB Press Conferences

"Ladies and gentlemen, the Vice President and I are very pleased to welcome you at the press conference."
"I will now report on the outcome of today's meeting of the Governing Council of the ECB, which was also attended by (. . .)"
"Based on its regular economic and monetary analysis the Governing Council decided"
"Let me now explain our assessment in greater detail, starting with the economic analysis."
"Turning to the monetary analysis"
"We are now at your disposal for questions."

In contrast to algorithmic stemmers, lemmatization requires specifying the part of speech of a word in a sentence in order to reduce the word to its dictionary form (lemma). A lemmatizer transforms all plurals into singular forms and past-tense verbs to present-tense verbs (e.g., "left" to "leave," "developments" to "development," but "stability" and "financial" are unaffected).

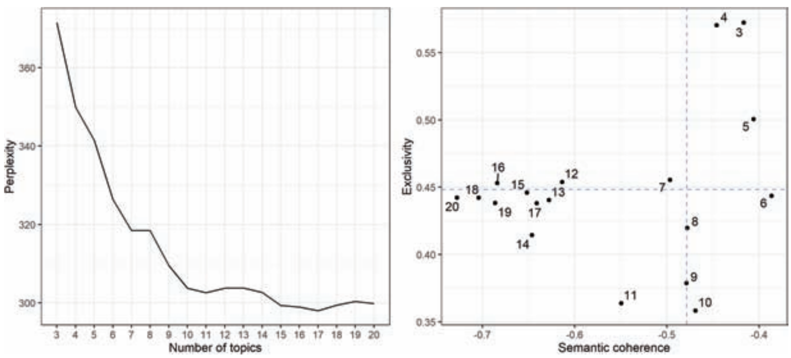
A lemmatizer is more accurate than a stemmer, and it is unlikely to over-conflate (Schofield and Mimno 2016). First, a lemmatizer finds a common form for irregular verbs and nouns ("analyses" — "analysis", "indices" — "index"), which an algorithmic stemmer cannot do. Second, a stemmer may remove too many endings and conflate terms with different meanings. For example, a stemmer (e.g., the Porter 2001 stemmer) would view the following pairs of words as equivalent while lemmatization would not: "import" and "important", "income" and "incoming", "emerging" and "emergency", "future" and "futures", "maturity" and "mature", "consistent" and "consist", "positive" and "position", "accounts" and "accountability".

A lemmatizer increases precision at the expense of recall. In contrast to a stemmer, it is not able to conflate semantically related words belonging to different parts of speech. For example, in the sentence: "With regard to fiscal policies, the Governing Council sees continued reasons for concern", the term "continued" is tagged as an adjective and its lemma is "continued". The Porter stemmer conflates "continue", "continuing", and "continued" to the same stem, "continu".

Appendix C. Model Selection

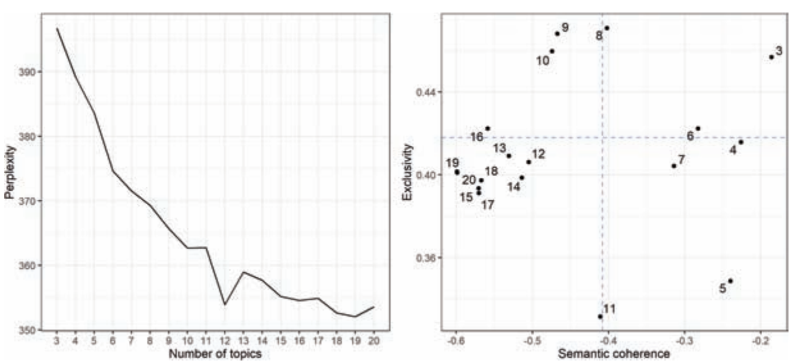
In Figures C.1–C.4 the left panel shows the average perplexity and the right panel presents the average semantic coherence versus the average exclusivity for models with different numbers of topics specified. Lower perplexity indicates better predictive performance of the model, while higher coherence and exclusivity indicate more interpretable topics, on average. The dashed lines mark the 2/3 quantile along each dimension (exclusivity, semantic coherence). We first discard the least performing solutions along the two dimensions separately to remove solutions that have, for example, extremely

Figure C.1. Selection of Number of Topics in the Decision Summary



Note: Selected number of topics is five.

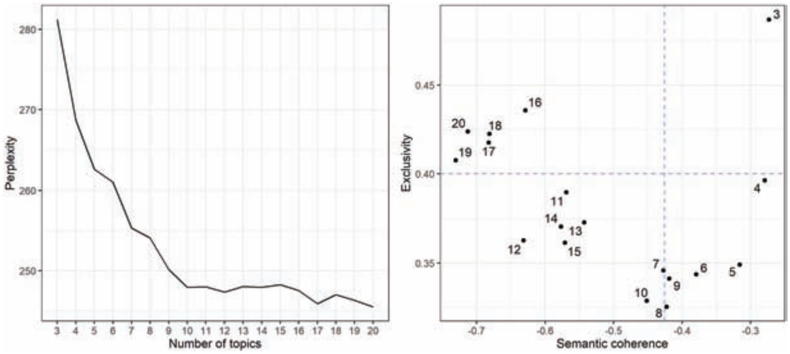
Figure C.2. Selection of Number of Topics in the Economic Analysis Section



Note: Selected number of topics is six.

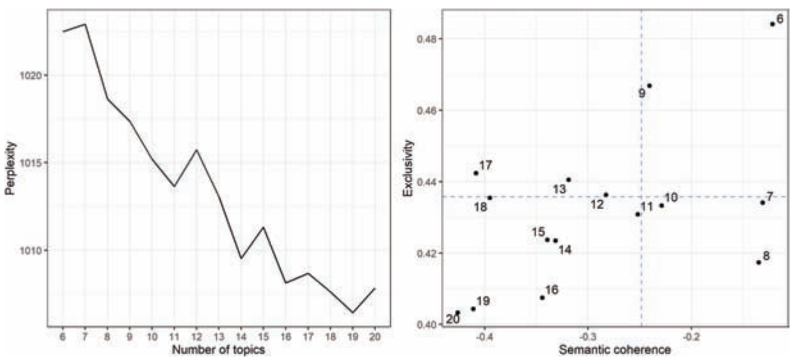
high coherence but very low exclusivity and vice versa. Therefore, we select one of the models located in the top right corner of the graph.

Figure C.3. Selection of Number of Topics in the Monetary Analysis Section



Note: Selected number of topics is four.

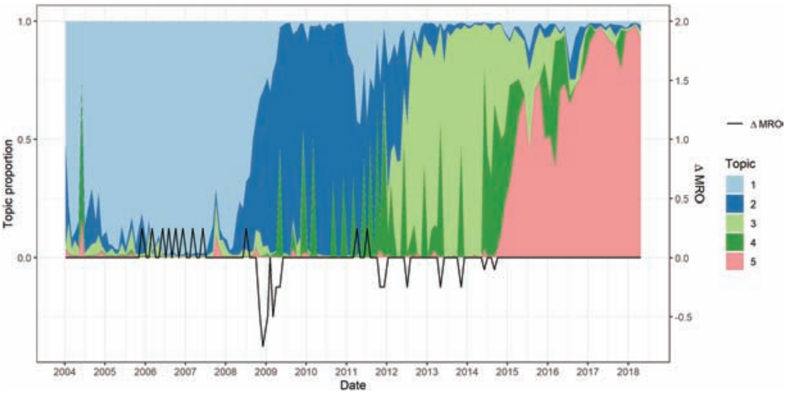
Figure C.4. Selection of Number of Topics in the Q&A



Note: Selected number of topics is nine.

Appendix D. Estimated Topics

Figure D.1. Topics in the Decision Summary over Time



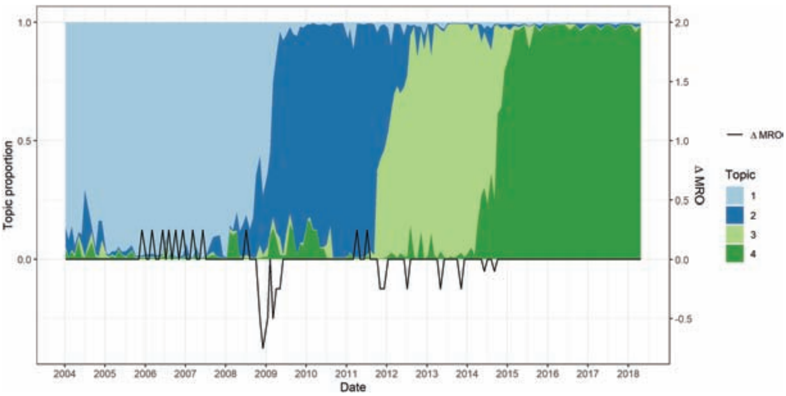
Note: This figure plots the proportion of the decision summary devoted to each topic along with the ECB MRO rate decisions. The topics were estimated using the LDA algorithm (Blei, Ng, and Jordan 2003). The sample comprises 156 transcripts of the section from the ECB press conferences between 2004 and 2018.

Table D.1. Top 10 Terms Describing Topics of the Decision Summary, Ranked by the FREX Score

Topic 1	Topic 2	Topic 3	Topic 4	Topic 5
upside_risk	likely	subdue	conduct	asset_purchase
vigilance	temporary	picture	operation	sustained
prerequisite	purchase_power	prolong	fixed_rate	path
monitor_closely_development	nature	government	procedure	net
solidly	non_standard_measure	weakness	maintenance_period	case
ongoing	construction	extend	tender	monthly
vigorous	inflationary_pressure	impact	full_allotment	beyond
nominal	moderate	dynamic	mros	app
real	allotment	base	three_month	run
timely_manner	acute	financial	least	financial_condition

Note: The FREX score gives high ranks to terms that are both frequent and exclusive

Figure D.2. Topics in the Monetary Analysis Section over Time



Note: This figure plots the proportion of the monetary analysis section devoted to each topic along with the ECB MRO rate decisions. The topics were estimated using the LDA algorithm (Blei, Ng, and Jordan 2003). The sample comprises 156 transcripts of the section from the ECB press conferences between 2004 and 2018.

Table D.2. Top 10 Terms Describing Topics of the Monetary Analysis Section, Ranked by the FREX Score

Topic 1	Topic 2	Topic 3	Topic 4
price_stability	challenge	resilience	recovery
upside_risk	fund	heightened	place
horizon	recapitalisation	country	across
medium_long_term	government	transmission	firm
expansion	full	adjusted	condition
monetary_expansion	advantage	step	annual_rate
price	address	fragmentation	begin
strength	outside	inflow	improvement
trend	different	establish	narrow
influence	steep	adequate	put

Note: The FREX score gives high ranks to terms that are both frequent and exclusive.

Appendix E. Descriptive Statistics

Table E.1. Descriptive Statistics

	Mean	**Std.**	**Min.**	**Max.**
ΔV_t	−0.012	0.066	−0.180	0.249
$\lvert s_t^{MRO}\rvert$	0.009	0.045	0.000	0.250
D_t^A	0.077	0.267	0.000	1.000
r_t^{DE}	−0.003	0.063	−0.204	0.288
$\lvert s_t^{JC}\rvert$	14.144	12.367	0.000	64.000
I_t^{DS}	0.784	0.185	0.399	0.988
I_t^{EA}	0.790	0.170	0.361	0.984
I_t^{MA}	0.889	0.126	0.519	0.993
I_t^{QA}	0.679	0.141	0.419	0.960

Note: This table displays descriptive statistics of variables used in the event-based regressions. ΔV_t denotes the daily percentage change in the VSTOXX index on the conference day t relative to the previous day, $|s_t^{MRO}|$ and $|s_t^{JC}|$ are absolute surprise components of the Main Refinancing Operations (MRO) rate and the U.S. jobless claims, respectively, D_t^A is an indicator for announcements regarding non-standard monetary policy measures, r_t^{DE} is a daily change in German two-year government bond yields and I_t^{DS}, I_t^{EA}, I_t^{MA}, I_t^{QA} denote the index values that capture changes in communication by section: decision summary, economic analysis, monetary analysis, Q&A.

Table E.2. Correlation Matrix

	ΔV_t	I_t^{DS}	I_t^{EA}	I_t^{MA}	I_t^{QA}
ΔV_t	1.000				
I_t^{DS}	0.040	1.000			
I_t^{EA}	−0.140	0.197	1.000		
I_t^{MA}	−0.166	0.420	0.448	1.000	
I_t^{QA}	0.104	0.281	−0.130	0.105	1.000

Note: This table displays Pearson correlation coefficients of our four topic-based communication variables and the VSTOXX index between 2004 and 2018 at a monthly frequency.

Appendix F. Robustness Check

Table F.1. Robustness Check: Different Number of Topics

	Dependent Variable: ΔV_t		
	(1)	**(2)**	**(3)**
$\|s_t^{MRO}\|$	−0.112	−0.144	−0.154
	[0.337]	[0.220]	[0.193]
δV_{t-1}	0.064	0.065	0.053
	[0.558]	[0.558]	[0.623]
r_t^{DE}	−0.159*	−0.162**	−0.146*
	[0.054]	[0.049]	[0.077]
$\|s_t^{JC}\|$	0.0004	0.0004	0.0003
	[0.372]	[0.316]	[0.522]
D_t^A	−0.008	−0.020	−0.004
	[0.691]	[0.320]	[0.858]
I_t^{DS}	0.027	−0.012	0.034
	[0.433]	[0.681]	[0.397]
I_t^{EA}	−0.015	0.007	−0.039
	[0.644]	[0.858]	[0.295]
I_t^{MA}	−0.130***	−0.094**	−0.130***
	[0.010]	[0.029]	[0.006]
I_t^{QA}	0.031	0.040	0.026
	[0.525]	[0.278]	[0.493]
$I_t^{QA} \times D_t^{Draghi}$	−0.053***	−0.051***	−0.042**
	[0.006]	[0.008]	[0.016]
Constant	0.090*	0.058	0.103*
	[0.098]	[0.173]	[0.084]
Observations	156	156	156
Adjusted R^2	0.112	0.103	0.117

Note: Column 1 reports the results where the baseline dimensionality is decreased by 1; column 2 reports the results where the baseline dimensionality is increased by 1; column 3 reports the regression results where the number of topics is selected first by discarding solutions below the 2/3 quantile along dimensions: coherence, exclusivity, and then selecting the model that strictly dominates other models in terms of both coherence and exclusivity; p-values are in brackets. *p < 0.1; **p < 0.05; ***p < 0.01.

References

Airoldi, E. M., and J. M. Bischof. 2016. "Improving and Evaluating Topic Models and Other Models of Text." *Journal of the American Statistical Association* 111 (516): 1381–1403.

Airoldi, E. M., D. Blei, E. A. Erosheva, and S. E. Fienberg. 2014. *Handbook of Mixed Membership Models and Their Applications.* CRC Press.

Altavilla, C., L. Brugnolini, R. S. Gürkaynak, R. Motto, and G. Ragusa. 2019. "Measuring Euro Area Monetary Policy." *Journal of Monetary Economics* 108 (December): 162–79.

Apel, M., and M. Grimaldi. 2012. "The Information Content of Central Bank Minutes." Working Paper No. 261, Sveriges Riksbank.

Asuncion, A., M. Welling, P. Smyth, and Y. W. Teh. 2009. "On Smoothing and Inference for Topic Models." In *Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence*, ed. J. Bilmes and A. Y. Ng, 27–34. AUAI Press.

Bischof, J. M., and E. M. Airoldi. 2012. Summarizing Topical Content with Word Frequency and Exclusivity. In *Proceedings of the 29th International Conference on Machine Learning,* 9–16. Omnipress.

Blei, D. M., A. Y. Ng, and M. I. Jordan. 2003. "Latent Dirichlet Allocation." *Journal of Machine Learning Research* 3 (January): 993–1022.

Blinder, A. S., M. Ehrmann, M. Fratzscher, J. De Haan, and D.-J. Jansen. 2008. "Central Bank Communication and Monetary Policy: A Survey of Theory and Evidence." *Journal of Economic Literature* 46 (4): 910–45.

Bloomberg L.P. 2018. The ECB Main Refinancing Operations Rate (EURR002W:IND), VSTOXX Index (V2TX), German 2-year bond yields (GTDEM2Y:GOV), the U.S. jobless claims (INJCJC:IND) 11/1/03 to 4/30/18. Retrieved July, 2018 from Bloomberg database.

Born, B., M. Ehrmann, and M. Fratzscher. 2014. "Central Bank Communication on Financial Stability." *Economic Journal* 124 (577): 701–34.

Bouma, G. 2009. "Normalized (Pointwise) Mutual Information in Collocation Extraction." In *Proceedings of the Biennial GSCL Conference,* 31–40. Gunter Narr Verlag.

Boyd-Graber, J., D. Mimno, and D. Newman. 2014. "Care and Feeding of Topic Models: Problems, Diagnostics, and Improvements." In *Handbook of Mixed Membership Models and Their Applications,* ed. E. M. Airoldi, D. Blei, E. A. Erosheva, and S. E. Fienberg. CRC Press.

Brand, C., D. Buncic, and J. Turunen. 2010. "The Impact of ECB Monetary Policy Decisions and Communication on the Yield Curve." *Journal of the European Economic Association* 8 (6): 1266–98.

Chang, J. 2015. *LDA: Collapsed Gibbs Sampling Methods for Topic Models.* R Package Version 1.4.2.

Chang, J., J. Boyd-Graber, S. Gerrish, C. Wang, and D. M. Blei. 2009. "Reading Tea Leaves: How Humans Interpret Topic Models." In *Proceedings of the 22nd International Conference on Neural Information Processing Systems,* Vol. 22, 288–96. Curran Associates, Inc.

Coenen, G., M. Ehrmann, G. Gaballo, P. Hoffmann, A. Nakov, S. Nardelli, E. Persson, and G. Strasser. 2017. "Communication of Monetary Policy in Unconventional Times." ECB Working Paper No. 2080.

Collobert, R., J. Weston, L. Bottou, M. Karlen, K. Kavukcuoglu, and P. Kuksa. 2011. "Natural Language Processing (Almost) from Scratch." *Journal of Machine Learning Research* 12 (August): 2493–2537.

Constâncio, V. 2018. "Past and Future of the ECB Monetary Policy." Speech at the "Conference on Central Banks in Historical Perspective: What Changed After the Financial Crisis?" organized by the Central Bank of Malta, Valletta, May 4. Retrieved from https://www.ecb.europa.eu/press/key/date/2018/html/ecb.sp180504.en.html.

Denny, M. J., and A. Spirling. 2018. "Text Preprocessing for Unsupervised Learning: Why It Matters, When It Misleads, and What to Do about It." *Political Analysis* 26 (2): 168–89.

Diessner, S., and G. Lisi. 2019. "Masters of the 'Masters of the Universe'? Monetary, Fiscal and Financial Dominance in the Eurozone." *Socio-Economic Review* 18 (2): 315–35.

DiMaggio, P., M. Nag, and D. Blei. 2013. “Exploiting Affinities between Topic Modeling and the Sociological Perspective on Culture: Application to Newspaper Coverage of US Government Arts Funding.” *Poetics* 41 (6): 570–606.

Dybowski, T. P., and B. Kempa. 2019. “The ECB’s Monetary Pillar after the Financial Crisis.” Technical Report, Center for Quantitative Economics (CQE), University of Muenster.

Eddelbuettel, D., and R. François. 2011. “Rcpp: Seamless R and C++ Integration.” *Journal of Statistical Software* 40 (8): 1–18.

Ehrmann, M., and M. Fratzscher. 2009. “Explaining Monetary Policy in Press Conferences.” *International Journal of Central Banking* 5 (2): 42–84.

Ehrmann, M., and J. Talmi. 2020. “Starting from a Blank Page? Semantic Similarity in Central Bank Communication and Market Volatility.” *Journal of Monetary Economics* 111 (May): 48–62.

European Central Bank. 2003. “The ECB’s Monetary Policy Strategy.” Press release (May 8). Retrieved from https://www.ecb.europa.eu/press/pr/date/2003/html/pr030508_2.en.html.

———. 2009. “The Current Euro Area Recovery from a Historical Perspective.” Monthly Bulletin (February).

Fligstein, N., J. S. Brundage, and M. Schultz. 2017. “Seeing Like the Fed: Culture, Cognition, and Framing in the Failure to Anticipate the Financial Crisis of 2008.” *American Sociological Review* 82 (5): 879–909.

Fratzscher, M., M. Lo Duca, and R. Straub. 2016. “ECB Unconventional Monetary Policy: Market Impact and International Spillovers.” *IMF Economic Review* 64 (1): 36–74.

George, C. P., and H. Doss. 2018. “Principled Selection of Hyperparameters in the Latent Dirichlet Allocation Model.” *Journal of Machine Learning Research* 18 (162): 1–38.

Griffiths, T. L., and M. Steyvers. 2004. “Finding Scientific Topics.” *Proceedings of the National Academy of Sciences* 101 (suppl 1): 5228–35.

Grimmer, J., and B. M. Stewart. 2013. “Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts.” *Political Analysis* 21 (3): 267–97.

Gürkaynak, R. S., B. Sack, and E. Swanson. 2005. “Do Actions Speak Louder Than Words? The Response of Asset Prices to

Monetary Policy Actions and Statements." *International Journal of Central Banking* 1 (1, May): 55–94.

Hansen, S., and M. McMahon. 2016. "Shocking Language: Understanding the Macroeconomic Effects of Central Bank Communication." *Journal of International Economics* 99 (Supplement 1): S114–S133.

Hansen, S., M. McMahon, and A. Prat. 2017. "Transparency and Deliberation within the FOMC: A Computational Linguistics Approach." *Quarterly Journal of Economics* 133 (2): 801–70.

Hornik, K., and B. Grun. 2011. "Topicmodels: An R Package for Fitting Topic Models." *Journal of Statistical Software* 40 (13): 1–30.

Jacobs, B. J., B. Donkers, and D. Fok. 2016. "Model-Based Purchase Predictions for Large Assortments." *Marketing Science* 35 (3): 389–404.

Jansen, D.-J., and J. De Haan. 2005. "Talking Heads: The Effects of ECB Statements on the Euro–Dollar Exchange Rate." *Journal of International Money and Finance* 24 (2): 343–61.

———. 2007. "The Importance of Being Vigilant: Has ECB Communication Influenced Euro Area Inflation Expectations?" CESifo Working Paper No. 2134.

Jasra, A., C. C. Holmes, and D. A. Stephens. 2005. "Markov Chain Monte Carlo Methods and the Label Switching Problem in Bayesian Mixture Modeling." *Statistical Science* 1 (February): 50–67.

Jegadeesh, N., and D. Wu. 2013. "Word Power: A New Approach for Content Analysis." *Journal of Financial Economics* 110 (3): 712–29.

———. 2017. "Deciphering Fedspeak: The Information Content of FOMC Meetings." Working Paper.

Lau, J. H., and T. Baldwin. 2016. "The Sensitivity of Topic Coherence Evaluation to Topic Cardinality." In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies,* 483–87. Association for Computational Linguistics.

Loughran, T., and B. McDonald. 2011. "When is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks." *Journal of Finance* 66 (1): 35–65.

Lucca, D. O., and F. Trebbi. 2009. "Measuring Central Bank Communication: An Automated Approach with Application to FOMC Statements." NBER Working Paper No. 15367.

Mimno, D., H. Wallach, E. Talley, M. Leenders, and A. McCallum. 2011. "Optimizing Semantic Coherence in Topic Models." In *Proceedings of the Conference on Empirical Methods in Natural Language Processing,* 262–72. Association for Computational Linguistics.

Newman, D., J. H. Lau, K. Grieser, and T. Baldwin. 2010. "Automatic Evaluation of Topic Coherence." In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics,* 100–108. Association for Computational Linguistics.

Peek, J., E. S. Rosengren, and G. Tootell. 2016. "Does Fed Policy Reveal a Ternary Mandate?" Working Paper No. 16-11, Federal Reserve Bank of Boston.

Picault, M., and T. Renault. 2017. "Words Are Not All Created Equal: A New Measure of ECB Communication." *Journal of International Money and Finance* 79 (December): 136–56.

Porter, M. F. 2001. "Snowball: A Language for Stemming Algorithms." Retrieved from http://snowball.tartarus.org/texts/introduction.html (accessed April 2018).

Quinn, K. M., B. L. Monroe, M. Colaresi, M. H. Crespin, and D. R. Radev. 2010. "How to Analyze Political Attention with Minimal Assumptions and Costs." *American Journal of Political Science* 54 (1): 209–28.

Roberts, M. E., B. M. Stewart, D. Tingley, C. Lucas, J. Leder-Luis, S. K. Gadarian, B. Albertson, and D. G. Rand. 2014. "Structural Topic Models for Open-Ended Survey Responses." *American Journal of Political Science* 58 (4): 1064–82.

Rosa, C., and G. Verga. 2007. "On the Consistency and Effectiveness of Central Bank Communication: Evidence from the ECB." *European Journal of Political Economy* 23 (1): 146–75.

Schofield, A., and D. Mimno. 2016. "Comparing Apples to Apple: The Effects of Stemmers on Topic Models." *Transactions of the Association for Computational Linguistics* 4: 287–300.

Shapiro, A. H., and D. Wilson. 2021. "Taking the Fed at Its Word: A New Approach to Estimating Central Bank Objectives Using

Text Analysis." Working Paper No. 2019-02, Federal Reserve Bank of San Francisco.

Stark, J. 2009. "Monetary Policy Before, During and After the Financial Crisis." Speech at University Tübingen, Tübingen, November 9. Retrieved from https://www.ecb.europa.eu/press/key/date/2009/html/sp091109.en.html.

Stephens, M. 2000. "Dealing with Label Switching in Mixture Models." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 62 (4): 795–809.

Tetlock, P. C. 2007. "Giving Content to Investor Sentiment: The Role of Media in the Stock Market." *Journal of Finance* 62 (3): 1139–68.

Tetlock, P. C., M. Saar-Tsechansky, and S. Macskassy. 2008. "More than Words: Quantifying Language to Measure Firms' Fundamentals." *Journal of Finance* 63 (3): 1437–67.

Tirunillai, S., and G. J. Tellis. 2014. "Mining Marketing Meaning from Online Chatter: Strategic Brand Analysis of Big Data Using Latent Dirichlet Allocation." *Journal of Marketing Research* 51 (4): 463–79.

van Dieijen, M., and R. L. Lumsdaine. 2018. "What Say They About Their Mandate? A Textual Assessment of Federal Reserve Speeches." Unpublished Working Paper, Erasmus University Rotterdam.

Vo, D.-V. 2019. "The Response of Exchange Rate to Central Bank Communication: A Textual Analysis Approach." Unpublished Working Paper. Retrieved from https://ssrn.com/abstract=3406603 or http://dx.doi.org/10.2139/ssrn.3406603.

Wallach, H. M. 2006. "Topic Modeling: Beyond Bag-of-Words." In *Proceedings of the 23rd International Conference on Machine Learning,* 977–84. Association for Computing Machinery.

Wallach, H. M., D. Mimno, and A. McCallum. 2009. "Rethinking LDA: Why Priors Matter." In *Proceedings of the 22nd International Conference on Neural Information Processing Systems,* 1973–81. Curran Associates Inc.

Wallach, H. M., I. Murray, R. Salakhutdinov, and D. Mimno. 2009. "Evaluation Methods for Topic Models." In *Proceedings of the 26th Annual International Conference on Machine Learning,* 1105–12. Association for Computing Machinery.