Australian National University
Crawford School of Public Policy

CAMA

Centre for Applied Macroeconomic Analysis

What Do Central Bankers Talk About? Evidence From the BIS Archive

CAMA Working Paper 35/2025
June 2025

Martin Feldkircher
Vienna School of International Studies
Centre for Applied Macroeconomic Analysis, ANU

Petr Koráb
Slovak Academy of Sciences
Text Mining Stories

Viktoriya Teliha
Vienna School of International Studies

Abstract

This paper analyzes the evolution of central bank topics using a corpus of over 20,000 speeches spanning nearly three decades and a range of topic models. We identify thirteen themes, including monetary policy, financial stability, digital payments, and climate-related finance. Examining their development over time, we classify these themes as “evergreens”, “waning threads”, or emergent “rising stars”, and show that early adoption and topic leadership are nearly equally shared between emerging and advanced economies’ central banks. In the aftermath of the Global Financial Crisis, topic focus converged worldwide, with a renewed emphasis on financial stability. Finally, static covariate regressions link topic prevalence to inflation regimes, central bank independence, and speech format, highlighting the impact of macroeconomic and institutional factors on communication priorities.

Keywords

monetary policy, financial stability, digital payments, climate-related finance

JEL Classification

C55, C88, E52, E58, D83

Address for correspondence:

(E) cama.admin@anu.edu.au

ISSN 2206-0332

What do central bankers talk about? Evidence from the BIS archive

MARTIN FELDKIRCHER[1,2], PETR KORÁB[3,4], and VIKTORIYA TELIHA*[1]

[1]Vienna School of International Studies (DA)
[2]Centre for Applied Macroeconomic Analysis (CAMA)
[3]Slovak Academy of Sciences, Institute of Economic Research (SAS)
[4]Text Mining Stories (TMS)

June 25, 2025

Abstract

This paper analyzes the evolution of central bank topics using a corpus of over 20,000 speeches spanning nearly three decades and a range of topic models. We identify 13 themes, including monetary policy, financial stability, digital payments, and climate-related finance. Examining their development over time, we classify these themes as "evergreens", "waning threads", or emergent "rising stars", and show that early adoption and topic leadership are nearly equally shared between emerging and advanced economies' central banks. In the aftermath of the Global Financial Crisis, topic focus converged worldwide, with a renewed emphasis on financial stability. Finally, static covariate regressions link topic prevalence to inflation regimes, central bank independence, and speech format, highlighting the impact of macroeconomic and institutional factors on communication priorities.

JEL Codes: C55, C88, E52, E58, D83

*Martin Feldkircher `martin.feldkircher@da-vienna.ac.at` (corresponding author), Petr Koráb `petr.korab@savba.sk` and Viktoriya Teliha, `viktoriya.teliha@da-vienna.at`. We would like to thank Xiaobao Wu, Florian Horký and participants of the DA's brown bag seminar for helpful comments.

1 Introduction and Literature Review

Central bank speeches have become a pivotal channel for communicating monetary policy perspectives, especially in an environment of prolonged near-zero interest rates and unconventional policy tools across advanced economies. Unlike policy statements or meeting minutes, speeches allow for greater narrative flexibility, enabling policymakers to articulate evolving policy themes, long-term concerns, and institutional priorities in their word (Woodford, 2005). Building on this narrative potential, text-based analysis of central bank communication has emerged as a rich field, offering insights into how institutions adapt their messaging across economic cycles, policy innovations, and geopolitical shifts.

Central bank communication plays a crucial role not only in guiding market expectations but also in reinforcing central bank accountability and institutional credibility (Blinder et al., 2008; Bholat et al., 2015; Haldane and McMahon, 2018). In recent years, the scope of central bank communication has broadened significantly, extending beyond traditional monetary policy to encompass themes such as financial stability, climate change, digital currencies, and public trust. As central banks face rising public scrutiny and increasingly complex mandates, the demand for more structured and transparent communication strategies has grown. However, translating policy discourse into actionable insights remains challenging due to the technical complexity of central bank language, its multi-layered audiences, and the institutional context in which it is embedded (Silva et al., 2025).

Against this backdrop, natural language processing (NLP) and machine learning methods — particularly transformer-based models and large language models (LLMs)—offer powerful tools to systematically analyze central bank texts. Recent studies have used such models to uncover semantic shifts in policy narratives, track thematic focus, and assess sentiment (Gambacorta et al., 2024; de Araujo et al., 2025; Hansen and Kazinnik, 2024). A recent IMF study proposes a multi-dimensional LLM-based framework to assess communication across multiple dimensions, further underscoring the potential of these methods (Silva et al., 2025).

This paper contributes to the growing literature by presenting the first systematic comparison of classical bag-of-words and modern contextualized topic modeling approaches applied to long-form central bank speeches. It is also the first study to examine nearly three decades of global central bank communication using a unified corpus drawn from more than 20,150 speeches and presentations available via the Bank for International Settlements (BIS) archive. By implementing and comparing three distinct modeling approaches—keyword-assisted LDA (Eshima et al., 2024), BERTopic (Grootendorst, 2022), and FASTopic (Wu et al., 2024) — we address persistent challenges in modeling long, formal, and thematically dense policy texts.

Our analysis bridges two complementary strands of research. The first applies unsupervised and hybrid topic modeling to central bank communication. Latent Dirichlet Allocation (LDA; Blei et al., 2003) and Structural Topic Models (STM; Roberts et al., 2014) remain widely used tools to extract latent themes and track thematic evolution. Feldkircher et al. (2021) identify

dominant themes in European central bank addresses, while Bohl et al. (2023) show how institutional mandates shape discourse at the ECB and Federal Reserve. More targeted applications have emerged as well — Feldkircher and Teliha (2024) trace climate-related narratives using keyword-assisted topic models, revealing that peer effects and government activism, rather than formal sustainability mandates, drive green communication across 112 central banks. Ahrens and McMahon (2021) use supervised models to derive monetary policy shocks from Federal Reserve speeches, while Park et al. (2023) show that unsupervised keyword scoring can detect early warning signs of financial instability. Notably, Kumar et al. (2024) represent the only known application of BERT embeddings for topic modeling in central bank communication to date.

The second research strand focuses on sentiment, tone, and novelty in central bank texts. Sentiment analysis, using both dictionary- and embedding-based methods, has become a powerful lens for interpreting central bank tone. Armelius et al. (2020) find that sentiment co-moves across central banks, with distance and Fed influence playing key roles. Recent work has shifted toward contextualized embeddings: Niţoi et al. (2023) manually annotate monetary-policy-relevant sentences and use a pre-trained BERT model to estimate sentiment, while Kim et al. (2025) combine discourse simplification with FinBERT (Araci, 2019) to analyze FOMC sentiment. Gambacorta et al. (2024) show that domain-trained BERT models improve prediction tasks on central bank data. Baerg and Binder (2024), in turn, caution that model performance and interpretability depend critically on research design, particularly when applying large language models (LLMs) to central bank communication. The Silva et al. (2025)'s recent classifier represents a significant step forward by providing a unified architecture for analyzing four dimensions of communication—topics, sentiment, stance, and audience—on a multilingual corpus across central banks, helping institutions benchmark their communication strategies and improve transparency.

Despite these advances, no study has yet compared classical bag-of-words models (like LDA or STM) with transformer-based models (like BERTopic or FASTopic) on long-form central bank speeches. These texts pose unique challenges due to their length, formal structure, and thematic complexity.

To fill this gap, we analyze over 20,000 speeches and presentations from the BIS archive. We implement and compare three models: keyword-assisted LDA (which uses domain-specific seed terms for topic discovery), BERTopic (which clusters BERT embeddings and uses word frequency for topic assignment), and FASTopic (which integrates BERT embeddings within an LDA-style framework). Each model is evaluated in a range of topic counts using semantic coherence (Mimno et al., 2011), normalized pointwise mutual information (Bouma, 2009), and topic diversity metrics. Preliminary results suggest that keyword-assisted LDA with 15 topics achieves the most favorable trade-off between coherence and diversity, outperforming both purely unsupervised and embedding-based models in terms of interpretability and thematic coverage.

Our main findings are threefold. First, we offer the first head-to-head comparison of classical bag-of-words and transformer-based topic models applied to long-form central bank speeches. Our results demonstrate that keyword-assisted models generate more coherent and interpretable themes than BERT-based alternatives, which struggle to decompose complex documents into multiple co-occurring topics. Second, we propose a thematic-evolution framework that identifies monetary policy and financial stability as persistent "evergreens" in central bank discourse, while revealing a balanced innovation landscape where emerging economy central banks pioneer approximately half of all policy topics, though advanced economies maintain advantages in sustained topic leadership. We document systematic attention synchronization triggered by global crises—particularly during the 2008–2012 Global Financial Crisis—when all central banks simultaneously elevated emphasis on financial stability discourse. Third, we show that communication strategies vary systematically across macroeconomic environments and institutional contexts: inflationary periods prompt increased financial literacy efforts, while more independent central banks engage more actively with politically sensitive topics, devoting relatively less attention to global coordination.

The remainder of the paper proceeds as follows. Section 2 describes the BIS speech corpus and our model methodology. Section 3 describes the topic model frameworks and evaluation metrics. Section 4 presents the main results based on our preferred model and section 5 concludes.

2 Data

We build on the data employed in Feldkircher and Teliha (2024) and use texts from the BIS archive (`www.bis.org/cbspeeches/index.htm`). The archive contains public speeches from central bankers and monetary authorities translated into English by expert staff. Only speeches that have been actively submitted by the authorities are contained in the archive. This implies that these speeches probably reflect only a subset of the full set of speeches given by central banks – a fact that has been recently led to the development of a new data base introduced in Campiglio et al. (2025) that on top of BIS archived speeches uses web-scrapped speeches from central bank websites and translates non-English speeches automatically using Google translate. While being more extensive in the sense that more central banks are covered, in this paper, we preferred to have the most up-to-date set of speeches in order to analyze topic evolution over time. Hence our data set contains all speeches up to 2025, whereas the one of Campiglio et al. (2025) ends in 2023.

The main body of our data has been obtained by downloading the PDFs from the archive. Since 2024, the BIS makes the speeches also available in Excel format and we concatenated the latest update on the set of our directly obtained speeches. From the speeches, we disregard non-English and duplicated texts.[1] We also discarded speeches from institutions other than central banks and dropped plain reports that have been submitted to the archive at the beginning of

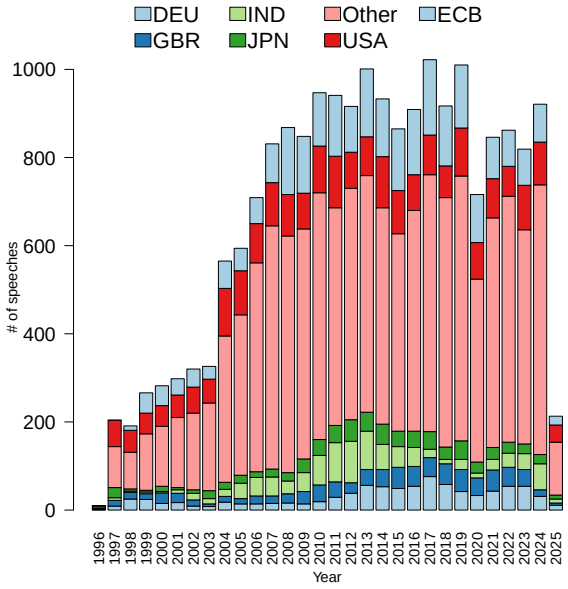
[1] All texts are in PDF format; from these, around 140 PDFs have been non-readable and had to be converted into plain text in order to be able to further process them.

the sample, mostly by the Bank of Japan.

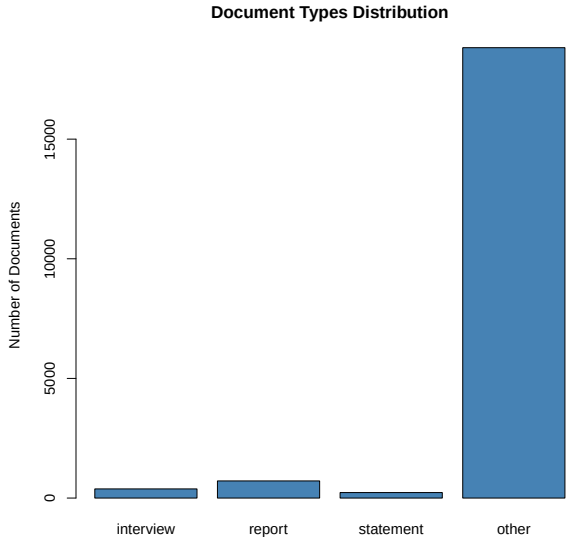
Our final corpus then consists of $D = 20,150$ documents, with the earliest dating back to 1997-01-06 and the latest being from 2025-04-28. The number of speeches are displayed in Figure 1.

Figure 1: Descriptive statistics of the BIS speech corpus

(a) Number of speeches per year



(b) Distribution of speech types



Notes: The left panel shows the temporal distribution of speeches from 1997-2025, with a notable increase starting in 2004 and a temporary decline in 2020 during the COVID-19 pandemic. The right panel categorizes speeches by type based on keyword analysis of titles. Speech types were identified using keywords such as "interview," "statement," and "presentation" to classify the communication format. The majority of documents fall into the general speech category, indicating diverse communication formats beyond formal statements and interviews.

The figure demonstrates a rising trend in the number of speeches. There has been a sharp increase in 2004 and a drop in speeches in 2020, the COVID period. The top contributors of speeches in our data set comprise the US Fed, the ECB, the German Bundesbank, the Bank of England, the Bank of Japan and the Indian Reserve bank. The documents we analyze differ significantly in length. The minimum length before any pre-processings is 47 tokens and the maximum length about 58,000, the average length of a document is about 3,000 tokens (amounting to about five to 8 pages). The figure also shows the type of speech. These were obtained by simply analyzing the speeches' titles for certain keywords such as "interview" and "questions" for an interview or "report", "review" for a presentation or discussion of a report and "statement" for introductory statements.[2] The figure shows that the data set contains mostly speeches that cannot be categorized easily by this keyword approach. Nevertheless, it is important to note that some speeches are presentations of reports, others explanations of

[2]The keywords used for interviews are "interview"; "q&a"; "questions"; "conversation with"; "dialogue"; "discusses with", for reports "statement"; "press release"; "announcement"; "declaration"; "communique"; "proclamation", for speeches "speech"; "discusses"; "considers"; "addresses"; "talks"; "gives his"; "gives her"; "looks at"; "remarks"; "lecture"; "presentation"; "briefing"; "address to"; "keynote"; "speaking"; "speaks on".

monetary policy decisions (statements), and others interviews.

For using topic models, preprocessing the text data is a crucial step, which also differs across the model classes. For the *keyATM* model, we assemble an extensive stopword list by merging a canonical English stopword set with domain-specific terms[3], thereby ensuring that frequently occurring but semantically empty words are removed. We then apply a compound-token identification routine to recognize and retain multi-word expressions such as country and city names, currencies, speech author names, and other fixed phrases, ensuring that these units are treated as single tokens. Following this, all text is subjected to a normalization sequence comprising the removal of punctuation, symbols, numerals, and URLs; retention of alphabetic tokens only, with a minimum length of three characters; and conversion of all tokens to lowercase. Finally, we perform lemmatization, which maps each word form to its base or dictionary form (e.g., "banks" and "bank" both become "bank"), thereby consolidating inflected variants into a single lexical unit.

The topic models using embeddings use a different pre-processing pipeline. For embedding-based topic modeling, we adopt a modified preprocessing strategy in order to preserve the contextual information required by transformer encoders. In particular, BERTopic is trained on the original, uncleaned text: as noted in the BERTopic documentation, "removing stop words as a preprocessing step is not advised, as the transformer-based embedding models that we use need the full context to create accurate embeddings." Accordingly, all standard cleaning operations—compound-token handling, punctuation and symbol removal, numeric and URL stripping, lowercase normalization, and lemmatization—are performed within BERTopic's custom preprocessing pipeline during model training rather than as a separate, preliminary step. This integrated approach ensures that the embedding model receives complete input context while still benefiting from our established normalization procedures. Following the recommendations of Wu et al. (2024), for FASTopic, the data is structurally cleaned as part of preprocessing, so the model receives a clean corpus as input. We keep the identical sequence of cleaning operations for all three models to minimize potential bias from text preprocessing.

Finally, we transform the cleaned and lemmatized text into bigrams. The main reason for that is that central bank speeches are replete with domain-specific collocations (e.g., "policy rate", "output gap") whose latent semantics cannot be recovered by isolated tokens. Incorporating bigram features directly into the modeling process preserves these multi-word expressions and yields topics that more closely align with the nuanced language of monetary policy (for recent monetary policy applications, see Feldkircher and Teliha, 2024; Feldkircher et al., 2024). Lau et al. (2014) empirically demonstrate that phrase-based coherence metrics substantially outperform unigram baselines in capturing human judgments of topic quality. Furthermore, Blei and Lafferty (2009) and their successors extend classical topic models by integrating phrase de-

[3]These comprise terms like "bis", "speeches", "online", "conference", "meeting", "notes", "remarks", "address", "thank", "thanks", "ladies", "gentlemen", "introduction", "report", "nationalbank"; textual representation of numbers, months, and years (e.g., "one", "two", "october", "feb", "annual", "percent"); academic degrees (e.g., "prof", "professor"); and their lexical and lowercased variants (e.g., "Annualy" → "annual").

tection, showing that multi-word units enhance the semantic richness of inferred topics. Early work by Wallach (2006) similarly argues that n-gram features—particularly bigrams—facilitate the interpretability of topics by anchoring them to meaningful lexical patterns rather than isolated terms. In our own trials, post hoc phrase augmentation in a unigram model failed to produce comparably coherent or interpretable topics, reinforcing the necessity of native bigram incorporation for analyzing central bank discourse.

3 Methodology & Evaluation

We use three topic models that fundamentally differ in how they model text. The first one is the keyword assisted topic model *keyATM* of Eshima et al. (2024), which belongs to the class of mixture topic models, going back to Blei and Lafferty (2007). The main idea of mixture topic models is that each document is a mixture of different topics and each *topic* is a distribution over words or tokens. The *keyATM* model extends this classical Latent Dirichlet Allocation (LDA) framework by incorporating prior information through user-specified *keywords*. This prior information helps guide the model toward interpretable and meaningful topics, particularly when the data are sparse or when certain topics are known to be of interest.

More specifically, in the baseline *keyATM*, the main idea is that for each token (in our case bigram) j in document d, a latent topic assignment variable is drawn according to:

$$z_{dj} \sim \text{Categorical}(\theta_d),$$

where θ_d is a K-dimensional vector of topic probabilities for document d with $\sum_{k=1}^{K} \theta_{dk} = 1$. Given the topic assignment z_{dj}, tokens in the document are generated from topic-specific token distributions:

$$w_{dj} \mid z_{dj} = k \sim \text{Categorical}(\phi_k),$$

where ϕ_k is a V-dimensional vector representing the probability distribution over the vocabulary for topic k.

A distinctive feature of *keyATM* is the treatment of *keyword-assisted* topics. Specifically, the model partitions the K topics into $\tilde{K} \leq K$ *keyword topics* and $K - \tilde{K}$ *non-keyword topics*. For the keyword topics ($k \leq \tilde{K}$), token generation follows a two-step process. First, a binary latent variable s_{dj} is drawn according to:

$$w_{dj} \mid z_{dj} = k \sim \begin{cases} \text{Categorical}(\phi_k), & \text{if } s_{dj} = 0, \\ \text{Categorical}(\tilde{\phi}_k), & \text{if } s_{dj} = 1, \end{cases}$$

where $s_{dj} \sim \text{Bernoulli}(\pi_k)$ with success probability π_k. In case $s_{dj} = 1$, the vocabulary that is used for token-distribution relates to the probability vector of the provided keywords, $\tilde{\phi}_k$.

π_k is hence a very influential parameter, since forcing it to 1 would imply that only keywords are allowed to generate the tokens and documents, which would be very restrictive. Following Eshima et al. (2024) we hence put a uniform, non-informative prior on $\pi_k \sim \text{Gamma}(\gamma_1, \gamma_2)$, with $\gamma_1 = \gamma_2 = 1$ for all k.

To assess topic evolution, we use the time-varying modification of this baseline set-up which introduces time by modeling the time periods as distinct, latent states $r = 1, \ldots, R$. Time variation is then introduced by making the parameter α that governs topic-document distributions, θ_d, state dependent, $\alpha_{rk} \sim \text{Gamma}(\eta_1, \eta_2)$ for $r = 1, 2, \ldots, R$ and $k = 1, 2, \ldots, K$. This approach allows us to estimate time-varying topic proportions per central bank, which we will use to analyze topic leaders and spillovers between central banks. Our approach to derive keywords and the list of all keywords is presented in the appendix, A.

The second topic model we consider is BERTopic from Grootendorst (2022). The model fundamentally differs from *keyATM* in at least two aspects. First, it uses the pre-trained BERT model of Devlin et al. (2018). Using BERT allows to get semantic meaning of words and phrases in a context-dependent way on a sentence level for each document, which are then aggregated to the document level. The second step involves a dimension reduction technique (UMAP) on the obtained document embeddings. Finally, using clustering algorithms (e.g., HDBSCAN) documents with similar document embeddings are assigned to the same cluster. This approach has two caveats. First, documents are not assumed to be comprised of different topics, each document is assigned to one and only one cluster. Second, the clusters have to be examined in order to label a topic. This is typically done by looking at TF-IDF weighted frequencies of the top words in each cluster.

The last topic model we consider is FASTopic of Wu et al. (2024). Here, again a pre-trained BERT model is used to gather sentence embeddings which are then aggregated into document embeddings. The novelty in FASTopic is that it then models the relation between topic and word embeddings through dual semantic-relation reconstruction (DSR). The advantage here is that documents are not clustered like in BERTopic but document-topic and topic-word distributions arise naturally from the framework used.

The choice of models span traditional, bag-of-word mixture topic models (*keyATM*), contextualized BERT-based document-cluster approaches (BERTopic) and a BERT-based approach that preserves the main idea of mixture topic models, namely that documents consist of multiple topics and topics are defined by word frequency.

Evaluation Metrics The most important choice in topic modeling is the number of allowed topics. For that purpose, we evaluate all models using three standard metrics.

First, we examine two measures for topic *coherence*, a concept which denotes the statistical tendency of top terms to co-occur. Coherence is not to be confused with topic *interpretability*, which refers to the ease with which a human can assign a clear label to each topic by inspecting its highest-probability terms; For instance, a topic containing the bigrams "quantitative easing"

and "interest rate" may score highly on coherence but still be hard to interpret if other top terms are obscure. Conversely, a "labor market" topic may exhibit lower coherence if its terms co-occur less frequently. Topic coherence measures are data based and can be easily calculated, whereas topic interpretability would need human judgment as input.

The first coherence measure we compute is the semantic coherence (SC) score of Mimno et al. (2011). For each model, we consider the top 10 terms per topic and calculate coherence based on how often these top terms occur together in the estimation corpus. The metric can be calculated as follows:

$$\text{SC} = \sum_{i=2}^{M} \sum_{j=1}^{i-1} \log \frac{D(w_i, w_j) + \epsilon}{D(w_j)}, \quad \epsilon = 1 \tag{1}$$

with $D(w_i, w_j)$ denoting the number of documents containing both tokens w_i and w_j, and $D(w_j)$ the number of documents containing w_j. SC hence considers joint occurrence whenever two terms occur in the same document. Values closer to zero indicate higher coherence. Token pairs for which co-occurrence is computed proceeds by going down the list of high-probability tokens and. That is for each token, SC quantifies how frequently it appears alongside each of the tokens ranked above it. This directional strategy—examining whether later terms co-occur with the earlier, more central ones—ensures that the strongest topic indicators truly anchor the overall theme, while avoiding the computational burden and noise sensitivity that arise from evaluating every possible pair. By mirroring the way readers naturally scan and interpret ordered word lists, this approach yields a coherence score that aligns closely with human judgments of topical consistency.

Second, we compute the Normalized Pointwise Mutual Information (NPMI) score of Bouma (2009), which is an alternative measure for topic coherence. NPMI is based on a normalization of token-pair association strength within a fixed context window of size W. That is, co-occurrence is not evaluated based on whether two tokens appear in the same document, but whether they appear within a window of pre-specified length. In our case, we choose a window size of $W = 5$ and denote the emerging total number of sliding windows by $D_{W,ref}$. In contrast to SC, co-occurrence of the terms is assessed using an external reference corpus. For our purpose, we decided to use central bank speeches contained in the newly created data base of Campiglio et al. (2025), where we dropped texts contained in our data base to ensure no duplicates are present for the evaluation. Finally, $D_{W,ref}(w_i)$ refers to the number of windows containing term w_i, and $D_{W,ref}(w_i, w_j)$ the number of windows in which both w_i and w_j co-occur. Then

$$P_W(w_i) \;=\; \frac{D_{W,ref}(w_i)}{D_{W,ref}}, \qquad P_W(w_i, w_j) \;=\; \frac{D_{W,ref}(w_i, w_j)}{D_{W,ref}}$$

and the NPMI score for a token pair (w_i, w_j) becomes

$$\mathrm{NPMI}_W(w_i, w_j) \;=\; \frac{\log\Big[P_W(w_i, w_j) \,/\, \Big(P_W(w_i)\, P_W(w_j)\Big)\Big]}{-\,\log P_W(w_i, w_j)}. \tag{2}$$

Higher values of NPMI_W indicate stronger semantic coherence, as they reflect that w_i and w_j co-occur more often than expected by chance within windows of size W. This measure is calculated per topic for all possible combinations of top terms (per topic, excluding self-pairs) and then averaged. Note that it is of ample importance to use exactly the same pre-processings in the reference corpus than in the estimation corpus, otherwise the top terms from the topic model cannot be matched in $D_{W,ref}$.

Our case warrants special attention since in this study, terms relate to bi-grams, not unigrams / words. On the one hand, incorporating bigrams as the fundamental units in both directional and NPMI-based coherence measures offers the potential to capture richer, phrase-level semantics—particularly in domains characterized by frequent multiword expressions—but it also exacerbates data sparsity and computational complexity. Because bigrams occur far less often than individual words, estimates of co-occurrence probabilities can become noisier. This problem can be addressed, however, by using a large corpus in order to achieve stable measurements, which is the case in this study. Calculation of the NPMI scores with bigrams are especially computationally costly, since the quadratic expansion of the candidate token set further amplifies the cost of counting co-occurrences within documents or sliding windows. Moreover it demands careful specification of how bigrams are extracted and aligned in context, which we did by synonym matching. The most effective synonym matching among the tested options for bigrams was the pairwise combination of lemmatized, lowercased words (e.g., "labor market" → "labour market", "economic growth" → "growth economic"). This matching strategy resulted in fewer than 10% of bigrams from all three models being unmatched by the tokens in the reference corpus.

Finally, we consider *topic diversity (TD)*, which quantifies the degree to which a topic model's set of topics covers distinct themes rather than repeating similar token clusters, and it serves as a crucial counterbalance to coherence. While coherence measures assess the internal semantic consistency of each topic, they may inadvertently favor models that produce many narrowly focused or overlapping topics. In contrast, diversity metrics—often computed as the proportion of unique token types among the top terms across all topics—ensure that the model's output spans the full breadth of the corpus. Topic diversity is calculated as:

$$\mathrm{TD} = \frac{\text{Number of unique terms in top } N \text{ terms across all topics}}{N \times K}, \tag{3}$$

and lies between 0 (all topics share the same top terms) and 1 (no overlap in top terms).

High diversity indicates that topics capture different facets of the data rather than converging on the same frequent terms, which is essential for applications such as document exploration, summarization, and downstream predictive tasks.

By jointly optimizing for both coherence and diversity, one can obtain topic models that are not only interpretable at the individual topic level but also collectively informative and non-redundant. The results are shown in Table 1, where we present the average topic coherence measure (averaged over all topics) for the SC and NPMI metrics. Since each model class employs different preprocessing approaches and vocabularies, NPMI and SC scores should not be compared directly across model types. Instead, these coherence measures are most appropriately used to select the optimal number of topics (K) within each model class. Table Table 1 also shows the *Shannon Entropy* (SE) of the topic distribution, based on assigning each speech to the topic with the highest prevalence. Higher values indicate a more even distribution of speeches across topics, reflecting a more balanced classification.

Table 1: Model evaluation metrics for optimal topic selection

Model	K	NPMI	SC	TD	SE
BERTopic	10	0.359	-24.90	0.700	1.339
BERTopic	15	0.331	-28.34	0.693	1.824
BERTopic	20	0.310	-29.85	0.665	2.297
BERTopic	50	0.295	-30.35	0.566	3.365
FASTopic	10	0.270	-24.71	1.000	2.278
FASTopic	15	0.236	-26.74	1.000	2.666
FASTopic	20	0.228	-27.69	1.000	2.960
FASTopic	50	0.220	-31.68	1.000	3.718
KeyATM	10	0.245	-80.32	0.850	1.920
KeyATM	15	0.237	-89.63	0.873	2.460
KeyATM	20	0.226	-93.34	0.820	2.590
KeyATM	50	0.214	-86.60	0.742	3.280

Notes: This table presents evaluation metrics for three topic modeling approaches across different numbers of topics (K). NPMI measures normalized pointwise mutual information (higher values indicate better coherence), SC represents semantic coherence (values closer to zero indicate better performance), and TD shows topic diversity (higher values indicate more diverse topics). The evaluation is based on the top 10 terms per topic across all model configurations. SE stands for Shannon Entropy.

BERTopic demonstrates optimal performance at smaller topic numbers, with $K = 10$ yielding the highest NPMI score (0.359) and most favorable SC value (-24.90) within this model class. While coherence peaks at $K = 10$, topic diversity exhibits a declining trend as K increases, decreasing from 0.700 to 0.566. Notably, BERTopic puts most mass on a few topics producing a very skewed distribution, which can be seen by rather low SE values.

FASTopic exhibits distinctive characteristics across all tested configurations. The model achieves peak NPMI performance at $K = 10$ (0.27). A remarkable feature of FASTopic is its perfect topic diversity (TD = 1.0) maintained across all values of K, indicating complete separation between topic terms with no overlap. FASTopic, however, produces a very even distribution, as can be seen by the comparably high SE values for all number of topics. This contradicts typical expectations of natural topic hierarchies where some topics should be more prevalent than others.

keyATM maximizes both NPMI scores (0.245) and SC scores (-80.32) for $K = 10$. Topic diversity scores vary between 0.742 and 0.873, with peak diversity observed at $K = 15$. The Shannon Entropy (SE) scores for *keyATM* fall between those of BERTopic and FASTopic, indicating that it strikes a balance between concentrated and evenly distributed topic assignments.

The three models exhibit distinct patterns regarding topic redundancy. BERTopic produces topics with the largest overlap between top terms, potentially indicating some redundancy in topic representation. In contrast, FASTopic creates essentially no overlap between topics, achieving perfect separation of topic vocabularies. *keyATM* demonstrates intermediate levels of topic overlap, balancing between the extremes observed in the other two approaches. Regarding document distribution patterns, BERTopic tends to assign a disproportionate share of documents to only a few topics, which may be unrealistic given the diversity of policy discourse. In contrast, FASTopic imposes an almost uniform allocation across topics, which can also be implausible for most corpora where topic prevalence is typically skewed. *keyATM* offers a more balanced middle ground, yielding distributions where some topics are broadly represented while others capture more specialized content—closely reflecting the structure of real-world speech corpora.

Based on the evaluation metrics, we recommend the following optimal configurations, for which we then manually judge topic interpretability by investigating the top 10 terms per identified topic: BERTopic and FASTopic with $K = 10$ which maximizes coherence and topic diversity; and *keyATM* with $K = 15$ which balances coherence and diversity measures.

We list the top ten terms for all three candidate models in Table 4 and provide a first, manual assessment of topic interpretability. The table shows, that topics identified by BERTopic are especially hard to interpret. For example, the bigrams *financial_stability, financial_market, financial_crisis* and *financial_sector* appear in the first four topics. This makes it hard to identify distinct topics. Our main explanation for this is the fundamental difference of BERT based models in uncovering semantically closed tokens and documents versus mixture models that unveil themes within documents based on token co-occurrence / frequency patterns. More specifically, in embedding-based topic discovery (e.g. BERTopic), each document is first encoded into a single high-dimensional vector that captures its overall semantic content; clustering then groups together documents with similar "global" themes, but is not suited to decompose a document into multiple, co-occurring subtopics. For example, for our corpus a monetary policy statement that always has the same structure, maybe with an opening that relates to the current economic environment, followed by an interest rate decision and concluding with an outlook, encoded with BERT would cluster similarly structured documents, but cannot unveil the three different topics within the document - a task for which topic mixture models have been explicitly designed. This fundamental modeling difference explains why embedding-based clusters often merge them into one group, whereas LDA disentangles their internal topical structure—a phenomenon also observed in Contextualized Topic Models, which combine BERT embeddings with word-level topic inference to recover both semantic coherence and multi-topic mixtures

(Bianchi et al., 2021). As an alternative strategy, not pursued in this study, one could use a two-step procedure, cluster documents based on their semantic relationships using BERT-based embeddings on the document level and then run an LDA type model for each cluster.

4 What do central bankers talk about

4.1 Evidence from the *keyATM* model

In this section, we will present the main results based on our preferred model, which is the *keyATM* model with 13 keyword-assisted topics and 2 residual topics.

We label the topics according to three considerations. First and foremost, we look at the word frequency per topic, that is, the top words that appear in each topic. These are displayed as word clouds in Figures 2 and 3. Second, we consider topics well defined when the seeded keywords appear among the top words per topic. The top ten words per topic are displayed in Table 5 with the ✓indicating that the word belongs to the set of seeded keywords. Last, we provide in the appendix in Table 6 the top five speeches per topic.

The topic modeling analysis reveals a coherent thematic structure across central bank communications. The first topic clearly revolves around issues related to **monetary policy**. The most prominent terms include *monetary_policy*, *interest_rate*, *central_bank*, *price_stability*, *inflation_target*, and *inflation_expectation*. Among the top terms, five originate from the seeded keywords, indicating successful topic identification. Notably, while our approach successfully identifies the monetary policy topic, it does not capture the varying approaches across different central banks, where some may favor more active policy implementations while others adopt more passive stances.

The second topic centers on **financial stability**, characterized by terms such as *financial_stability*, *financial_system*, *financial_market*, and *financial_sector*. Four of the top ten terms derive from seeded keywords. The third topic encompasses broader **finance** issues, featuring terms like *financial_institution*, *financial_inclusion*, *financial_product*, and *financial_literacy*. This topic addresses financial inclusion and literacy concerns, with four of its top terms originating from the keyword set. Some overlap exists with the financial stability topic and the banking supervision topic.

The fourth topic focuses on **banking supervision**, defined by terms including *bank_system*, *risk_management*, *balance_sheet*, *capital_requirement*, *bank_sector*, *stress_test*, *bank_supervision*, and *credit_risk*. Seven of the ten most prominent terms belong to the seeded keywords, with minimal cross-topic overlap. The fifth topic addresses **payment** systems, characterized by terms such as *payment_system*, *financial_service*, *settlement_system*, *payment_service*, *retail_payment*, *service_provider*, *payment_settlement*, *new_technology*, and *digital_euro*. This well-seeded topic shows limited overlap with other topics.

The sixth topic relates to the **labor market**, though it appears somewhat mixed in content. It

Figure 2: Word clouds for topics 1-8 from *keyATM* analysis

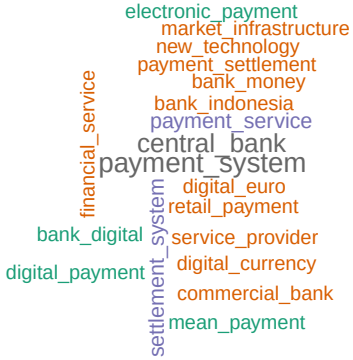
(a) Monetary policy (1_mp)

(b) Financial stability (2_fs)

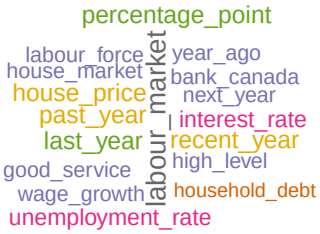
(c) Finance (3_finance)

(d) Banking supervision (4_bsuper)

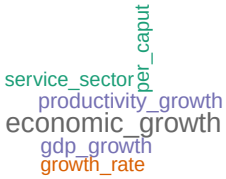
(e) Payment (5_payment)



(f) Labor market (6_labor)



(g) Productivity (7_productivity)



(h) Islamic finance (8_islamic)

Notes: Word clouds visualizing the most prominent terms for topics 1-8 identified by the *keyATM* model. Word size corresponds to term probability within each topic. These topics represent core central banking functions including monetary policy, financial stability, banking supervision, and specialized areas such as Islamic finance. The visualization confirms successful keyword seeding, with domain-specific bigrams like "monetary_policy," "financial_stability," and "risk_management" appearing prominently in their respective topics.

contains seeded keywords like *labour_market* and *wage_growth*, alongside temporal terms such as *recent_year*, *past_year*, and housing-related terms including *house_price* and *house_market*. The seventh topic focuses on **productivity**, defined by terms like *economic_growth*, *productivity_growth*, *growth_rate*, and *gdp_growth*. This topic contains three seeded keywords and shows some overlap with the financial stability topic. The eighth topic addresses **Islamic finance**, featuring terms such as *islamic_finance*, *financial_system*, *capital_market*, and *islamic_financial*.

The remaining topics – shown in Figure 3 – include several regional and policy-specific themes. The **fiscal policy** topic is well-represented by terms including *fiscal_policy*, *economic_policy*, *public_finance*, *private_sector*, *public_debt*, *sovereign_debt*, and *government_bond*. The **monetary union** topic captures European Central Bank communications with terms like *monetary_union*, *member_state*, *european_central*, *govern_council*, and *bank_union*.

A **global economy** topic emerges with terms such as *exchange_rate*, *foreign_exchange*, *current_account*, *emerge_market*, *global_financial*, and *capital_flow*. The **green economy** topic, though smaller, features terms like *climate_change*, *climate_risk*, and *financial_risk*. Additionally, a regional topic captures **US economy** terminology specific to Federal Reserve communications, including *federal_reserve*, *reserve_bank*, *reserve_system*, *board_governor*, and *fund_rate*.

Finally, two residual topics emerge that primarily contain procedural and temporal language rather than substantive policy content, featuring terms such as *year_ago*, *can_make*, *decision_make*, *many_year*, *question_whether*, and *answer_question*. These topics likely capture the conversational and administrative aspects of central bank communications rather than specific policy themes.

Having identified and labeled the topics, we can analyze them further. In Figure 4, we show the proportions of the assigned topics, aggregated over time. The figure shows monetary policy and financial stability as the most prevalent topics. Other topics that received a lot of attention are the finance topic and the banking supervision topic. The remaining topics are equally often addressed in speeches with the exception of the payment topic, which covers aspects of digital payments and is hence a more recent topic.

The dynamic nature of our model also allows us to discover time variation in topic prevalence. This is displayed in Figure 5, where we categorized the topics into "evergreens", those topics that are always present in the corpus, ones with falling and ones with rising popularity. The figure demonstrates that central banks always devote some focus to the topics that align with their core mandates. This includes foremost monetary policy, financial stability and banking supervision. Other, mandate related topics comprise of the labor market and productivity topics, as well as islamic finance. Most of these topics are addressed with a stable proportion over time. An exception to this is financial stability, for which topic proportion spiked in the aftermath of the global financial crisis.

We also see two topics that get nowadays less attention, for one we have less texts devoted to the topic monetary union. This topic was heavily dominated by speeches of the ECB and its

Figure 3: Word clouds for topics 9-15 from *keyATM* analysis

(a) Fiscal policy (9_fiscal)

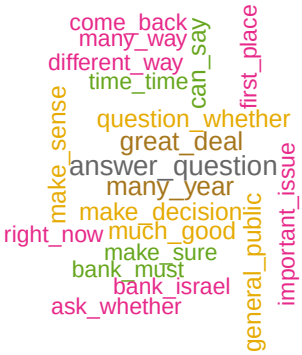
(b) Monetary union (10_munion)

(c) Global economy (11_global)

(d) Green economy (12_gr)

(e) US economy (13_federal)

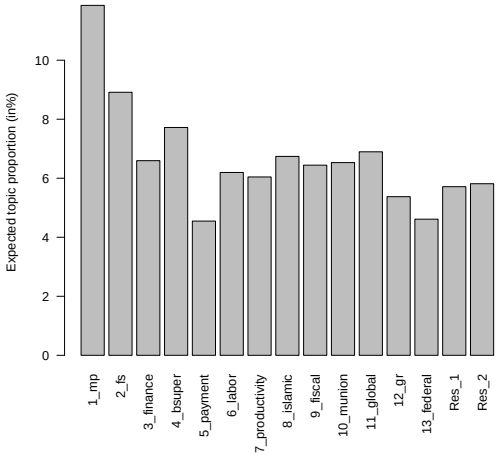
(f) Residual 1



(g) Residual 2

Notes: Word clouds for the remaining topics (9-15) identified by the *keyATM* model, including regional/institutional topics and residual categories. Topics 9-13 capture broader policy themes (fiscal policy, monetary union, global economy, green finance, US-specific terms), while the two residual topics primarily contain procedural and conversational language rather than substantive policy content. The prominence of region-specific terms (e.g., "federal_reserve," "monetary_union") demonstrates the model's ability to identify institutional and geographic patterns in central bank communication.

Figure 4: Overall topic prevalence in central bank speeches



Notes: This figure displays the average proportion of each topic across the entire corpus of 20,150 central bank speeches from 1997-2025. Monetary policy and financial stability emerge as the most prevalent themes, reflecting central banks' core mandates. The relatively low prevalence of specialized topics like Islamic finance and payment systems reflects their regional or temporal specificity, while the prominence of general finance and banking supervision topics indicates their broad relevance across different central banking systems.

predecessor the European Monetary Institute (EMI) before the creation of the euro. Second, the topic on the global economy gets less attention now then it used to get in the beginning of the sample period. This development could coincide with the global trend of diminishing globalisation. Last, we have two topics that are nowadays addressed more frequently by central banks, namely the green economy and payment systems. The rise of climate-related speeches we see aligns with findings of Feldkircher and Teliha (2024). The payment topic also evolves around digitalization and the research and attention towards e.g., central bank digital currencies has increased significantly over the last years.

As a robustness check, we also estimated a 15-topic FASTopic model, which combines BERT embeddings with mixture modeling approaches. We also used an LLM for automatic topic labeling. Results (presented in Appendix B) show substantially similar thematic structures, confirming that our findings are not driven by the specific choice of keyATM.

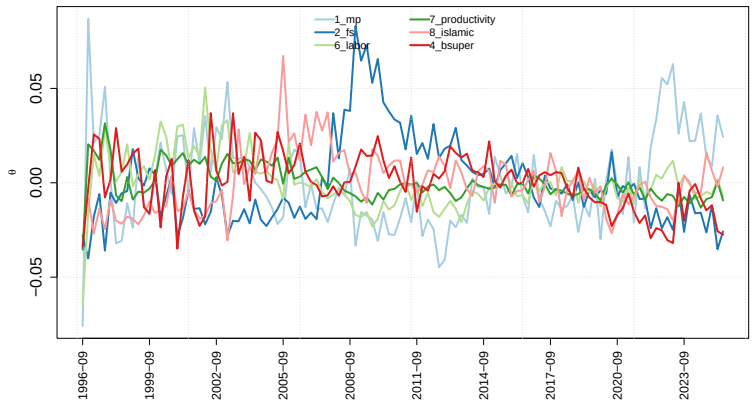
Summing up, we find that the *keyATM* model yields plausible results in terms of topic proportions, time trends, and top speeches. In the next two sections, we provide further analysis to first see, which central banks are topic leaders and second how macroeconomic and institutional variables influence topic prevalence.

4.2 Spillovers of topics and ideas

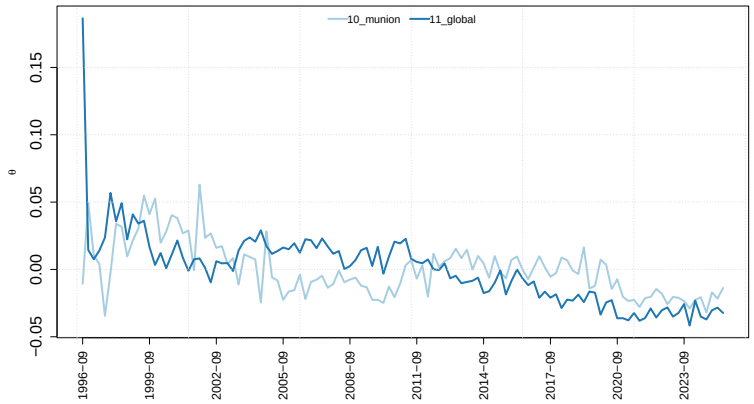
In this section, we present a cross-sectional analysis of our main results. In particular, we want to gather information on which central banks engaged most strongly in which topic and whether there are some banks that set the topics that later other central banks adopt. For that purpose, we had to aggregate the topic prevalence into regional clusters since we do not have

Figure 5: Temporal evolution of topic prevalence (demeaned trends)

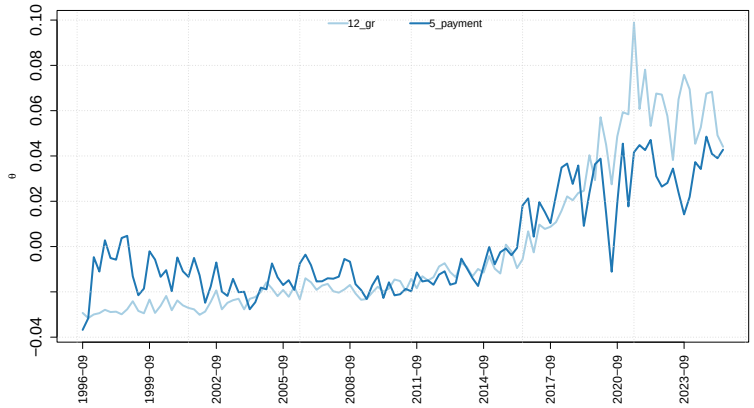
(a) Evergreen topics: Stable prevalence over time



(b) Declining topics: Falling popularity



(c) Emerging topics: Rising popularity



Notes: Temporal trends of topic proportions from the *keyATM* model, showing demeaned values to highlight deviations from average prevalence. Evergreen topics maintain relatively stable attention throughout the sample period, with financial stability showing a notable spike during the 2008-2012 Global Financial Crisis. Declining topics reflect institutional and global changes: monetary union discussions decreased after Euro establishment, while global economy topics waned amid reduced international integration. Rising topics capture emerging priorities: green economy attention surged post-2015 Paris Agreement, while payment systems gained prominence with digital currency developments.

otherwise a continuous set of speeches for the different central banks. We also limit the time period ranging from 1999-01-01 to 2024-12-31. In our analysis, we gather information on seven major central banks, namely the Bank of England, the ECB, the Deutsche Bundesbank, the US Fed, the Bank of Japan, the Bank of Canada and the Reserve Bank of Australia and seven regions, Western Europe, Eastern Europe, Latin America, Asia, Middle-East and North Africa (MENA), Africa and Pacific. Table 2 summarized the regions and also displays the number of observations with at least one speech available. The table shows that data for Latin American, African, Pacific and central banks from the MENA regions is scarce, with missing data reaching up to 70%.

We then proceed by aggregating monthly topic proportions by summing rather than averaging individual speech contributions within each bank-topic-month combination. This choice reflects both practical and theoretical considerations. Practically, topic frequency captures the intensity of institutional focus, as sustained emphasis on particular themes signals strategic priorities. Theoretically, this approach acknowledges that larger central banks naturally command greater influence in shaping global policy discourse and can exert peer pressure on smaller institutions to address particular topics. Empirical evidence supports this mechanism: Deyris (2023) and Feldkircher and Teliha (2024) demonstrate that reputational costs associated with neglecting emerging themes like green finance incentivize smaller central banks to adopt communication strategies pioneered by major institutions.

Finally, we employ two complementary approaches to the regionally aggregated topic prevalence data. Our first aim is to identify which central banks positioned themselves as *topic leaders*. For that purpose, we utilize cross-sectional percentile ranks which measures absolute topic dominance, providing results with direct economic interpretation as competitive rankings among central banks.

More specifically and for each time period t and topic k, we rank all central banks i by their aggregated topic emphasis π_{ikt} and compute cross-sectional percentile ranks as $P_{ikt} = (N_{kt} - \text{Rank}_{ikt})/(N_{kt} - 1)$, where Rank_{ikt} denotes the ordinal rank of bank i among all banks at time t for topic k, and N_{kt} represents the number of banks with available data. A central bank emerges as a topic leader at the first time period τ when $P_{ikt} \geq 0.8$, defining emergence as initial entry into the top quintile of institutions. This threshold balances between identifying genuine leaders while maintaining sufficient observations for robust statistical inference. Implementation requires minimum coverage of three banks per time-topic observation to ensure meaningful percentile calculations, and we restrict analysis to periods where major central banks maintain comparable data availability to address concerns about spurious early emergence from limited comparison sets.

The results are provided in the top panel of Figure 6, where we exclude the region-specific US topic and the two residual topics on the grounds that interpretation and spillover potential is limited. The graph demonstrates that for example, the topic of monetary union was heavily discussed in Europe (Eastern and Western Europe, ECB and Germany) right from the beginning

Table 2: Regional groupings, data coverage, and topic specialization patterns

Region	Data Availability	Most Emphasized Topic
Major Central Banks		
USA, ECB, JPN, GBR, DEU, AUS, CAN	USA: 310/312 (2, 0.6%) ECB: 309/312 (3, 1.0%) JPN: 252/312 (60, 19.2%) GBR: 251/312 (61, 19.6%) DEU: 244/312 (68, 21.8%) AUS: 246/312 (66, 21.2%) CAN: 285/312 (27, 8.7%)	USA: 4_bsuper (0.899) ECB: 10_munion (1.60) JPN: 7_productivity (0.396) GBR: 6_labor (0.359) DEU: 10_munion (0.595) AUS: 6_labor (0.524) CAN: 6_labor (0.331)
Western Europe		
AUT, BEL, CHE, DNK, ESP, FIN, FRA, IRL, ISL, ITA, LUX, NLD, NOR, PRT, SWE, MLT	Western Europe: 310/312 (2, 0.6%)	9_fiscal (1.18)
Eastern Europe		
BGR, CZE, EST, HRV, HUN, LTU, LVA, POL, ROU, SVK, SVN, UKR, MKD, BIH, SRB, ALB, CYP, GRC, XKX, RUS, TUR	Eastern Europe: 276/312 (36, 11.5%)	11_global (0.548)
Latin America		
MEX, ARG, BRA, CHL, COL, ECCB	Latin America: 179/312 (133, 42.6%)	11_global (0.345)
Asia		
CHN, HKG, KOR, IND, SGP, THA	Asia: 305/312 (7, 2.2%)	3_finance (1.05)
MENA		
ARE, SAU, ISR	MENA: 91/312 (221, 70.8%)	7_productivity (0.216)
Africa		
ZAF, NGA, KEN	Africa: 243/312 (69, 22.1%)	3_finance (0.414)
Pacific		
NZL, FJI	Pacific: 201/312 (111, 35.6%)	3_finance (0.231)

Notes: Regional aggregation scheme for lead-lag analysis covering January 1999 to December 2024. Data availability shows months with at least one speech, with missing data percentages in parentheses. Most emphasized topics represent cumulative topic proportions over the analysis period, revealing institutional specialization patterns: advanced economies focus on core mandates (banking supervision, monetary union), while emerging regions emphasize global integration and financial development themes. High missing data rates for MENA (70.8%) and Latin America (42.6%) reflect limited BIS archive representation from these regions.

of the joint sample period, which is around 1999/2000. By contrast other major central banks such as the US Fed or the Bank of Japan did not pay particular attention towards the topics until 2005 or 2010 respectively. The topics of the green economy, financial inclusion and banking supervision have been emphasized early from almost all central banks with the exception of the Australian Reserve bank. By contrast, the Reserve Bank was in the lead regarding the labor market topic, which it emphasized right from the beginning of the sample period.

Analysis of chronological topic pioneering (which bank was the first top perfomer per topic) reveals a remarkably balanced innovation landscape, with emerging economy central banks establishing first-mover leadership in 6 out of 12 policy topics (50.0%), while Western and major economy banks pioneered the remaining 6 topics. The ECB emerged as the most prolific topic pioneer, establishing first-mover advantage in 4 policy areas (monetary union, monetary policy, payment systems, and fiscal policy coordination). Regional emerging economy banks demonstrated strong early specialization, with Africa pioneering leadership in global economy integration, green economy initiatives, and finance topics, while Asia established first-mover positions in financial stability, banking supervision, and Islamic banking frameworks.

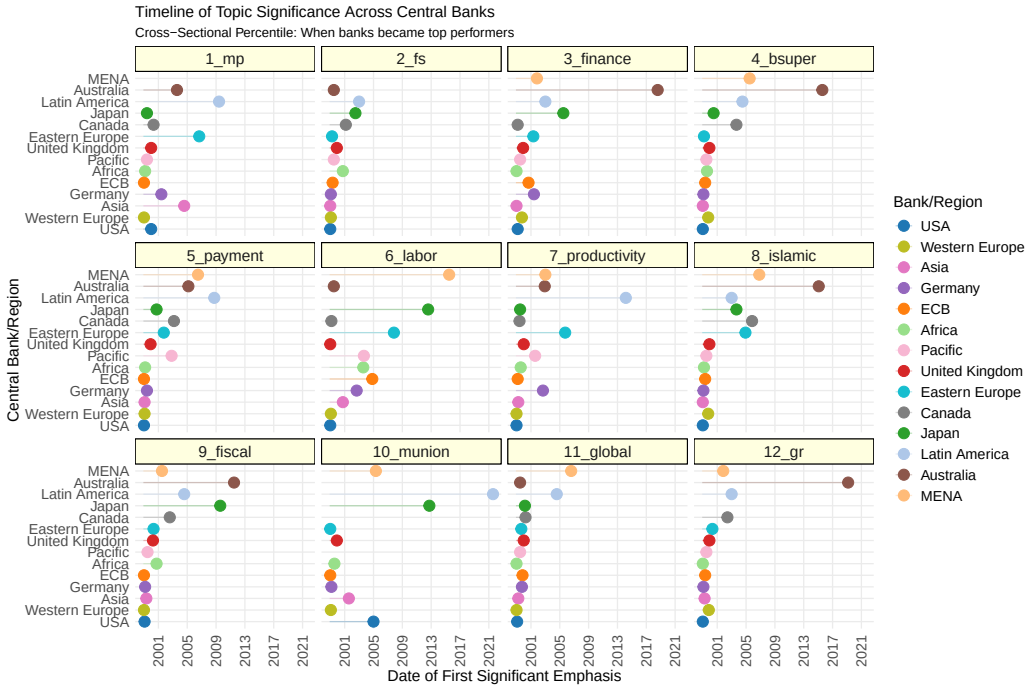
Beyond pioneering patterns, substantial heterogeneity exists in subsequent topic adoption and performance quality across central banking systems. Fundamental policy areas such as monetary policy and financial stability achieved widespread early adoption across most central banks, reflecting their core mandates. However, more specialized or emerging policy domains reveal distinct regional patterns. Among major central banks, some institutions demonstrate relatively slower adoption in specific policy areas, with the analysis suggesting differentiated emphasis patterns where certain advanced economy banks show lower relative performance in emerging policy domains such as green economy frameworks and specialized financial sector oversight.

Emerging market central banks often exhibit delayed or lower-intensity adoption in specific policy areas, particularly those requiring substantial institutional infrastructure or international coordination. These patterns manifest across several domains, including lower adoption intensity of productivity-focused policies among Latin American and Eastern European institutions, delayed emphasis in payment system innovation across MENA and Latin American regions, limited relevance and adoption of monetary union coordination in MENA and Latin American contexts, slower development of advanced finance topics in MENA and Latin American banking systems, and variable adoption patterns of global economy integration across MENA and Latin American institutions. These patterns reflect not only institutional capacity differences but also varying policy priorities, economic structures, and regional integration frameworks that shape central bank policy emphasis over time.

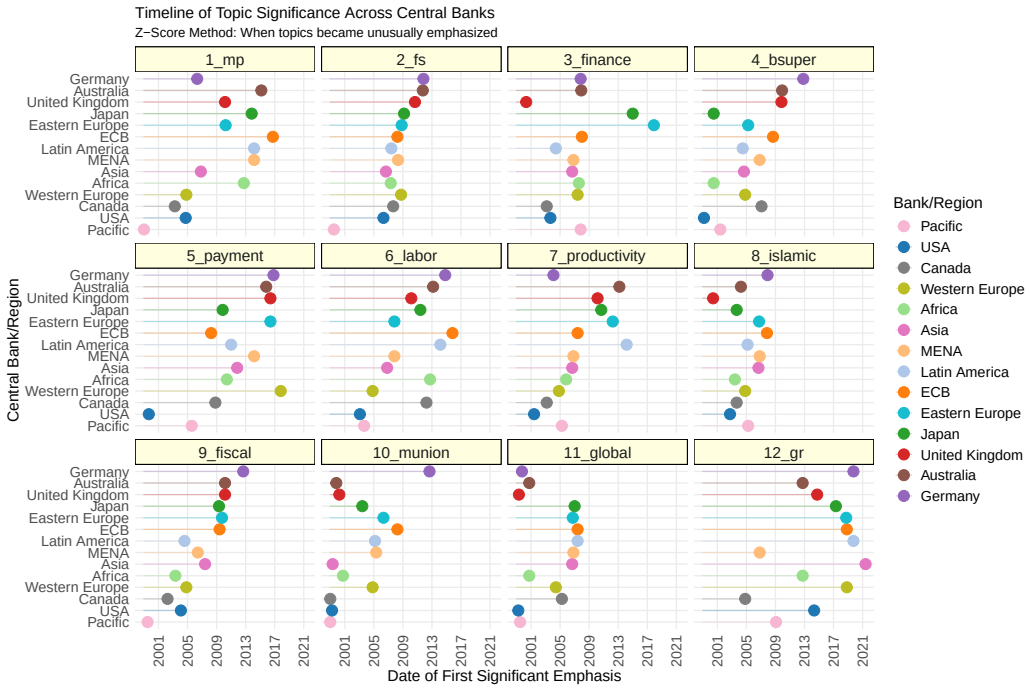
The second measure we employ is the Z-score methodology which identifies periods of unusual policy emphasis by measuring when a central bank's attention to a specific topic deviates significantly from its historical baseline, capturing the timing of *policy attention shifts* rather than relative performance rankings. For each central bank i and policy topic j, let θ_{ijt} represent

Figure 6: Leadership patterns and crisis-driven convergence in central bank communication

(a) Cross-sectional leadership analysis (percentile ranks)



(b) Crisis-driven attention synchronization (Z-score analysis)



Notes: Analysis of topic leadership and attention synchronization patterns from 1999-2024. Top panel: Cross-sectional percentile method identifies when central banks first achieved top-quintile performance ($\geq$80th percentile) relative to peers for each topic, revealing stable leadership hierarchies concentrated in early sample years. Bottom panel: Z-score method measures when banks exhibited unusually high topic emphasis relative to their own historical baselines (Z-score >1.5), demonstrating crisis-driven synchronization particularly during 2008-2012 Global Financial Crisis. The contrast between methods reveals that while relative performance rankings remain stable, systemic crises trigger synchronized shifts in communication priorities across the global central banking system.

the measured topic prevalance at time t, and we calculate the historical mean and standard deviation using all available observations up to time $t-1$ as $\mu_{ij}^{(t-1)} = \frac{1}{t-1}\sum_{\tau=1}^{t-1} X_{ij\tau}$ and $\sigma_{ij}^{(t-1)} = \sqrt{\frac{1}{t-1}\sum_{\tau=1}^{t-1}(X_{ij\tau} - \mu_{ij}^{(t-1)})^2}$. The Z-score at time t measures how many standard deviations the current emphasis deviates from the historical baseline as $Z_{ijt} = \frac{\theta_{ijt} - \mu_{ij}^{(t-1)}}{\sigma_{ij}^{(t-1)}}$, where $Z_{ijt} > 0$ indicates above-average emphasis relative to the central bank's historical pattern. We define unusual emphasis as occurring when the Z-score exceeds the historical mean by at least two standard deviations.

The results are displayed in Figure 6 in the bottom panel and reveal fundamentally different patterns of policy attention timing that provide robust evidence for crisis-driven synchronization. This approach shows a more temporally distributed pattern of emergence, with only 22% of unusual emphasis events occurring in the early period (1999-2003) compared to 33% during the extended crisis and post-crisis period (2008-2015). Most significantly, the Z-score analysis demonstrates strong crisis convergence, with 23.8% of all unusual emphasis events concentrated in the 2008-2012 period—a six-fold increase compared to the percentile method. More specifically, the topic that is perfectly synchronized with a crisis event, is financial stability. Here, we see clearly that all central banks put more emphasis - compared to their historical, central bank specific baseline - on communication related to financial stability. This shift happened around 2008/09, the period of the global financial crisis (GFC). We also see that most banks put more emphasis during that period on topic 3, finance which includes the themes of financial literacy and financial inclusion. We also see that most central banks put unusual high emphasis on the green topic around 2016, which is shortly after the Paris Agreement was signed in December 2015, a finding which is in line with Feldkircher and Teliha (2024). For the majority of banks, the topics productivity and global economy gained unusual attention around 2005. In general, the crisis-period convergence provides strong evidence that systemic financial crises accelerate policy attention synchronization across the global central banking system. The temporal concentration of unusual emphasis during 2008-2012—representing nearly a quarter of all such events in the sample—suggests that crisis conditions compel central banks to simultaneously reevaluate and adjust their policy priorities, particularly for global coordination and fiscal policy topics. This synchronization of policy attention timing, combined with the stable performance hierarchies revealed by the percentile analysis, supports a nuanced understanding of policy diffusion where crisis events accelerate the spread of policy attention while preserving underlying performance leadership structures.

4.3 Covariate analysis

This analysis employs an augmented version of the baseline model of Eshima et al. (2024) that incorporates covariates to model topic prevalence within a static framework. The model uses a Dirichlet-Multinomial regression following Mimno and McCallum (2008):

$$\boldsymbol{\theta}_d \sim \text{LogisticNormal}(\boldsymbol{\mu}_d, \boldsymbol{\Sigma}) \tag{4}$$

$$\boldsymbol{\mu}_d = \mathbf{X}_d\boldsymbol{\beta} \tag{5}$$

linking topic prevalence $\boldsymbol{\theta}_d$ for document d to a set of covariates contained in design matrix $\mathbf{X}_d$.

We examine four key determinants of topic emphasis: macroeconomic conditions (CPI inflation levels), institutional characteristics (Central Bank Independence scores from Romelli, 2022), temporal shifts (pre- and post-Global Financial Crisis periods following Bohl et al., 2023), and communication formats (statements, interviews, and reports). We transform both variables into high ($CBIE_{high}, Dp_{high}$) and low ($CBIE_{low}, Dp_{low}$) dummy variables by setting the high dummy equal to 1 for observations above the 75th cross-sectional percentile of the respective variable and the low dummy equal to 1 for observations below the 25th cross-sectional percentile, effectively capturing the top and bottom quartiles while excluding the middle 50% of the distribution. We then add dummy variables for the communication types, such that the final design matrix looks like $\mathbf{X}_d = (CBIE_{high}, CBIE_{low}, Dp_{high}, Dp_{low}, D_{GFC}, D_{interview}, D_{report}, D_{statement})$.

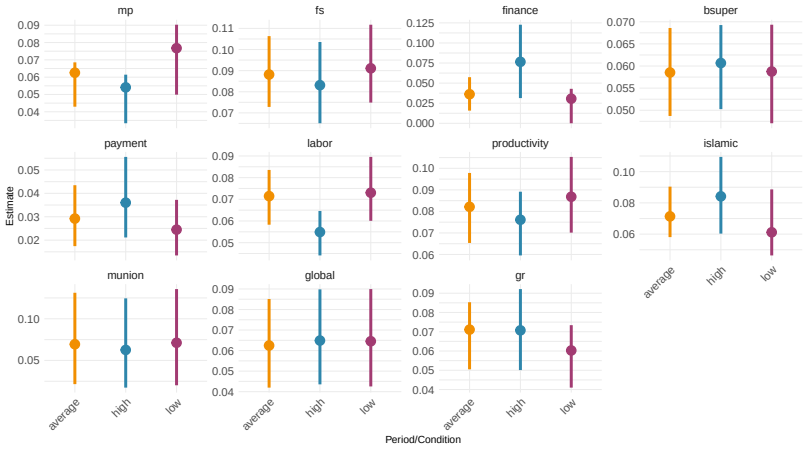
We start by investigating how topic prevalence differs for high and levels of inflation and central bank independence shown in Figure 7.

High versus low inflation periods reveal distinct communication allocation patterns. For example, when central banks experience high inflation, less communication is devoted to the core topic of monetary policy, whereas in low inflation regimes, more space in communication is geared towards monetary policy. This finding could reflect the experience of major central banks over the last decades of rather low inflation and their attempt at explaining to the public new monetary policy measures to bring up inflation to target (such as quantitative easing and other forms of unconventional monetary policy tools). We also see less emphasis on the labor market topic, whereas more emphasis is put on the finance topic, which covers also financial literacy. This finding could imply that in times of high inflation, central banks want the public to be well aware of how financial products and risks work, to make better borrowing and investment decisions, which should reduce the likelihood of financial crises and bubbles.

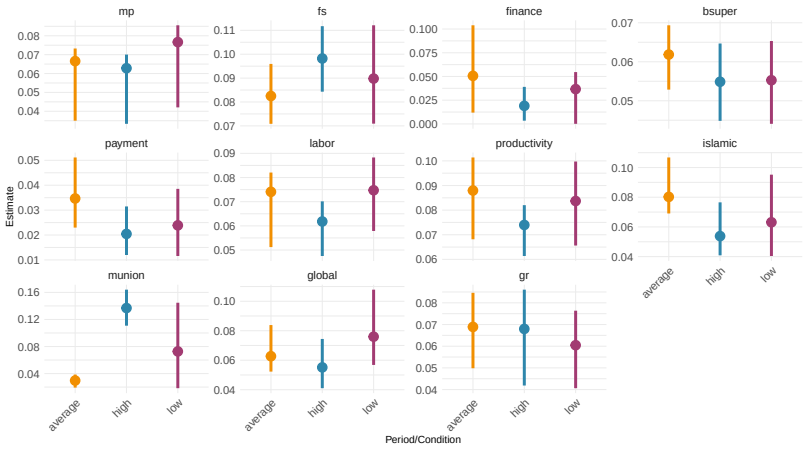
The CBIE analysis reveals systematic differences in communication patterns based on institutional independence levels. In general, the more independent banks tend to have narrower mandates (Dikau and Volz, 2021), but that does not a priori constrain them in the breadth of topics they address. Empirical evidence on this has been recently provided by Leek and Bischl (2025) who document that the more independent central banks are, the more policy pressure they are exposed to, which they then address in their communication, resulting into less focus on solely mandate related topics. Our results show that highly independent central banks have a high topic prevalence for the monetary union topic. This finding could be driven by the ECB which is a very independent central bank and a topic leader in monetary union related issues. Surprisingly, we see that more independent banks devote less attention to the

Figure 7: Impact of macroeconomic and institutional factors on topic allocation

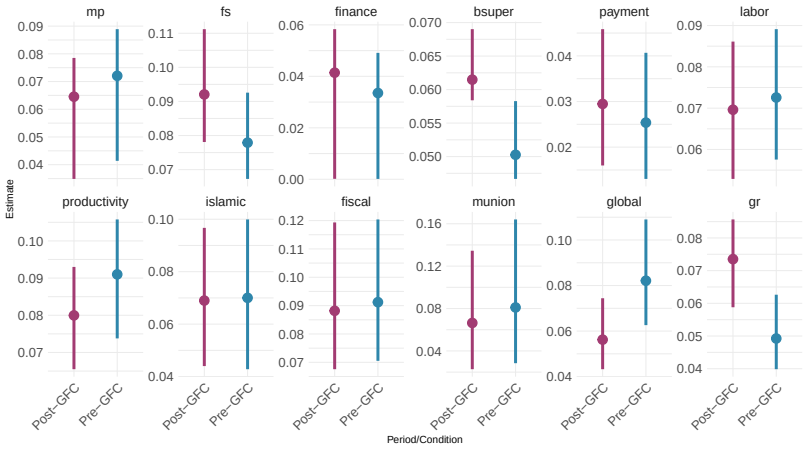
(a) High vs. low inflation regimes



(b) Central bank independence levels



(c) Pre- vs. post-Global Financial Crisis



Notes: Marginal posterior means and 90% credible intervals for document-topic distributions across different covariate values from the *keyATM* covariate model. Inflation regimes are defined by top/bottom percentiles of CPI inflation rates. Central Bank Independence (CBIE) scores from Romelli (2022) are similarly categorized into high/low levels. The Global Financial Crisis comparison contrasts periods before and after 2008. Results show systematic adaptation of communication strategies to macroeconomic conditions, institutional constraints, and crisis events, with significant shifts in topic emphasis across all dimensions analyzed.

global economy and also to the finance topic, which also includes issues related to financial inclusion - which is clearly a theme that is high on the agenda of many emerging markets central banks. The bottom panel of Figure 8 shows differences in topic prevalence for the pre- and post GFC period. Post-crisis communication shows substantially intensified emphasis on financial stability, banking supervision, and the green economy, whereas less emphasis is put on global coordination topics.

Finally, we present an analysis regarding the different communication formats shown in Figure 8.

The analysis across communication formats reveals stark differences in both topic intensity and distribution. Monetary policy is predominantly addressed in discussions and presentations of reports, and public explanations of monetary policy decisions through official statements. Productivity is also frequently covered in presentations and statements, while Islamic finance remains underrepresented in these formats. In interviews, there is a stronger emphasis on monetary policy, whereas less attention is devoted to topics such as Islamic finance, as well as technical subjects like banking supervision, payment infrastructure, and financial stability.

5 Conclusion

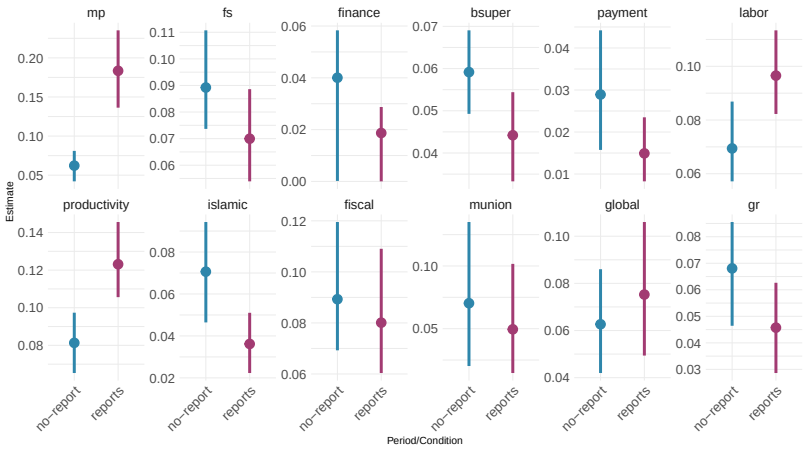
This paper provides the first large-scale benchmarking of classical and transformer-based topic models applied to nearly three decades of global central bank communication, drawing on over 20,000 speeches from the BIS archive. We evaluate three distinct modeling approaches—keyword-assisted LDA (keyATM), contextual document clustering (BERTopic), and hybrid transformer-topic modeling (FASTopic)—and assess their performance using semantic coherence, diversity, and interpretability metrics. Our results demonstrate that keyATM achieves the best balance between interpretability and thematic coverage, outperforming transformer-based models, which struggle to decompose long-form, multi-thematic speeches into distinct subtopics.

Methodologically, our study adds to the growing literature on transformer-based text analysis (Gambacorta et al., 2024; de Araujo et al., 2025; Hansen and Kazinnik, 2024), by demonstrating the continued relevance of structured probabilistic models in policy-relevant domains. While LLMs and contextual embeddings offer new possibilities in text understanding, our results caution against their direct application to document-level topic discovery in domains where interpretability, thematic separation, and compositional structure are crucial.

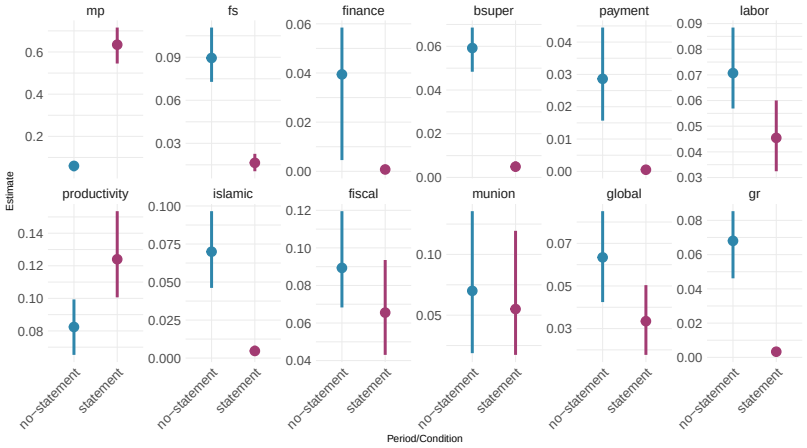
Our findings show that central bank communication is structured around a set of enduring policy "evergreens"—most notably monetary policy, financial stability, and banking supervision—that persist across institutions and time. Topics such as monetary union and global economic integration have declined in prominence, reflecting institutional consolidation and waning globalization, while others like digital payments and climate-related finance have recently emerged. These results align with recent literature showing the increasing diversification of central bank mandates and communicative scope in response to new policy challenges (see, in the context

Figure 8: Communication format effects on topic allocation

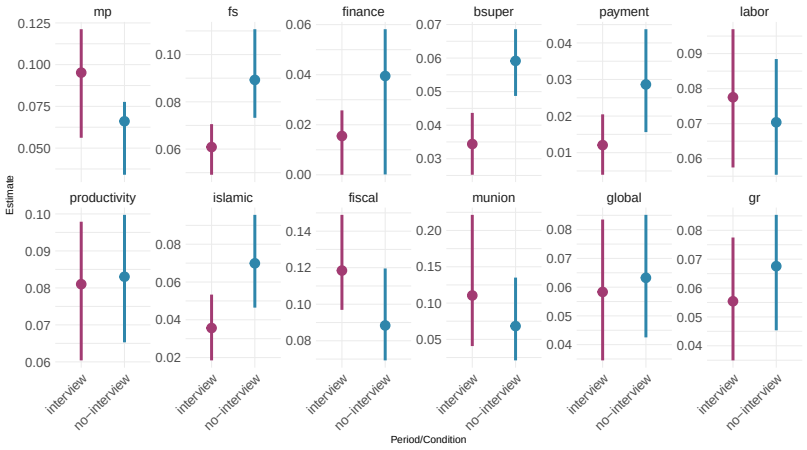
(a) Report presentations



(b) Official statements



(c) Media interviews



Notes: Marginal posterior means and 90% credible intervals showing how communication format affects topic allocation. Speech types were classified using keyword analysis of titles: reports include presentations and discussions of official documents; statements encompass policy announcements and formal declarations; interviews cover media interactions and Q&A sessions. The analysis reveals systematic differences in topic emphasis across formats, with monetary policy and productivity topics prominently featured in formal presentations and statements, while interviews show different emphasis patterns, particularly reduced attention to technical topics like banking supervision and payment systems.

of green central banking Dikau and Volz, 2021; Feldkircher and Teliha, 2024; Campiglio et al., 2025).

Our topic diffusion analysis reveals two complementary dimensions of central banking leadership. First, chronological topic pioneering shows a remarkably balanced innovation landscape, with emerging economy central banks establishing first-mover leadership in approximately half of all policy topics, challenging assumptions about advanced economy dominance in policy innovation. However, subsequent performance analysis demonstrates that advanced economies maintain advantages in achieving sustained topic leadership and superior emergence quality. This suggests that while innovation capability is globally distributed, the depth and persistence of policy emphasis reflects institutional capacity differences.

Most significantly, we document distinct patterns of policy attention synchronization during crisis periods. While stable leadership hierarchies persist under normal conditions—with the vast majority of top-performer emergences concentrated in the early sample period—systemic shocks trigger rapid attention convergence across the global central banking system. Our z-score analysis reveals a substantial concentration of unusual policy emphasis shifts during the 2008-2012 financial crisis, representing a marked increase in synchronization compared to baseline periods. This crisis-driven convergence particularly affected financial stability discourse, where all central banks simultaneously elevated emphasis relative to their historical patterns around 2008-2009. Similar event-driven coordination appears following international policy milestones, as evidenced by synchronized attention to green finance topics around 2016 following the Paris Agreement.

Finally, our covariate analysis further demonstrates that macroeconomic and institutional contexts shape the thematic focus of central bank communication. During periods of high inflation, central banks emphasize financial literacy topics more than monetary policy itself—perhaps to better equip households to navigate volatile financial environments. More independent central banks communicate more frequently on politically sensitive themes like monetary union, while devoting less attention to global coordination or financial inclusion. Communication format also matters: official statements and reports focus more on monetary policy and technical content, whereas media interviews emphasize broader themes and omit complex topics like banking supervision.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the author(s) used ChatGPT to improve language, expression, and grammar, and Claude to streamline code. After using these tools, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.

References

Ahrens, Maximilian and Michael McMahon, "Extracting Economic Signals from Central Bank Speeches," *Proceedings of the Third Workshop on Economics and Natural Language Processing*, 2021.

Araci, Dogu, "Finbert: Financial sentiment analysis with pre-trained language models," *arXiv preprint arXiv:1908.10063*, 2019.

Armelius, Hanna, Christoph Bertsch, Isaiah Hull, and Xin Zhang, "Spread the Word: International Spillovers from Central Bank Communication," *Journal of International Money and Finance*, 2020, *103*, 102–116.

Baerg, Nicole and Carola Binder, "Automated detection of emotion in central bank communication: A warning," *National Institute Economic Review*, 2024, *269*, 82–91.

Bholat, David, Nicholas Broughton, John Ter Meer, and David Walczak, "Enhancing Central Bank Communications: The Role of Summaries, Visuals, and Readability," Staff Working Paper 2015/08, Bank of England 2015.

Bianchi, Federico, Silvia Terragni, and Dirk Hovy, "Pre-training is a Hot Topic: Contextualized Document Embeddings Improve Topic Coherence," in "Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)" Association for Computational Linguistics 2021, pp. 759–766.

Blei, D. and J. Lafferty, "A Correlated Topic Model of Science," *The Annals of Applied Statistics*, 2007, *1*, 17–37.

Blei, David M. and John D. Lafferty, "Topic Models," in Anuj Srivastava and Mehran Sahami, eds., *Text Mining: Classification, Clustering and Applications*, Cambridge, UK: Chapman & Hall/CRC, 2009, pp. 71–93.

_ , Andrew Y. Ng, and Michael I. Jordan, "Latent Dirichlet Allocation," *The Journal of Machine Learning Research*, 2003, *3*, 993–1022.

Blinder, Alan S., Michael Ehrmann, Marcel Fratzscher, Jakob De Haan, and David-Jan Jansen, "Central Bank Communication and Monetary Policy: A Survey of Theory and Evidence," *Journal of Economic Literature*, December 2008, *46* (4), 910–45.

Bohl, Martin T., Dimitrios Kanelis, and Pierre L. Siklos, "Central bank mandates: How differences can influence the content and tone of central bank communication," *Journal of International Money and Finance*, 2023, *130*, 102752.

Bouma, Gerlof, "Normalized (Pointwise) Mutual Information in Collocation Extraction," in "Proceedings of the Biennial GSCL Conference" GSCL Tübingen, Germany 2009, pp. 31–40.

Campiglio, Emanuele, Jérôme Deyris, Davide Romelli, and Ginevra Scalisi, "Warning words in a warming world: central bank communication and climate change," Working Papers hal-04978072, HAL March 2025.

de Araujo, Daniel K. G., Nikola Bokan, Filippo A. Comazzi, and Michele Lenza, "Word2Prices: Embedding Central Bank Communications for Inflation Prediction," BIS Working Papers 1253, Bank for International Settlements 2025.

Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," *CoRR*, 2018, *abs/1810.04805.*

Deyris, Jérôme, "Too green to be true? Forging a climate consensus at the European Central Bank," *New Political Economy*, 2023, *28* (5), 713–730.

Diessner, Sebastian and Giulio Lisi, "Masters of the "masters of the universe"? Monetary, fiscal and financial dominance in the Eurozone," *Socio-Economic Review*, 2020, *18* (2), 315–335.

Dikau, Simon and Ulrich Volz, "Central bank mandates, sustainability objectives and the promotion of green finance," *Ecological Economics*, 2021, *184*, 107022.

Dunning, Ted, "Accurate methods for the statistics of surprise and coincidence," *Computational linguistics*, 1994, *19* (1), 61–74.

Eshima, Shusei, Kosuke Imai, and Tomoya Sasaki, "Keyword-Assisted Topic Models," *American Journal of Political Science*, 2024, *68* (2), 730–750.

Feldkircher, Martin and Viktoriya Teliha, "Speeches in the green: The political discourse of green central banking," *Energy Economics*, 2024, *135*, 107646.

_ , Paul Hofmarcher, and Pierre L. Siklos, "What do central banks talk about? A European perspective on central bank communication," *Focus on European Economic Integration*, 2021, *Q2/21*, 61–81.

_ , _ , and Pierre Siklos, "One Money, One Voice? Evaluating Ideological Positions of Euro Area Central Banks," *European Journal of Political Economy*, 2024, *85*, 102582.

Gambacorta, Leonardo, Byeungchun Kwon, Taejin Park, Pietro Patelli, and Sonya Zhu, "CB-LMs: language models for central banking," BIS Working Papers 1215, Bank for International Settlements, Monetary and Economic Department October 2024.

Grootendorst, Maarten, "BERTopic: Neural topic modeling with a class-based TF-IDF procedure," 2022.

Haldane, Andrew G. and Michael McMahon, "Central Bank Communications and the General Public," *AEA Papers and Proceedings*, 2018, *108*, 578–583.

Hansen, Stephen A. L. and Samuel Kazinnik, "Can ChatGPT Decipher Fedspeak?," Available at SSRN 2024.

Kardos, Márton, "Turftopic: Topic Modelling with Contextual Representations from Sentence Transformers," *Journal of Open Source Software*, 2025, *Forthcoming.*

Kim, Wonseong, Christina Niklaus, Choong Lyol Lee, and Siegfried Handschuh, "DisSim-FinBERT: Text Simplification for Core Message Extraction in Complex Financial Texts," 2025.

Kumar, Rohit, Sourabh Bikas Paul, and Nikita Singh, "Words that Move Markets- Quantifying the Impact of RBI's Monetary Policy Communications on Indian Financial Market," 2024.

Lau, Jey Han, David Newman, and Timothy Baldwin, "Machine Reading Tea Leaves: Automatically Evaluating Topic Coherence and Topic Model Quality," in Shuly Wintner, Sharon Goldwater, and Stefan Riezler, eds., *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics*, Association for Computational Linguistics Gothenburg, Sweden Apr 2014, pp. 530–539.

Leek, Lauren and Simeon Bischl, "How central bank independence shapes monetary policy communication: A Large Language Model application," *European Journal of Political Economy*, 2025, *87*, 102668.

Mimno, David and Andrew McCallum, "Topic models conditioned on arbitrary features with Dirichlet-multinomial regression," in "Proceedings of the Twenty-Fourth Conference on Uncertainty in Artificial Intelligence" UAI'08 AUAI Press Arlington, Virginia, USA 2008, p. 411–418.

_ , Hanna Wallach, Edmund Talley, Miriam Leenders, and Andrew McCallum, "Optimizing Semantic Coherence in Topic Models," in Regina Barzilay and Mark Johnson, eds., *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, Association for Computational Linguistics Edinburgh, Scotland, UK. July 2011, pp. 262–272.

Niţoi, Mihai, Maria-Miruna Pochea, and Ştefan Constantin Radu, "Unveiling the sentiment behind central bank narratives: A novel deep learning index," *Journal of Behavioral and Experimental Finance*, 2023, *38*, 100809.

Park, Jihye, Hye Jin Lee, and Sungzoon Cho, "Hot topic detection in central bankers' speeches," *Expert Systems with Applications*, 2023, *230*, 120563.

Roberts, Margaret E., Brandon M. Stewart, and Dustin Tingley, "Structural Topic Models for Open-Ended Survey Responses," *American Journal of Political Science*, 2014, *58* (4), 1064–1082.

Romelli, Davide, "The political economy of reforms in Central Bank design: evidence from a new dataset," *Economic Policy*, 01 2022, *37* (112), 641–688.

Scott, Mike and Christopher Tribble, *Textual patterns: Key words and corpus analysis in language education*, Vol. 22, John Benjamins Publishing, 2006.

Silva, Thiago Christiano, Kei Moriya, and Romain M. Veyrune, "From Text to Quantified Insights: A Large-Scale LLM Analysis of Central Bank Communication," IMF Working Paper 2025/109, International Monetary Fund 2025. Accessed June 9, 2025.

Wallach, Hanna M., "Topic Modeling: Beyond Bag-of-Words," in "Proceedings of the 23rd International Conference on Machine Learning" Pittsburgh, PA 2006, pp. 977–984.

Woodford, Michael, "Central Bank Communication and Policy Effectiveness," in "Proceedings - Economic Policy Symposium - Jackson Hole" Federal Reserve Bank of Kansas City 2005, pp. 399–474.

Wu, Xiaobao, Thong Nguyen, Delvin Ce Zhang, William Yang Wang, and Anh Tuan Luu, "FASTopic: Pretrained Transformer is a Fast, Adaptive, Stable, and Transferable Topic Model," 2024.

A Keyword selection for the keyATM

In this section, we briefly explain how we selected keywords for the seeded LDA model. All selected keywords for the *keyATM* models with $K = 10$ and $K = 15$ are presented in Table 3. Our approach two keyword selection proceeds in two steps. First, we select a list of general topics. For that purpose, we referred to central banks' mandates, following the approach outlined in Bohl et al. (2023). This rationale underpins the inclusion of topics such as monetary policy, financial stability, payment infrastructure, and banking supervision. In some jurisdictions, this also encompasses issues related to the labor market, the green economy (Dikau and Volz, 2021), financial inclusion, and Islamic finance.On top of that and owing to our global speech data base, we considered additional topics based on their recognized relevance to central bankers. For example, we included a topic on globalization, given that—particularly since the global financial crisis—there has been heightened attention to globalization and its implications for monetary policy and financial stability.[4] We also included fiscal policy, which typically falls outside the purview of central banks. Nonetheless, central bankers frequently discuss its ramifications, particularly with respect to the effects of interest rate changes on fiscal policy (for the ECB, see Diessner and Lisi, 2020).

Second, we use an iterative keyness analysis to come up with a list of keywords. More specifically, for each topic, we first select a representative speech and compute χ^2 keyness scores by comparing observed word frequencies contained in that representative speech against expectations from the remainder of the corpus. We then examine the speeches with the highest *keyATM* scores in initial runs, update the keyword lists based on their χ^2 keyness profiles, and repeat until convergence. This procedure follows the methodological precedents of Scott and Tribble (2006) and Dunning (1994) and has been applied in Feldkircher and Teliha (2024) for a similar corpus. The total number of seeded topics closely aligns with the count considered in Bohl et al. (2023). In addition to the seeded topics, we also accommodate a number of unseeded, residual topics. As the table shows, for $K = 10$, we use a smaller set of five core and mandate related topics and five unseeded topics. For $K = 15$, we use 13 pre-seeded topics and 2 residual topics. For both $K = 20$ and $K = 50$

B Robustness analysis: FASTopic

In this section, we present alternative results using the unsupervised FASTopic model with 15 topics, comparing its performance against KeyATM with 13 seeded topics and 2 residuals.

To ensure objective and consistent topic interpretation, we employ *GPT-4o-mini* for automated topic labeling. This LLM-based approach represents a methodological advancement that minimizes subjective bias inherent in manual labeling processes (Kardos, 2025). Following model training, we apply the automated labeling framework to generate coherent topic descriptions.

Figure 9 presents word clouds displaying the top 10 terms for each identified topic. The analysis

[4]For instance, the US Federal Reserve System hosts the Dallas Fed Globalization Institute.

Table 3: Keywords for topic identification in *keyATM* model

Topic	**Label**	**Keywords**	**K=10**	**K=15/20/50**
Monetary policy	1_mp	monetary_policy, interest_rate, price_stability, inflation_target, inflation_expectation	✓	✓
Financial stability	2_fs	financial_stability, financial_sector, financial_system, financial_market	✓	✓
Finance	3_finance	financial_institution, financial_inclusion, financial_product, financial_literacy	✓	✓
Banking supervision	4_bsuper	bank_system, balance_sheet, financial_service, risk_management, capital_requirement, credit_risk, bank_supervision, stress_test	✓	✓
Payment	5_payment	payment_service, payment_system, retail_payment, settlement_system		✓
Labor market	6_labor	unemployment_rate, labour_market, wage_growth		✓
Productivity	7_productivity	productivity_growth, economic_growth, gdp_growth		✓
Islamic finance	8_islamic	islamic_finance, islamic_financial, islamic_bank		✓
Fiscal policy	9_fiscal	fiscal_policy, public_finance, government_debt, sovereign_debt, budget_deficit		✓
Monetary union	10_munion	monetary_union, euro_area, member_state		✓
Global economy	11_global	exchange_rate, global_financial, capital_flow, global_economy, foreign_exchange, current_account	✓	✓
Green economy	12_gr	climate_change, climate_risk, impact_climate, climate_policy, green_finance, sustainable_finance		✓
US economy	13_federal	federal_reserve, governor_federal, reserve_system, federal_fund, fund_rate		✓

Notes: This table presents the keyword sets used to seed the *keyATM* model for topic identification across multiple model specifications. Keywords are domain-specific bigrams derived from central banking terminology and refined through iterative keyness analysis. Checkmarks (✓) indicate which topics are seeded with expert-selected keywords in each model specification. For K=10, five topics are seeded focusing on core central bank functions (topics 1-4) and global economic themes (topic 11), with five additional unseeded topics. For K=15, K=20, and K=50, all thirteen substantive topics shown are seeded with keywords covering the full range of central bank mandates and policy areas. The remaining topics in each specification (2, 7, and 37 topics respectively) are unseeded and capture additional thematic content without predetermined keywords.

reveals distinct thematic clusters: Topic 1 encompasses **demographic and economic development** concerns, featuring terms such as *population growth*, *living standards*, *social security*, and *labour productivity*. Topics 2 and 10 address **financial stability and banking regulation**, characterized by terminology including *capital requirements*, *Basel Committee*, *regulatory capital*, and *systemic risk*.

Monetary policy constitutes a central theme across multiple topics (3, 13, and 15), with Topic 13 specifically addressing **US monetary policy frameworks**. Key terms include *price stability*, *inflation expectations*, *federal reserve*, *deposit rates*, *money supply*, and *board governors*. Topic 4 represents a heterogeneous category of **economic policy and consumption dynamics**, encompassing diverse terms ranging from *private consumption* and *fixed investment* to *fresh food* and *overseas economy*—a complexity that would challenge traditional manual categorization.

Topics 5 and 8 concentrate on **eurozone monetary policy and economic governance**, featuring prominent terms such as *single currency*, *supervision authority*, and *stability growth*. Topic 6 focuses on **financial literacy and consumer protection**, emphasizing *financial education*, *financial inclusion*, and *consumer protection* initiatives.

Inflation and energy market dynamics characterize Topic 7, incorporating terms like *commodity prices*, *oil prices*, *high inflation*, and *house prices*. Topics 9 and 12 address **international trade and foreign exchange markets**, featuring terminology related to *trade investment*, *foreign investment*, *external debt*, *capital flows*, and *foreign exchange* operations.

Topic 11 captures the emerging domain of **digital finance**, including *digital currency*, *bank money*, *payment systems*, *crypto assets*, and *retail payments*. Finally, Topic 14 represents a hybrid category combining **climate finance and Islamic banking**, incorporating terms such as *climate risk*, *sustainable finance*, *Islamic banks*, and *Islamic financial* instruments.

Figure 10 demonstrates that the most specialized topics—**sustainable Islamic finance and climate risk** and **monetary policy dynamics**—represent less than 6% of total speech content, indicating their niche but significant policy relevance. The FASTopic model with LLM-based labeling produces topic categorizations largely comparable to KeyATM's 15-topic configuration. Notably, only two non-residual KeyATM topics—**labour market** and **productivity**—lack direct correspondence in the alternative model's output.

While automated topic labeling using large language models offers scalable solutions for topic interpretation, our findings suggest limitations for less coherent topic clusters. Specifically, topics with low internal consistency (e.g., the sustainable Islamic finance and climate risk cluster) may receive overly general or uninformative labels that limit their utility for granular policy analysis.

C Additional results

Figure 9: FASTopic Word Clouds for 15-Topic Model

Notes: This figure displays the top 10 bigrams for each topic in the 15-topic FASTopic model. Word size corresponds to term frequency within each topic cluster.

Table 4: Comparative analysis of BERTopic and FASTopic model performance

BERTopic 10

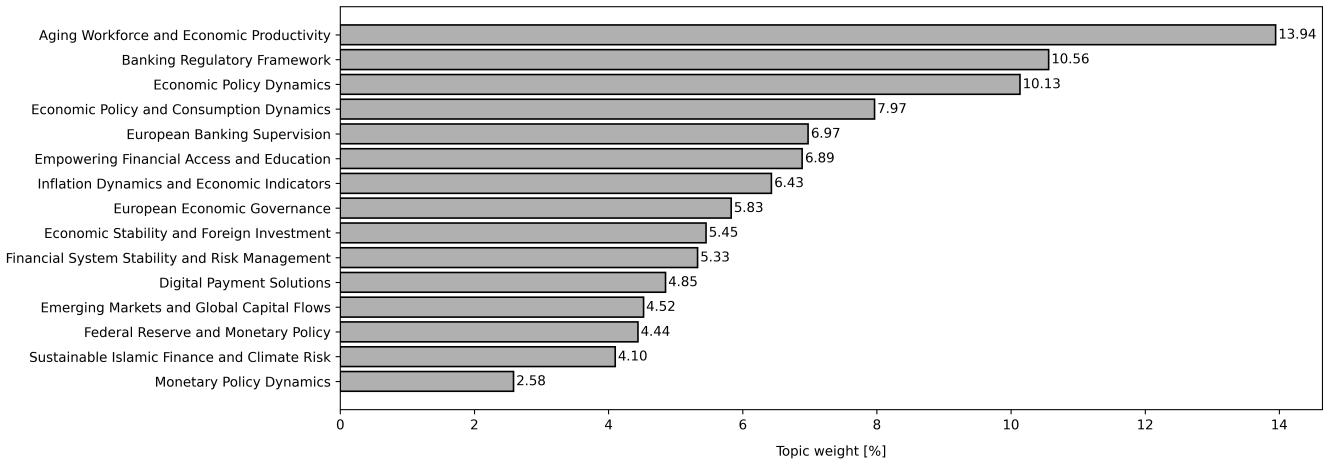
Rank	**Topic 1**	**Topic 2**	**Topic 3**	**Topic 4**	**Topic 5**	**Topic 6**	**Topic 7**	**Topic 8**	**Topic 9**	**Topic 10**
1	financial institution	monetary policy	monetary policy	digital currency	monetary policy	insurance industry	reference rate	population ageing	diversity inclusion	market participant
2	monetary policy	exchange rate	federal reserve	payment service	policy rate	insurance sector	rate benchmark	social security	gender diversity	fx market
3	financial market	price stability	exchange rate	retail payment	exchange rate	life insurance	market participant	pension fund	gender equality	global code
4	financial sector	financial stability	financial stability	federal reserve	inflation expectation	risk management	libor transition	demographic change	men woman	foreign exchange
5	financial service	financial market	financial market	financial service	foreign exchange	risk margin	riskfree rate	life expectancy	gender balance	exchange market
6	financial inclusion	fiscal policy	balance sheet	service provider	inflation forecast	insurance company	benchmark reform	retirement age	male female	fx global
7	economic growth	financial crisis	financial crisis	financial stability	inflation target	internal model	dollar libor	dependency ratio	diversity policy	wholesale fx
8	financial stability	governing council	inflation target	crossborder payment	domestic demand	insurance market	alternative reference	workingage population	labor force	principle code
9	exchange rate	member state	price stability	instant payment	economic activity	life insurer	financial benchmark	fertility rate	share woman	code conduct
10	economic activity	climate change	labour market	financial sector	price stability	insurance product	alternative rate	ageing population	career progression	adherence code

FASTopic 10

Rank	**Topic 1**	**Topic 2**	**Topic 3**	**Topic 4**	**Topic 5**	**Topic 6**	**Topic 7**	**Topic 8**	**Topic 9**	**Topic 10**
1	financial inclusion	payment service	service provider	financial literacy	retail payment	payment system	bank service	bank money	financial education	digital currency
2	climate change	sustainable finance	free trade	international trade	human capital	foreign direct	population age	public debt	living standard	small open
3	capital requirement	stress test	bank supervision	risk management	basel committee	basel iii	bank supervisor	risk bank	credit risk	climate risk
4	federal reserve	board governor	governor federal	reserve system	fund rate	federal fund	open market	labor market	house price	housing market
5	govern council	economic monetary	european commission	single currency	single market	european bank	structural reform	stability growth	sovereign debt	european level
6	monetary policy	price stability	inflation target	inflation expectation	policy rate	exchange rate	inflation rate	low inflation	fiscal policy	real rate
7	productivity growth	wage growth	energy price	core inflation	price inflation	labour market	high inflation	underlie inflation	real wage	inflation dynamic
8	domestic demand	emerge economy	growth rate	private consumption	monetary ease	global economic	account deficit	activity price	fix investment	economy continue
9	financial institution	financial sector	financial system	bank sector	financial stability	international financial	global financial	capital market	capital flow	cross border
10	islamic financial	islamic finance	islamic bank	financial centre	financial industry	money laundering	money launder	aml cft	corporate governance	anti money

Notes: Comparison of top 10 terms for each topic identified by BERTopic (10 topics) and FASTopic (10 topics) models.

Figure 10: Topic Representation in FASTopic Model



Notes: This figure presents the relative weights of each topic in the 15-topic FASTopic model, indicating the proportion of total corpus content attributed to each thematic cluster.

Table 5: Complete topic keyword analysis for *keyATM* model

	1_mp	**2_fs**	**3_finance**	**4_bsuper**
1	monetary_policy ✓	central_bank	reserve_bank	bank_system ✓
2	interest_rate ✓	financial_system ✓	financial_inclusion ✓	financial_institution$^{(3)}$
3	central_bank	financial_stability ✓	financial_sector$^{(2)}$	risk_management ✓
4	inflation_target ✓	financial_market ✓	financial_service$^{(4)}$	bank_sector
5	price_stability ✓	financial_crisis	commercial_bank	capital_requirement ✓
6	inflation_expectation ✓	financial_sector ✓	financial_product ✓	credit_risk ✓
7	economic_activity	market_participant	bank_india	bank_supervision ✓
8	policy_rate	global_financial$^{(11)}$	financial_literacy ✓	stress_test ✓
9	inflation_rate	money_market	financial_education	balance_sheet ✓
10	oil_price	systemic_risk	private_sector	large_bank
	5_payment	**6_labor**	**7_productivity**	**8_islamic**
1	central_bank	labour_market ✓	economic_growth ✓	financial_institution$^{(3)}$
2	payment_system ✓	house_price	gdp_growth ✓	financial_service$^{(4)}$
3	payment_service ✓	recent_year	growth_rate	financial_system$^{(2)}$
4	settlement_system ✓	past_year	financial_institution$^{(3)}$	islamic_finance ✓
5	retail_payment ✓	unemployment_rate ✓	last_year	capital_market
6	service_provider	productivity_growth$^{(7)}$	sustainable_growth	islamic_financial ✓

Continued on next page

Table 5 – continued from previous page

	1_mp	2_fs	3_finance	4_bsuper
7	digital_euro	percentage_point	per_caput	international_financial
8	bank_money	bank_canada	governor_bank	financial_centre
9	digital_currency	house_market	growth_potential	financial_industry
10	payment_settlement	bank_england	potential_growth	islamic_bank ✓
	9_fiscal	**10_munion**	**11_global**	**12_gr**
1	fiscal_policy ✓	central_bank	exchange_rate ✓	climate_change ✓
2	economic_policy	monetary_policy(1)	foreign_exchange ✓	financial_sector(2)
3	private_sector	price_stability(1)	current_account ✓	climate_risk ✓
4	public_finance ✓	monetary_union ✓	global_economy ✓	can_help
5	sovereign_debt ✓	member_state ✓	emerge_market	last_year
6	public_debt	european_central	capital_flow ✓	look_forward
7	long_term	govern_council	global_financial ✓	financial_risk
8	government_bond	bank_union	advance_economy	bank_ireland
9	public_sector	structural_reform	interest_rate(1)	work_together
10	low_interest	european_commission	foreign_currency	sustainable_finance ✓
	13_federal	**Other_1**	**Other_2**	
1	federal_reserve ✓	bank_can	interest_rate(1)	
2	reserve_bank	work_good	year_ago	
3	reserve_system ✓	long_time	can_make	
4	board_governor	many_year	decision_make	
5	governor_federal ✓	year_ago	even_though	
6	fund_rate ✓	general_public	public_policy	
7	labor_market	look_back	policy_committee	
8	federal_fund ✓	start_point	great_deal	
9	balance_sheet(4)	university_press	make_good	
10	work_paper	look_like	make_sure	

Notes: Complete topic analysis showing the top 10 keywords for each of the 15 topics identified by the *keyATM* model. Keywords marked with ✓ indicate terms from the original seed keyword sets, demonstrating successful topic identification. Superscripts (e.g., (4)) indicate cross-references to topics where identical terms appear, showing natural semantic overlap between related policy areas. The two residual topics (Other_1 and Other_2) capture procedural and conversational language patterns rather than substantive policy themes, confirming the model's ability to separate content-based from structural linguistic patterns.

Table 6: Representative speeches for each topic from *keyATM* analysis

	Document Title (Speaker)
Topic 1_mp: Monetary Policy	
1	The bank's semiannual report on currency and monetary control (Haruhiko Kuroda)
2	The bank's semiannual report on currency and monetary control (Haruhiko Kuroda)
3	ECB press conference - introductory statement (Mario Draghi)
4	The bank's semiannual report on currency and monetary control (Haruhiko Kuroda)
5	The bank's semiannual report on currency and monetary control (Kazuo Ueda)
Topic 2_fs: Financial Stability	
1	Sub-prime crisis, liquidity tensions and central banks - one year on (Jose Gonzalez-Paramo)
2	Macroprudential policy for non-bank financial intermediation (Luis de Guindos)
3	Reflections on the current financial market correction (Jean-Claude Trichet)
4	Reflections on the current financial markets correction (Jean-Claude Trichet)
5	Margins and haircuts as a macroprudential tool (Vitor Constancio)
Topic 3_finance: Finance and Financial Inclusion	
1	Strategy adopted for financial inclusion (Deepali Pant Joshi)
2	Financial sector development in Zambia (Caleb Fundanga)
3	Launch of the financial literacy project in Karnataka (V Sharma)
4	Financial inclusion - journey so far and road ahead (Deepali Pant Joshi)
5	Moving financial capability forward - innovation scale and impact (Deepali Pant Joshi)
Topic 4_bsuper: Banking Supervision	
1	Basel II (Susan Bies)
2	Capital and risk management (Susan Bies)
3	Basel II and commercial real estate (Susan Bies)
4	Implementing Basel II - choices and challenges (Susan Bies)
5	Addressing challenges raised by Basel II implementation (Susan Bies)
Topic 5_payment: Payment Systems	
1	Emission data - a mission for statistics (Fritzi Kohler-Geib)
2	Managing the known unknowns (John Williams)
3	The role of TIPS for the future payments landscape (Ignazio Visco)

Continued on next page

Table 6 – continued from previous page

	Document Title (Speaker)
4	The role of TIPS for the future payments landscape (Ignazio Visco)
5	The global code of conduct for the foreign exchange market (Guy Debelle)
Topic 6_labor: Labor Market	
1	The state of the labour market (Guy Debelle)
2	Employment and wages (Guy Debelle)
3	What's up (and down) with households? (Luci Ellis)
4	Cyclical and structural changes in the labour market (Christopher Kent)
5	The labour market and spare capacity (Philip Lowe)
Topic 7_productivity: Economic Growth and Productivity	
1	Energy (Alan Greenspan)
2	Energy (Alan Greenspan)
3	Economic activity, prices, and monetary policy in Japan (Toyoaki Nakamura)
4	Energy (Alan Greenspan)
5	Demographic changes and macroeconomic performance - Japanese experiences (Masaaki Shirakawa)
Topic 8_islamic: Islamic Finance	
1	The development of Islamic finance (Zeti Akhtar Aziz)
2	Fostering leadership role in Islamic finance (Zeti Akhtar Aziz)
3	Brief overview of the Islamic financial system (Zeti Akhtar Aziz)
4	The international dimension of Islamic finance (Zeti Akhtar Aziz)
5	Focus on human capital development in Malaysia (Zeti Akhtar Aziz)
Topic 9_fiscal: Fiscal Policy	
1	Greece - achievements, challenges, risks and a strategy for the future (Yannis Stournaras)
2	Financial crises - impact and challenges for Spain (Miguel Ordonez)
3	Financial crises - impact and challenges for Spain (Miguel Ordonez)
4	Greece - a comeback to the financial markets? Are we near the finishing line? (Yannis Stournaras)
5	The macroeconomic outlook for Spain in a global context and upcoming fiscal challenges (Pablo de Cos)
Topic 10_munion: Monetary Union	
1	The euro as one of the main aspects shaping European identity today (Jose Gonzalez-Paramo)

Continued on next page

Table 6 – continued from previous page

	Document Title (Speaker)
2	Maastricht and the future of Europe (Willem Duisenberg)
3	ECB and ESCB - ten-year anniversary (Jean-Claude Trichet)
4	Euro star event to raise public awareness and facilitate changeover to the euro in Estonia (Jean-Claude Trichet)
5	Address at the opening ceremony of the euro exhibition (Jose Gonzalez-Paramo)
Topic 11_global: Global Economy	
1	Developing countries and international capital markets (Philipp Hildebrand)
2	Strong Swiss franc and large current account surplus - a contradiction? (Thomas Jordan)
3	Global challenges, global solutions - some remarks (Duvvuri Subbarao)
4	The evolving international payments imbalance of the United States and its effect on Europe and the rest of... (Alan Greenspan)
5	Global imbalances, external sector vulnerability and capital flows - selected countries overview (Ana Mitreska)
Topic 12_gr: Green Economy	
1	Developing the ecosystem for energy transition (Ravi Menon)
2	Speech at economics conference "Creating pathways towards the green economy" (Gediminas Simkus)
3	Creating pathways towards the green economy (Gediminas Simkus)
4	Remarks at a panel to launch the third annual America's pledge report (Mark Carney)
5	Towards net zero in Asia (Ravi Menon)
Topic 13_federal: US Federal Reserve	
1	The Federal Reserve's review of its monetary policy strategy, tools, and communication practices (Richard Clarida)
2	The Federal Reserve's review of its monetary policy strategy, tools, and communication practices (Richard Clarida)
3	The Federal Reserve's review of its monetary policy strategy, tools, and communication practices (Richard Clarida)
4	The Federal Reserve's review of its monetary policy strategy, tools, and communication practices (Richard Clarida)
5	The Federal Reserve's review of its monetary policy strategy, tools, and communication practices (Richard Clarida)
Other_1: Procedural Language	

Continued on next page

Table 6 – continued from previous page

	Document Title (Speaker)
1	Scientific research - what researchers should do and should not (W Wijewardena)
2	The road to independence: a Manley perspective (Jwala Rambarran)
3	The road to independence - a Manley perspective (Jwala Rambarran)
4	Czechs and Austrians - even closer than they appear (Mojmir Hampl)
5	A fascinating train journey (Norman Chan)
Other_2: Conversational Content	
1	The resilience of free societies and the impact of free markets (Alan Greenspan)
2	The role and resilience of open and competitive markets (Alan Greenspan)
3	Address at the presentation of the 20th King of Spain prize in economics (Jose Escriva)
4	The ten suggestions (Ben Bernanke)
5	Trusting in money – from Kirkcaldy to the MPC (Mervyn King)

Notes: Representative speeches for each topic showing the five documents with highest topic loadings from the *keyATM* analysis. Topics are ranked by their prominence within each document, demonstrating the model's ability to identify speeches that best exemplify each thematic category. The selection includes speeches from major central bankers across different time periods and institutions, validating the model's capture of both temporal and institutional patterns in central bank communication. Residual topics (Other_1 and Other_2) contain speeches with primarily procedural or ceremonial content rather than substantive policy themes.