Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

Essays on Central Bank Communication

DISSERTATION

Zur Erlangung der Würde eines Doktors der
Wirtschafts- und Sozialwissenschaften
Doctor rerum politicarum
(gemäß der Promotionsordnung vom 18. Januar 2017)

vorgelegt von

Sami Diaf

Hamburg, den 7. Oktober 2026

Vorsitzender: Prof. Dr. Ole Wilms

Erstgutachter: Prof. Dr. Ulrich Fritsche

Zweitgutachter: Prof. Dr. Alexander Szimayer

Datum der Disputation: 7. Oktober 2026

Preface

The idea of writing this dissertation came while working as a Research Associate at the University of Hamburg. I joined the second phase of the SPP 1859 $Experience\text{-}Expectation\ Nexus$ project (2019-2022) with great enthusiasm, eager to investigate several macroeconomic hypotheses using text mining and machine learning techniques. This was the natural progression of the path I began few years ago while at the University of Hildesheim, studying advanced data analytics.

Throughout the project and beyond, I benefited from a fruitful exchange of ideas and an enlightening guidance from the project's Principal Investigator Prof. Dr. Ulrich Fritsche, whose broad expertise and sharp acumen helped me testing and improving many existing text-as-data algorithms as well as devising new research directions, particularly in central bank corpora and narrative economics.

I also thank my colleague Dr. Florian Schütze for the productive cooperation and continuous brainstorming in several aspects of text mining and economic analysis. This resulted in a rich collaboration, with several papers being presented at international conferences and later published, confirming the validity of concepts we used to enrich the literature in economic uncertainty and central bank multilinguality.

I am grateful to the support I received from the Faculty of Business, Economics and Social Sciences at the University of Hamburg, in particular Prof. Dr. Alexander Szimayer who kindly accepted to review this work and Prof. Dr. Ole Wilms as Chairman of the Jury.

I benefited from numerous remarks and observations given during research seminars by members of the Applied Economics Chair at the University of Hamburg. All of them are thanked here for their fruitful comments and discussions.

I dedicate this work to my family and people who kept encouraging me to push the limits further. In particular, to my Father who will surely appreciate this work.

Abstract

Technological innovations brought over the last three decades have transformed people's lives and perceptions around the world. The term digitization has been used to summarize all transformations that systematically changed the way people see and perceive, what was previously known and learned.

Digitization fostered computational innovations and constant improvements, through intensive research and developments. Economists, like other scientists, were impacted by this wave with a noticeable shift in the mainstream of modern economics. It triggered a modernization wave through constant incorporation of innovations from other disciplines.

Monetary policy was, for a long time, mostly taught at colleges and universities in a theoretical fashion. However, a little was said about its practice and communication. The $secrecy$ adopted by central bankers at that time proved to be harmful in times of crises, prompting them to become more open to explain monetary policy decisions. Central bank communication provided the opportunity to implement this, supported by a variety of documents made available to the public, such as minutes, transcripts, speeches and addresses of monetary committee members around the world.

Therefore, analyzing this new type of data became a necessity to understand policy practices and to discern the inner functioning of monetary committees. Hence, $text\ mining$ tools were first adopted to test the tonality of the speeches, then derive partisanship at the member-level to test differences between central bankers.

Earlier essays on central bank communication highlighted the challenges of handling such corpora with classic text mining techniques. They suggested advanced tools that later came under the name of $natural\ language\ processing$, which encompasses machine learning techniques and neural network architectures used to understand and model human speeches.

This thesis employs text mining techniques and natural language processing to test several hypotheses related to central bank communication. First, the traditional preference debate over interest rate and inflation was tackled using keyword-based topic models. They helped extracting granular topics, able to devise robust econometric inference for speaker-level preferences, based on their official addresses and contributions during internal meetings.

Second, multilinguality has been a key communication feature in multicultural countries. Testing the language-based information variations in central bank documents is crucial to uncover language-preferences related to specific topics of interest at monetary committees.

Third, the clarity of central bankers' addresses was investigated under the umbrella of textual ambiguity, which proved to be a stylized fact of modern monetary practices, along consensus. Ambiguity is also seen as a herding strategy that could explain the lack of differences in voting rolls, and was studied using a probabilistic framework.

Uncovering topics is probably the most widely used application by economists via topic models. The aim was not to straightforwardly apply machine learning algorithms but to derive new ones, able to shed more light on given addresses and unfold embedded information that cannot be uncovered via traditional techniques. Text covariates, for instance, could be used to successfully update the loss function in a macroeconometric model built for policy mix assessments.

A substantial effort was made to improve topic models, as a subclass of machine learning algorithms, and emphasize the interpretability of the results. Issues raised from multilingual corpora were resolved using advanced embedding techniques, whose deployment is a tour de force in modern text analytics. It permitted to tackle multilingual corpora and test cross-lingual hypotheses, in a way that could not be achieved using traditional monolingual tools.

As part of the widespread interest in narrative economics, network analysis was employed to enhance existing, and often agnostic, analytical methods. Highlighting the links between terms forming informative representations proved to be genuine in guiding the inference and reveal hidden causality. Several network structures, as for communities and cliques, helped augmenting applications and giving detailed results, matching the essence of monetary policy practice.

Finally, latest developments in NLP brought Large Language Models (LLMs) as a robust tool at the hand of practitioners dealing with complex tasks. Attempts to optimize existing LLMs for central bank corpora have focused on sentiment analysis and policy preferences, known to be a high-level task requiring strong linguistic and policy expertise. For this aim, elements from Random Matrix Theory were used to check the bona fide of LLMs trained or fine-tuned using central bank corpora. Subsequently, an experiment on uncertainty and tonality measurement was conducted using two general-purpose LLMs. Results were not found to be in-line with monetary policy surprises, and confirm the need to adopt an extensive, often expensive, strategy involving complex prompting and multitasking. However, informative hidden insights could be drawn from LLMs, thanks to their reasoning capabilities especially when investigating latent uncertainty buildup.

This thesis summarizes a journey that began with text mining techniques and gradually went through developments in text-as-data field before reaching large scale models. While the aim was to test several hypotheses related to monetary policy communication, the work frequently needed improvements in machine learning and network analysis to extract information more effectively and facilitate the interpretability of results, without neglecting their economic context.

Zusammenfassung

Die technologischen Innovationen der letzten drei Jahrzehnte haben das Leben und die Wahrnehmung der Menschen weltweit grundlegend verändert. Der Begriff „Digitalisierung" fasst all diese Transformationen zusammen, die die Art und Weise, wie Menschen das zuvor Bekannte und Gelernte sehen und wahrnehmen, dramatisch verändert haben.

Die Digitalisierung förderte computergestützte Innovationen sowie stetige Verbesserungen durch intensive Forschung und Entwicklung. Auch Ökonomen und andere Wissenschaftler waren von dieser Entwicklung betroffen. Ein deutlicher Wandel im modernen Mainstream der Wirtschaftswissenschaften führte zu einer Modernisierung dieser Disziplin durch die kontinuierliche Integration neuester Innovationen und Forschungsergebnisse.

Lange Zeit wurde die Geldpolitik an Hochschulen und Universitäten hauptsächlich theoretisch gelehrt, ihre Praxis und Kommunikation hingegen vernachlässigt. Die von Zentralbankern praktizierte Geheimhaltung erwies sich in Krisenzeiten jedoch als schädlich und veranlasste sie, ihre geldpolitischen Entscheidungen transparenter zu gestalten. Zu diesem Zweck bot die Kommunikation der Zentralbanken die Möglichkeit, diese Transparenz umzusetzen. Dies wurde durch eine Vielzahl öffentlich zugänglicher Dokumente unterstützt, darunter Protokolle, Transkripte, Reden und Ansprachen von Mitgliedern währungspolitischer Ausschüsse weltweit.

Um die geldpolitischen Praktiken zu verstehen und die internen Abläufe der währungspolitischen Ausschüsse zu ergründen, war die Analyse dieser neuen Datenart unerlässlich. Zunächst kamen Text-Mining-Tools zum Einsatz, um den Tonfall bzw. die Polarität der Reden zu analysieren. Anschließend wurde die Parteizugehörigkeit der einzelnen Mitglieder ermittelt, um Abweichungen zwischen den Zentralbankern aufzudecken.

In früheren Arbeiten zur Zentralbankkommunikation wurde auf die Schwierigkeiten bei der Verarbeitung solcher Korpora mit klassischen Text-Mining-Techniken hingewiesen. Es wurde auf die Notwendigkeit fortgeschrittenerer Methoden hingedeutet, die später unter dem Begriff „Natural Language Processing" (NLP) bekannt wurden. NLP umfasst verschiedene Techniken des maschinellen Lernens und neuronale Netzwerkarchitekturen, die zum Verständnis und zur Modellierung menschlicher Sprache eingesetzt werden.

In dieser Arbeit werden sowohl Text-Mining-Techniken als auch NLP verwendet, um verschiedene Hypothesen zur Zentralbankkommunikation zu überprüfen. Zunächst wurde die traditionelle Präferenzdebatte über Zinssätze und Inflation mithilfe schlüsselwortbasierter Themenmodelle untersucht. Das Ziel bestand darin, detaillierte Themen zu extrahieren, die es ermöglichen, robuste ökonometrische Schlussfolgerungen zu den Präferenzen einzelner Sprecher auf Basis ihrer offiziellen Reden und Beiträge in internen Sitzungen zu ziehen.

Zweitens wurde die Verständlichkeit der Reden von Zentralbankern im Hinblick auf textuelle Mehrdeutigkeit untersucht. Neben dem Konsensprinzip erwies sich diese als charakteristisches Merkmal moderner geldpolitischer Praktiken. Mehrdeutigkeit wird auch als Herdenverhalten betrachtet. Dieses könnte die fehlenden Unterschiede in den Abstimmungslisten erklären und wurde mithilfe eines probabilistischen Modells untersucht.

Drittens ist Mehrsprachigkeit ein zentraler Bestandteil der Kommunikation von Zentralbanken in multikulturellen Ländern und Währungsunionen. Die Prüfung des sprachbasierten Informationsgehalts ist daher entscheidend, um sprachliche Präferenzen in Bezug auf spezifische Themen von

Interesse in geldpolitischen Ausschüssen aufzudecken.

Die Identifizierung von Themen ist wahrscheinlich die am weitesten verbreitete Anwendung von Themenmodellen in der Ökonomie. Ziel war es nicht, maschinelles Lernen Algorithmen einfach anzuwenden, sondern neue zu entwickeln, die Reden besser verständlich machen und eingebettete Informationen aufdecken, die mit traditionellen Methoden nicht zugänglich sind. Textkovariaten könnten beispielsweise verwendet werden, um die Verlustfunktion in einem makroökonometrischen Modell zur Bewertung eines Politikmix zu aktualisieren.

Es wurden erhebliche Anstrengungen unternommen, um die Themenmodelle als Unterklasse der Algorithmen des maschinellen Lernens zu verbessern und die Interpretierbarkeit der Ergebnisse zu optimieren. Probleme, die bei der Analyse mehrsprachiger Korpora auftraten, konnten mithilfe fortschrittlicher Einbettungstechniken gelöst werden. Der Einsatz dieser Techniken stellte eine Meisterleistung in der modernen Textanalyse dar. Dadurch wurde die Bearbeitung mehrsprachiger Korpora und die Prüfung sprachübergreifender Hypothesen auf eine Weise ermöglicht, die mit traditionellen einsprachigen Werkzeugen nicht möglich war.

Im Zuge des breiten Interesses an narrativer Ökonomie wurde die Netzwerkanalyse eingesetzt, um bestehende, oft agnostische Analysemethoden zu verbessern. Die Hervorhebung der Verbindungen zwischen Begriffen, die informative Repräsentationen bilden, erwies sich als hilfreich für die Schlussfolgerungsführung und die Aufdeckung verborgener Kausalzusammenhänge. Verschiedene Netzwerkstrukturen, wie beispielsweise Communities und Cliques, trugen zur Erweiterung der Anwendungen und zur Bereitstellung detaillierter Ergebnisse bei, die dem Wesen der geldpolitischen Praxis entsprechen.

Schließlich haben sich jüngste Entwicklungen im Bereich NLP, die Sprachmodelle (LLMs) als robustes Werkzeug für Anwender komplexer Aufgaben etabliert. Bei den Optimierungsversuchen bestehender LLMs für Zentralbankkorpora lag der Fokus auf der Stimmungsanalyse und der Ermittlung politischer Präferenzen, die bekanntermaßen anspruchsvolle Aufgaben sind, für die fundierte linguistische und politikwissenschaftliche Expertise erforderlich ist. Zu diesem Zweck wurden Elemente der Zufallsmatrixtheorie herangezogen, um die Validität von LLMs zu überprüfen, die mit Zentralbankkorpora trainiert oder feinabgestimmt wurden. Anschließend wurde ein Experiment zur Messung von Unsicherheit und Tonalität mit zwei universellen LLMs durchgeführt. Die Ergebnisse stimmten nicht mit geldpolitischen Überraschungen überein, was die Notwendigkeit einer umfassenden, oft kostspieligen Strategie mit komplexen Prompts und Feinabstimmung bestätigt.

Diese Arbeit fasst einen Prozess zusammen, der mit Text-Mining-Techniken begann und sich schrittweise über die Entwicklung im Bereich „Text als Daten" bis hin zu groß angelegten Modellen erstreckte. Das Ziel bestand darin, verschiedene Hypothesen zur geldpolitischen Kommunikation zu testen. Oft waren jedoch Verbesserungen im Bereich des maschinellen Lernens und der Netzwerkanalyse erforderlich, um Informationen besser zu extrahieren und die Interpretierbarkeit der Ergebnisse zu verbessern, ohne dabei den ökonomischen Kontext zu vernachlässigen.

Contents

1 General Introduction 7
- 1.1 Central bank communication . . . 7
- 1.2 Text-as-data . . . 8
- 1.3 Text mining . . . 9
- 1.4 Natural language processing . . . 9

2 Scaling central bankers' policy preferences 11
- 2.1 Introduction . . . 11
- 2.2 Central bank communication . . . 14
- 2.3 Loss function . . . 15
- 2.4 Topic models . . . 15
- 2.5 Data and results . . . 16
  - 2.5.1 Federal Reserve governors' speeches . . . 16
  - 2.5.2 FOMC transcripts . . . 20
- 2.6 Individual preferences estimates . . . 20
- 2.7 Conclusion . . . 22

3 Multilinguality in central banking 23
- 3.1 Unfolding language and policy preferences of swiss central bankers . . . 23
  - 3.1.1 Introduction . . . 23
  - 3.1.2 Central banking communication . . . 24
  - 3.1.3 Topic models . . . 24
  - 3.1.4 Application . . . 25
  - 3.1.5 Conclusion . . . 27
- 3.2 Multilingual topic modeling on the Bundesbank . . . 30
  - 3.2.1 Introduction . . . 30
  - 3.2.2 Top2Vec . . . 30
  - 3.2.3 Bayesian factor analysis . . . 30
  - 3.2.4 Results . . . 31

4 Ambiguity in central bank communication 35
- 4.1 Introduction . . . 35
- 4.2 Central bank communication . . . 37
- 4.3 Ambiguity in text mining . . . 37
- 4.4 Application . . . 39
- 4.5 Conclusion . . . 40

5 Investigating uncertainty in central bank corpora 45
- 5.1 Introduction . . . 45
- 5.2 Methodology . . . 47
  - 5.2.1 Topic models . . . 47

5.2.2 Central bank communication 48
5.3 Data and results 48
5.4 Conclusion 53

6 Uncovering causality links in central bank corpora 55
6.1 Introduction 55
6.2 Text-as-data 57
6.3 Causal trees 58
6.4 Methodology 60
6.5 Application 60
6.5.1 Federal Reserve speeches 61
6.5.2 German parties' manifestos 61
6.6 Conclusion 62

7 LLMs in central bank communication 65
7.1 The use of LLMs in central bank communication: too deep to fail ? 65
7.1.1 Introduction 65
7.1.2 Random matrix theory 67
7.1.3 Central bank communication 69
7.1.4 Application 70
7.1.5 Conclusion 71
7.2 Investigating uncertainty using LLMs 73
7.2.1 Introduction 73
7.2.2 Uncertainty in text analysis 74
7.2.3 Uncertainty measurements 74
7.2.4 Uncovering uncertainty factors 75

8 A policy-mix DSGE for an oil exporting country 79
8.1 Introduction 79
8.2 Related literature 81
8.3 Description of the model 82
8.3.1 Households 82
8.3.2 Firms 83
8.3.3 Fiscal policy 85
8.3.4 Monetary policy 86
8.3.5 The rest of the world 87
8.4 Estimation and calibration 87
8.5 Results 87
8.6 Conclusion 90

9 Conclusion 93

A Bibliography 95

B Erklärung 109

List of Figures

2.1 Graphical model of the base KeyATM (Eshima et al., 2020). 16
2.2 Graphical model of the dynamic KeyATM (Eshima et al., 2020). 17
2.3 Proportion of keywords in the governors' speeches corpus. 18
2.4 Evolution of keyword topic scores (monthly averaged) in the FED speeches. 18
2.5 Evolution of non-keyword topic scores (monthly averaged) in the FED speeches. . . 19
2.6 Biplot of estimated inflation and unemployment individual preferences. 21

3.1 Crisis-scores per document and recession periods (in shaded pink). Blue curve represents the LOESS smoothed curve 28
3.2 Correspondence analysis between the built-in indices (Scores) and documents' languages. 28
3.3 Estimated Bundesbank executive board members' ideal points using $TopicShoal$ with projected topic contributions. 32

4.1 Estimated ambiguity scores over time (Line in blue is the Loess curve. 40

5.1 Aggregated monthly uncertainty index for the Federal Reserve System (Board of Governors). The blue line indicates LOESS smoothed curve. 52
5.2 Aggregated monthly world uncertainty index on the whole corpus. The blue line indicates LOESS smoothed curve. 52

6.1 Example of a dependency parsing output (tokenization, part of speech tagging and dependency relations) from a simple sentence, using UDPipe (Straka et al., 2016). . . 58
6.2 Synthetic examples of three causal patterns (McGowan et al., 2024), depending on the directions and the links between nodes. 59
6.3 Causality of some macroeconomic terms in the Federal Reserve governors' speeches corpus (1996-2024). Pre-defined words are colored in green, while other connecting nodes are in white. 63
6.4 Output of $Relatio$ using the same pre-defined terms as in Figure 6.3 63
6.5 Causality of economic-related words in the German parties' manifestos (1990-2021). Pre-defined words are colored in green, while other connecting nodes are in white. . 64
6.6 Causality of labor market-related words in the German parties' manifestos (1990-2021). Pre-defined words are colored in green, while other connecting nodes are in white. 64

7.1 Marchenko-Pastur distribution simulated 1,000 times on the correlation matrix of an initial random matrix with $\frac{T}{N}$=5 and $\sigma^2 = \frac{2}{3}$. α is the PL-exponent of the Marchenko-Pastur fit over the interval [λ_-,λ_+]. 68
7.2 PL metrics estimated from CentralBankRoBERTa model 71
7.3 PL metrics estimated from ECBERT-base. 71
7.4 Uncertainty estimates from the Federal Reserve minutes (2000-2025) 75
7.5 Tonality estimates from the Federal Reserve minutes (2000-2025) 76

7.6 Uncertainty estimates from the Federal Reserve minutes (2000-2025) and monetary policy surprises. 76
7.7 Uncertainty drivers (using OpenAI) from the Federal Reserve minute released on April 29^{th}, 2020 77

8.1 Impulse response functions following a 10% drop in international oil prices. 92

List of Tables

2.1 Keywords used to label monetary policy goals for both corpora. 17
2.2 Top 20 words of KeyATM for governors' speeches (keywords have ticks and numbers in brackets indicates keywords of specified topics). 19
2.3 Top 20 words of KeyATM for FOMC transcripts (keywords have ticks and numbers in brackets indicates keywords of specified topics). 20
2.4 Governors' inflation and unemployment preferences estimated using FED governors' speeches and FOMC transcripts. 22

3.1 Top 20 words of the learned topics using $Top2vec$. 26
3.2 Top 20 terms related to "uncertainty", and their correlation given by the cosine similarity of their respective vectors. 26
3.3 Top 20 terms related to "crisis", and their correlation given by the cosine similarity of their respective vectors. 27
3.4 Correlations between different built-in indices (variables are cosine similarity scores between the terms and each document of the corpus. 27
3.5 Top 10 words of the topics learned by $Top2Vec$ on the Bundesbank speeches corpus. 31
3.6 Estimated topic intensity β_i using TopicShoal on Bundesbank executive board members' corpus. 32

4.1 Estimated ambiguity average and variance per speaker. 39
4.2 List of communities learned using $CommunityFish$ algorithm. 43

5.1 Top 20 topics learned by $Top2Vec$. 49
5.2 Top 20 topics correlated with the term $uncertainty$. 50
5.3 Top 20 terms related to $uncertainty$, $uncertainty\ and\ knightian$ and $risk$, and their correlations given by the cosine similarity. 50
5.4 Results of a plain LDA showing the learned 10 topics and their most relevant terms. 51

7.1 Correlation matrix of uncertainty and tonality outputs with monetary policy surprises estimates. 75

8.1 Calibration of the DSGE model's parameters. 88
8.2 Calibration of the DSGE loss function. Weights are estimated from their respective frequencies in the IMF Article IV releases (2002-2018). 89
8.3 Normal and text-based DSGE loss functions (as the sum of estimated variances) . . . 89

Chapter 1

General Introduction

The world witnessed during the second half of the twentieth century an unprecedented wave of technological innovations and sophistication that brought constant improvements to almost every aspect of daily life. Similar to the industrialization that spread during last decades of the seventeenth century, innovation started first with rise of electronics and the gradual adoption of machines as powerful computing tools, which paved the way to automation and later an accelerated digitization.

Machines, at their turn, underwent dramatic changes and went from being fully agnostic task-scheduled to reliable, cost-effective with human-like reasoning. Their current capabilities exceeds, for many tasks, human capabilities.

The current artificial intelligence (AI) trend revolutionized our daily routines and impacted the research community which relies more and more on advanced computing efficiencies and the use of what is known as neural networks.

In social sciences, particularly in economics, the adoption of AI was the natural continuation of previous works that harnessed the growing amount of data in several fields of applications. The field of macroeconomic analysis was among the first to make a transition from statistics to computer sciences to improve the analytics on several macroeconomic fields of research, particularly monetary policy and central banking.

1.1 Central bank communication

Central banking, as a subfield of monetary policy, was long dominated by secrecy and central bankers were adamant when it comes to explain decisions taken by monetary policy committees around the world. Goodfriend (1986) coined the term $monetary\ mystique$ to these practices which put central bankers far from the public in general, and from macroeconomic researchers and other linguistics in particular.

The adoption of transparency at central banks was a cornerstone in modern monetary policy. Official sought ways to enhance the credibility of monetary institutions, their effectiveness and reinforce their accountability (Macklem and Vardy, 2023). A clear central bank communication is an advantage to make policies more understandable by the public to cement their trust and reinforce the legitimacy of central banks, especially during episodes of crises and turmoil. The wide adoption of inflation targeting in the 1990s by several central banks is a basic example of an explicit monetary policy, that brought clarity to missions and objectives.

The release of addresses, minutes and other verbatim proved to be an important step toward those goals and available documents' were seen as an information goldmine for economists trying to test several hypotheses regarding the impact on economic agents, or simply to understand the inner functioning of the decision process at central banks.

Researchers were given practical opportunities to test the $bona\ fide$ of several text analytics tools. The focus went primary to the central bank-financial market nexus and potential impact of monetary decision on economic agents. This exercise of information extraction went hand-in-hand with the constant improvements of computational text analysis and the adoption of sophisticated statistical and machine learning algorithms for text data.

However, this enthusiasm was shortly curbed, as difficulties in text analysis were raised by researchers, particularly in linguistics. Even if central bankers tried to inform the public about important items, their messages proved to be not as informative as it was thought to be. Meade (2005) pointed out a consensus attitude at monetary policy committees, hindering an individual understanding of central bankers' behaviors. Later, Baerg (2020) confirmed ambiguity as being a stylized fact in modern central bank communication, stemming from uncertainty which is considered as a forward-guidance tool in today's monetary practice (Greenspan, 2004). In other terms, secrecy changed its form from being institution-based to being speaker-based.

In addition to linguistic features, the automatic use of many text mining methods proved to be harmful and lead to questionable results. This pushed researchers to proceed by post-hoc adjustments (Eshima et al., 2023) to put the results in-line with the real facts. In particular, such shortcomings stem from bag-of-word, agnostic applications that could be improved by many semi-supervised strategy. These are mostly the core applications, the reader will learn in the next chapters.

1.2 Text-as-data

Text-as-data denotes the whole ecosystem built around data and the applications used to extract textual information. Earlier approaches were limited to simple word counts and laid down the foundations for the statistical learning, as a mathematical formalism used to seek solutions to many real-life problems.

First text mining approaches used latent semantic analysis (Deerwester et al., 1990) to artificially uncover topics via agnostic matrix factorization. Computing advances helped shifting the field of text mining from statistics to machine learning. Blei et al. (2003) set up the popular Latent Dirichlet Allocation (LDA) as the workhorse and the standard go-to solution for almost all text data. It models the document-word observable linkage via the creation of hidden variables as proxies of topics. Here, a topic is informally defined as a probability distribution over words, rather than a group of words having a semantic meaning (Mcauliffe and Blei, 2008).

The need of interpretable results pushed researchers to adopt neural networks as the basis of what is known as Natural Language Processing (NLP). This emerging field of research comprises all tasks, whose execution needs neural-based architectures. Examples are semantic role labeling and dependency parsing to name a few.

1.3 Text mining

Researchers were first interested in the word frequencies and used a probabilistic approach to identify most important words in a collection of documents. Elements of statistical learning were used to uncover synthetic themes, or topics, able to categorize documents based on learned themes. This field of research began at the end of 1980s when devising Latent Semantic Analysis (Deerwester et al., 1990), as a single-value-decomposition of the document-term-matrix.

Perhaps the most known and popularized method in this subclass of text-as-data is the Latent Dirichlet Allocation (LDA) (Blei et al., 2003), which stands for a mixed-membership algorithm that learns hidden structures, topics, able to explain the duality documents-words. From this milestone, several other extensions were created to remedy shortcomings of LDA as for dynamic corpora, semantic structure, attention given to specific terms and other guided/supervised inference.

Aside from uncovering topics, which has been the workhorse of many academic projects, researchers tried to rank or scale parties or speakers based on word frequencies. This direction, known as scaling techniques, permits to uncover unidimensional or multimensional scales for comparative studies using either probabilistic or non-probabilistic methods.

1.4 Natural language processing

The advent of neural networks at the end of 1980s and its exponential development in early 2010s permitted researchers to save time in complex computational tasks and put into practice ideas that turned out to be decisive in modern text analytics.

Generally speaking, most NLP solutions feature a neural architecture or call built-in frameworks on the basis of an existing neural linkage, given as a pre-trained model. The tour de force in this subfield was the ability to represent terms via vectors (Mikolov et al., 2013), known as word embeddings. The embedding scheme is simply assigning dense vectors to words, based on their occurrences in the trained documents. In other terms, word embeddings is the process to estimate similarity in context, but not similarity in meaning. For instance, the words $European$ and $Union$ have a high correlation in their respective word vectors, simply because the words occur together forming the entity $European\ Union$, but does not necessarily means that $European$ is close in meaning to $Union$.

The upcoming chapters feature detailed descriptions of the text mining and NLP algorithms used to study central bank corpora. Theoretical concepts are given to support the choice of the algorithms/methods used for a given task, later reinforced by practical comparisons with classic methods used by researchers.

Chapters 2 and 4 use a mix of scaling techniques and topic models, for the sake of interpretability, while Chapter 8 uses simple word-frequency estimates for the weight loss function used to compute policy parameters in a DSGE model.

Chapters 3 and 5 leverage NLP applications for multilingual central bank and uncertainty measurement using semantic search models (Angelov, 2020). This new direction in topic models is built on clustering documents and words in the same embedding space, later partitioned to find locally-based topics. This proved to be genuine in testing the semantic color of uncertainty and its difference from other related terms as for risk and crisis.

Chapter 6 focuses on the revealing causal links between specific terms using a dual approach based on language morphologies and simultaneous equation models. Because the use of dependency parsing calls pre-trained, neural-based models (Straka et al., 2016), it falls into the category of NLP.

Chapter 7 uses latest developments in testing neural networks and call elements from Random Matrix Theory (Martin et al., 2021) to evaluate two LLMs developed for central bank corpora. The aim is to uncover potential fitting issues that alter their predictive potential and demonstrate their reasoning strength when extracting latent uncertainty factors.

Chapter 2

Scaling central bankers' policy preferences[1]

Abstract

This work analyzes central bank information flow and proposes a novel strategy to estimate individual-level preferences, toward monetary policy objectives, by quantifying the narratives within its different communication channels for the case of the United States. While most of the text-as-data literature related to central banking used unsupervised topic models to quantify narrative signals, I propose a semi-supervised, keyword-based approach built upon groups of words linked to monetary objectives in order to get a coherent, dynamic estimation of topics' prevalence, whose scores could determine individual policy preferences for inflation and unemployment rates. The corpus of Federal Reserve governors' speeches (1996-2020) identified three non-keyword topics matching financial stability, financial innovation and the banking regulation, whose dynamics follow the chairman's tenure, considered as informative policy signals toward financial markets. Governors' preferences toward monetary policy objectives were better estimated using FOMC transcripts (1994-2016) whose narratives strictly match monetary policy practices and help ranking members on a partisanship scale, with a spectrum linked to the members' educational background. Though released with a five-year delay, the FOMC transcripts, as a proxy of internal communication, offer a better picture of the partisanship prevailing within monetary policy committees in the United States, that cannot be learned from Governors' addresses, but remain unable to capture non-conventional, but not less important topics.

Credits

Supported by the Deutsche Forschungsgemeinschaft (DFG), SPP 1859, Project: Exploring the experience-expectation nexus in macroeconomic forecasting using computational text analysis and machine learning.

2.1 Introduction

Narrative economics has gained the interest of researchers in both economics and linguistics as a considerable amount of documents has been released and made public, allowing to test hypotheses with the help of a variety of tools and techniques applied to text data. Monetary policy, as a pillar of the economic landscape, remains an important field of application where narratives help understanding central banking practices, enhance their predictability and gauge their potential impact on financial markets (Blinder et al., 2008).

[1] Paper published in the WiSo-HH Working Paper Series, number 69, November 2022. https://www.wiso.uni-hamburg.de/forschung/forschungslabor/working-papers/pdfs-wp/wp69.pdf

The seminal paper of Shiller (2017) confirmed this shift and drove the interest of researchers to explore more economic corpora in order to extract relevant information and use them as a complement to existing economic data. Such directions have been previously, at different levels, implemented for the case of monetary policy where the literature targeted qualitative aspects of central banking communication. Released text documents related to central banks' scope were investigated using a variety of techniques borrowed from statistics and machine learning fields, with a distinctive line of research that specializes on financial markets responses to this communication (Gentzkow et al., 2019a).

The 1990s marked a turnaround in monetary policy practices with the gradual adoption of transparency in central banking policies, instead of secrecy that was prevailing for a long time as stated by Blinder et al. (2008), considering that independent central banks should be more accountable. This effort came as an answer to earlier public misunderstanding of decisions made by monetary committees and their potential negative impacts on the economy (Goodfriend, 1986). Hence, the availability of monetary committees' statements and internal meetings constitutes a reliable source to embark in this line of research, although several contributions revealed a certain consensus or cacophony among central bankers that clearly emerged from these corpora, as a noise altering its informative power (Vissing-Jorgensen, 2019).

Today, central bank communication is seen as an informative channel of the monetary policy and a powerful tool for its forward guidance, although the use of unconventional instruments brought challenges about the effectiveness of communication at the European Central Bank (ECB) as well as at the FED, as reported by Whelan (2018), who highlighted communication issues related to modern central banking and cited examples of quantitative easing and forward guidance on policy rates.

The abundant literature related to central bank corpora demonstrated a keen interest on extracting narratives for storytelling and building indices as to define a hawkish/dovish scale of central bankers, based on different assumptions (Baerg and Lowe, 2020) that usually combine topic models and econometric inferences.

Noticeable is that central banking corpora are embedded with characteristics, or $stylized\ facts$, making the use of traditional text analysis tools biased, as for consensus among members (Meade, 2005) and ambiguity, whose occurrence in the speeches may be seen as an adopted communication strategy (Baerg, 2020). This makes the application of specific sentiment analysis and usual bag-of-word methods questionable, because messages tend to be neutral in tonality with less informative power in comparison with other corpora. Further individual characteristics of central bankers were found, in the case of the United States, to have a potential impact on dissenting votes at monetary policy meetings (Lähner, 2018).

While modern monetary policy practices around the world deal with specific objectives assigned to central banks, empirical works overlooked this feature and did not tailor the applications to take into account the importance of policy goals in the narratives, as for unemployment and inflation rates. Moreover, the used methods to study central banking corpora mostly belong to the bag-of-word family and are seen as naive applications, regarding the complexity and the specifications required when the discourse is ambiguous, not easily readable and understand by the public (Haldane and McMahon, 2018). The dominant technology used by researchers in automated content analysis has been probabilistic topic models, particularly latent Dirichlet allocation (LDA), which helps uncovering groups of words occurring together known as topics (Blei et al., 2003). Such heuristic methods usually fail to deliver coherent topic structures and were criticized for their sensitivity to starting values, inability to explore sequential corpora (Boyd-Graber et al., 2014), and requiring post-hoc adjustments of the learned topics (Eshima et al., 2020).

To the best of my knowledge, there was no published work that quantified central banking narratives, at the individual level, toward monetary policy objectives. The existing literature described attempts to either rank central bankers on an ideological scale based on hybrid approaches (Baerg and Lowe, 2020; Shapiro and Wilson, 2019), or using external corpora from newspaper and magazines (Istrefi, 2017) that are often embedded with information bias. This work proposes a novel strategy to explore central bankers' partisanship and preferences through quantifying monetary narratives in the United States by uncovering policy signals using the corpus of governors' speeches (1996-2020) and FOMC transcripts (1994-2016). I argue that both corpora offer a complete picture of the duality internal/external central banking communication as speeches have an external narrative purpose while transcripts, although available on a five-year delay, help understanding the inner debate between members and hence, their ideological preferences.

I adopt a semi-supervised approach, namely Keyword-Assisted Topic Model (KeyATM), implemented by Eshima et al. (2020), so that each monetary policy objective will be assigned a number of keywords that enhance the resulting topics' coherence, even if it remains a bag-of-word method. This strategy is convenient to estimate dynamic preferences of central bankers' and track the emphasize they gave during their tenures, to easily quantify their preferences with respect of the prevailing economic conjuncture and internal debates. Unlike usual topic models, keywords are set as an attention mechanism, and not as a weighting scheme, so to build meaningful groups of words matching the hypotheses to be tested. In other terms, keywords indicate the topic model which direction to take without biasing the sampling process, yielding robust topic scores that could be used to estimate stable, ideal points toward monetary policy objectives.

Results of KeyATM were used to estimate inflation and unemployment preferences for each governor in both corpora (FED speeches and FOMC transcripts) to make an effective comparison of the estimates, using an asymmetric loss function adopted by Shapiro and Wilson (2019), and investigate potential partisanship among members at the internal and external communication levels. The resulting topic prevalence scores were used to estimate unbiased FED governors' inflation and unemployment preferences or $ideal\ points$ that proved to be more scattered at the FED level, rather than at FOMC meetings. A particular consensus of long-term members around the chairman during Greenspan and Bernanke tenures (Lähner, 2018) was found, in addition of highlighting an overall preference gap for unemployment at 6%.

The analysis of governors' speeches confirm the dominance of unemployment and inflation topics during the periods (1996-2006) and (2012-2020) respectively, with a high variability over time and across members, depending on the prevailing economic conjuncture and the chairman's mandate. The interest rate topic was debated with force during and after the 2008 financial turmoil. Moreover, non-keyword topics were found to match second level, but not less important topics as for financial stability, financial innovation and the banking regulation, whose occurrences are linked to the 2007-2008 financial crisis. FOMC transcripts provide a more accurate estimates of individual objective preferences compared to the speeches, because of the inner debate structure that differentiates members.

The paper is organized as follows: the introduction will be followed by section 2 that gives a literature overview about central banking communication and topic models, and section 3 gives details about the estimation and results.

2.2 Central bank communication

Woodford (2005) cited three types of intertwined signals contained in central banking communication: macro signals about the outlook, policy signals about the potential policy path and public signals directed to private agents to coordinate when prices are strategic complements and agents seek to coordinate. Blinder et al. (2008) noticed the relatively high impact of short-run central banking communication (official statements, reports and minutes) on financial markets, which tends to create news, while a better central banking communication helps reducing noise.

Noticeable is that most of the central banking communication literature targeted, with the help of automated text analysis tools (Gentzkow et al., 2019a), the impact of text releases on financial markets or the predictability of interest rates (Hubert, 2017) with an explicit focus on qualitative communication produced by central banks, aimed to achieve policy objectives (manage expectations) by what Blinder et al. (2008) qualified as *creating news* and *reducing noise*. Other attempts (Haldane and McMahon, 2018) measured readability of central banking publications using linguistic complexity metrics and found a relatively higher public accessibility than political speeches and suggested the use of more innovative channels to reach a higher audience.

For the case of the United States, the Congress assigned the Federal Reserve the objective to conduct the monetary policy to support the goals of maximum employment, stable prices, and moderate long-term interest rates (Federal Reserve, 2020). The secrecy around central banks (Goodfriend, 1986) was a trademark practice until the adoption of a transparency guidance aimed at reaching a wider audience and reduce economic uncertainty.

FOMC transcripts were scrutinized with modern text mining tools to find informative patterns (Hansen et al., 2017) that might explain several longstanding facts about its inner functioning, or to confirm ideological stances between members (McGregor, 1996) and voting preferences were identified by Havrilesky and Gildea (1991) as stemming from three potential sources: training background/career experience, state-of-the-economy measures and partisanship measures.

Attempts to estimate policy preferences among central bankers used mostly voting data of policy rate, as for Eijffinger et al. (2013) who constructed a hawkish-dovish scale for five east-European central bankers or Hix et al. (2010) for the Bank of England's Monetary Policy Committee (MPC). Bennani et al. (2019) examined European Central Bank's ad-hoc communication and constructed an informative index to predict future monetary policy movements. In a similar way, Andersson et al. (2006) found the published speeches at the Riksbank to be an important determinant for long-term interest rates, than inflation reports, while Lähner (2018) studied dissenting voting behavior of FOMC members and concluded that the decision of releasing FOMC transcripts to the public in 1993 had a significant impact on voting habits, with long-serving members not inclined to cast negative votes, in comparison with new members.

In discerning central bankers' intonation to assigned goals and objectives, Masayuki and Yosuke (2019) compared corpora of two Bank of Japan governors and found one governor's strategy as *delphic*, with semantic ambiguity. Istrefi (2017) constructed an index to rank FOMC members in hawkish-dovish scale, based on human reading of 20,000 narrative records in US newspapers from 1960 until 2015 and concluded that 70% of FOMC members exhibited policy preferences in both directions, while the remaining are considered as swingers. Shapiro and Wilson (2019) gauged negative sentiment indices in the FOMC corpus to estimate a central bank objective function that found a lower implicit inflation target (1.5) than the commonly-assumed value (2).

Hartmann and Smets (2018) used topic models to investigate the prevailing topics in speeches of

ECB Board members during the period 1999–2017, while Edison and Carcel (2020) used a similar approach to study topics' evolution related to FOMC statements during the period 2003-2012 and suggested to identify topic priorities within the corpus. In a similar methodology, Baerg and Lowe (2020) proposed a strategy combining topic models and document scaling to estimate a policy preference, built upon the Taylor rule, based on a FOMC meeting transcripts (2005-2008) that permitted to rank its members on a hawkish-dovish scale, similar to McGregor (1996) who identified *partisan ideology* and *partisan loyalty* as two sources fueling governors' voting behavior at the FOMC meetings, conditional to a controlled state of the economy.

2.3 Loss function

Shapiro and Wilson (2019) adopted an asymmetric loss function between inflation gap and the loss to estimate inflation preferences (inflation target) at the Federal Reserve, requiring a non-linear estimation on a linear specification of the loss function.

$$\alpha_{it} = \delta_1(\pi_t - \pi_t^\star) \times \mathbb{1}[\pi_t > \pi_t^\star] + \delta_2(\pi_t - \pi_t^\star) \times \mathbb{1}[\pi_t \leq \pi_t^\star] + \mathrm{X}_t\Lambda + \epsilon_{it} \tag{2.1}$$

where X_t is a vector if non-inflation variables, α_{ij} is the inflation topic score estimated with a topic model for the speaker i at the time t and $\pi^\star$ is the inflation preference to be estimated, along δ_1 and δ_2 using a grid search approach which tests all possible values from 0 to 3 (with a stepsize of 0.1) and selects the one that minimizes the sum of squared residuals (SSR) using an ordinary least squares estimation (Shapiro and Wilson, 2019).

2.4 Topic models

Most of text-as-data applications in social sciences are built on automated content analysis (Gentzkow et al., 2019a), namely the latent Dirichlet allocation (LDA) developed by Blei et al. (2003), to extract features and use them as explanatory variables for further inferences. LDA is an unsupervised, mixed-membership Bayesian model that learns latent features constituted of words, called topics, assuming documents being mixtures of topics, which are in their turn mixtures of words drawn from a Dirichlet distribution. Despite several limitations, LDA remains the go-to algorithms when dealing with text data, although it often requires post-hoc adjustments (Eshima et al., 2020). Several extensions were later proposed to deal with specific cases (Boyd-Graber et al., 2014) as to consider sequences, hierarchies and sentiments in the corpus.

Wilson and Chew (2010) improved the precision of LDA models by adopting a weighting scheme for terms, preventing highly frequent words from dominating the resulting topics and Eshima et al. (2020) generalized the concept of keyword topic models, first established by Jagarlamudi et al. (2012) based upon seeded LDA (Lu et al., 2011). In this line of research, groups of words are set as labels to highlight the interest of the study (keyword topics) along with other topics (non-keyword topics) that could be learned in parallel. This strategy, known as KeyATM, differs from the standard post-hoc topic transformations that might alter the scientific objectivity (Eshima et al., 2020), and was found to outperform existing topic models used so far in practice. Moreover, the designed KeyATM incorporates covariates and models time trends, which helps tailoring the results and confront them with metadata.

The base KeyATM assumes D documents in the corpus, each document having total of N_d words, where w_{di} represents the i^{th} word appearing in the document d. A total of K topics will be learned and $\tilde{K}$ represents the number of keyword topics ($\tilde{K} \leq K$), where each keyword topic k has L_k

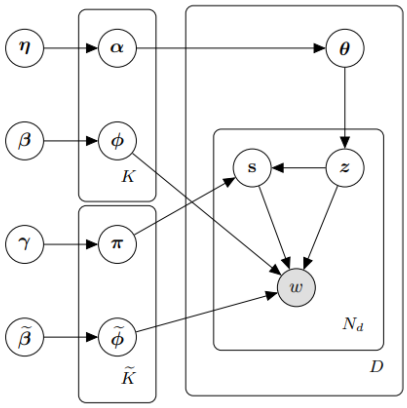


Figure 2.1: Graphical model of the base KeyATM (Eshima et al., 2020).

keywords.

The data generation process, as represented in Figure 2.1, first draws for each word i in document d, a latent topic variable $z_{di} \in \{1, ..., K\}$ from the topic distribution of the document: $z_{di} \sim Categorical(\theta_d)$ where θ_d is a K-dimension vector of topic probabilities whose elements sum to one.

If the sampled topic is one of the non-keyword topics, then we draw $w_{di}|z_{di} \sim Categorical(\phi_k)$, with ϕ_k is a V-dimension vector of word probabilities whose elements sum to one.

In case of a sampled topic is a keyword topic, a Bernoulli random variable s_{di} is drawn with success probability π_k for word i in document d. If this variable is equal to 1, then word w_{di} is drawn from the set of keywords for the topic based on probability vector $\tilde{\phi}_k$. In contrast, if s_{di} is equal to 0, then we sample the word from the standard topic-word distribution of the topic ϕ_k.

For sufficiently large data, it was shown that the choice of hyperparameter priors does not matter (Eshima et al., 2020), except for the prior of π_k whose values controls the influence of keywords which is assumed to follow a probability distribution Beta(1,1). Hence, the authors consider base KeyATM model as a mixture of two distributions: one for positive probabilities for keywords only, and one with positive probabilities for all words. The inference is performed via collapsed Gibbs sampling algorithm, as provided by the paper (Eshima et al., 2020).

The dynamic version of KeyATM, as pictured in Figure 2.2, was found to yield more stable topic proportions than the Structural Topic Model (STM) implemented by Roberts et al. (2013), as it encompasses a Hidden Markov Model (HMM) structure that defines T latent states ($h_{1:T}$) for the period of study, allowing the generative mechanism of topic probabilities of documents θ_d to evolve over time, via a forward-and-backward sampling procedure for $h_{1:T}$.

2.5 Data and results

2.5.1 Federal Reserve governors' speeches

Federal Reserve governors' speeches were scraped from the Federal Reserve website, comprising 1,488 speeches from January 1996 to December 2020, pre-processed to remove numbers, symbols, stopwords, as well as lemmatization, using $spaCy$ (Honnibal et al., 2020) English language model. Rare words whose frequencies are less than 3 were excluded from the corpus, in order to reduce the

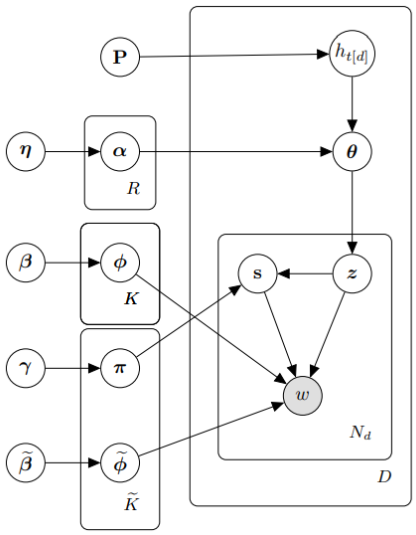


Figure 2.2: Graphical model of the dynamic KeyATM (Eshima et al., 2020).

size of the document-term matrix. A time variable was created to gather all speeches given between two FOMC meetings, serving as a period variable for the dynamic KeyATM inference.

I chose three groups of keyword related to the monetary policy objectives in the United States, as stated by the Congress, displayed in Table 2.1: $inflation$, $interest\ rate$ and $unemployment$ topics and three non-keyword topics that capture other word structures in the corpus[2]. Figure 2.3 displays the proportions of the chosen keywords, from a total of 15,557 words in the cleaned corpus.

Keyword Topic	Words
Inflation	inflation, rate, price, targeting, target.
Interest rate	interest, rate, borrowing, lending borrower, lender, mortgage.
Unemployment	unemployment, rate, employment, job, labor.

Table 2.1: Keywords used to label monetary policy goals for both corpora.

Dynamic version of KeyATM was applied using the standard hyperparametrization[3] provided by Eshima et al. (2020), considering two discrete latent states for time periods and setting the number of iterations to 5,000.

Table 2.2 shows top 20 words in keyword topics, where the word $rate$, although initially included in all keyword topics, seems to match inflation and unemployment topics rather than interest rate topic. Inflation topic exhibits strong ties to classic central banking policy and top words suggest a preference for long-term outlooks, while unemployment topic is linked to economic indicators (growth, productivity) and the real sector (investment, demand, price). Interest rate topic is close to retail banking and housing sector. Non-keyword topics[4] (Table 2.3) exhibit interesting structures, that could be identified as: financial stability (nk topic 3), financial innovation (nk topic 1) and banking regulation (nk topic 2).

[2]Increasing the number of non-keyword topic might assign keywords to non-keyword topics.

[3]Default parameters: $\gamma_1 = \gamma_2 = 1$, $\beta = 0.01$, $\tilde{\beta} = 0.1$, $\mu = 0$, $\sigma = 1$. One half of the samples will be discarded as burn-in.

[4]nk for non-keyword

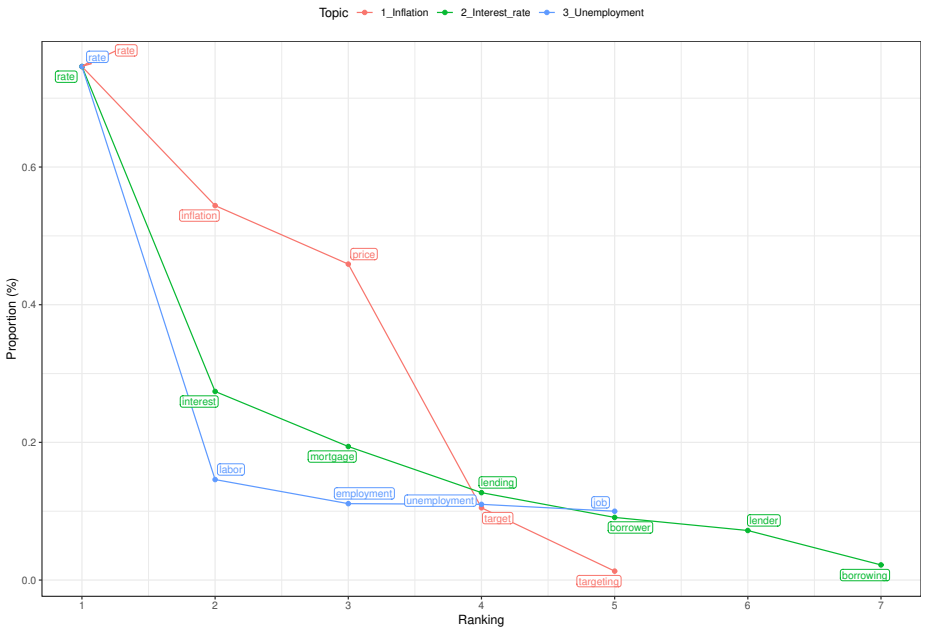


Figure 2.3: Proportion of keywords in the governors' speeches corpus.

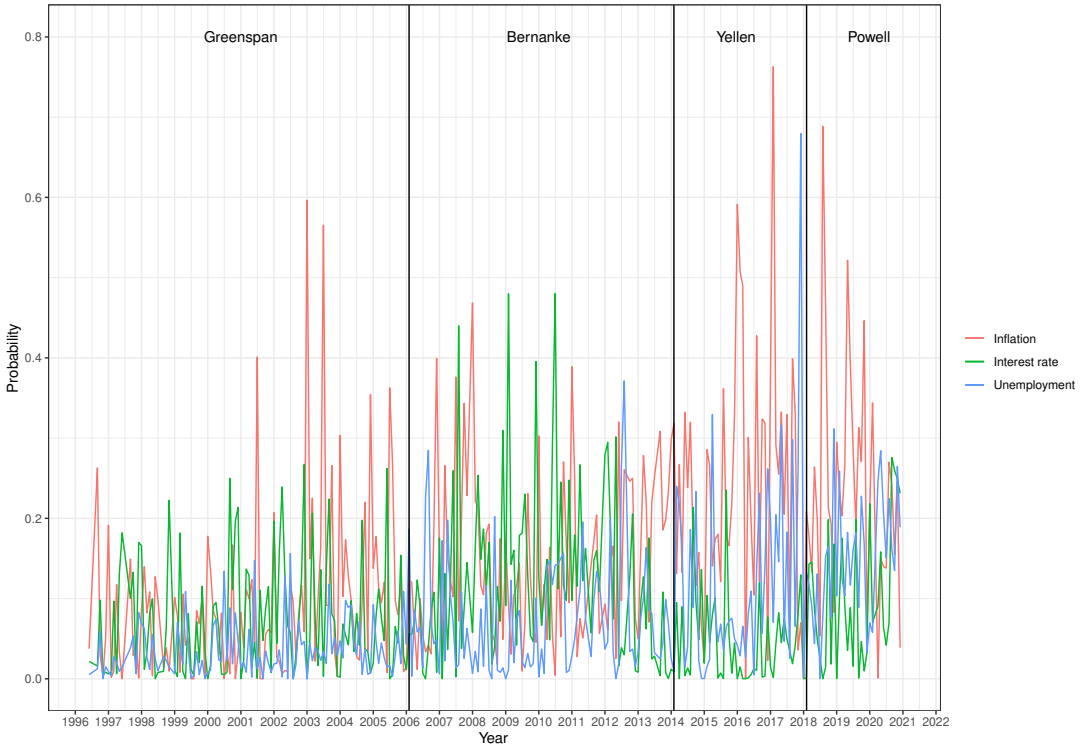


Figure 2.4: Evolution of keyword topic scores (monthly averaged) in the FED speeches.

Figures 2.4 and 2.5 show the monthly averages of the topics scores, respectively for keyword and non-keyword topics. Inflation topic (Figure 2.4) was the hot topic in earlier speeches until the beginning of the financial crisis (end-2007) while the interest rate gained the attention of Federal Reserve members in the aftermath of the crisis until the second round of the Quantitative Easing program (end-2012), which marked a clear dominance of inflation in the subsequent speeches. Figure 2.5 shows the financial innovation and banking regulation were the most non-keyword topics contained in the pre-2007 period, which marked a growing concern about the financial stability, but just second to financial innovation. By tenure, financial innovation and unemployment were dominant during Greenspan's time

Inflation	Interest rate	Unemployment	nk topic 1	nk topic 2	nk topic 3
policy	community	rate ✓	country	bank	financial
inflation ✓	mortgage ✓	price [1]	economy	risk	market
rate ✓	federal	year	market	banking	bank
monetary	loan	growth	economic	will	risk
federal	reserve	increase	united	capital	crisis
reserve	business	percent	states	management	asset
bank	credit	will	world	financial	credit
text	consumer	market	one	institution	liquidity
economic	financial	economy	u.s	regulatory	firm
return	small	rise	trade	basel	reserve
central	payment	high	government	process	system
price ✓	lending ✓	economic	financial	organization	federal
long	income	productivity	system	change	capital
interest [2]	return	recent	international	market	institution
term	development	low	exchange	information	fund
fomc	work	labor ✓	foreign	system	large
economy	service	decline	technology	firm	stability
see	text	long	global	new	term
target ✓	borrower ✓	demand	new	business	return
expectation	housing	business	many	practice	stress

Table 2.2: Top 20 words of KeyATM for governors' speeches (keywords have ticks and numbers in brackets indicates keywords of specified topics).

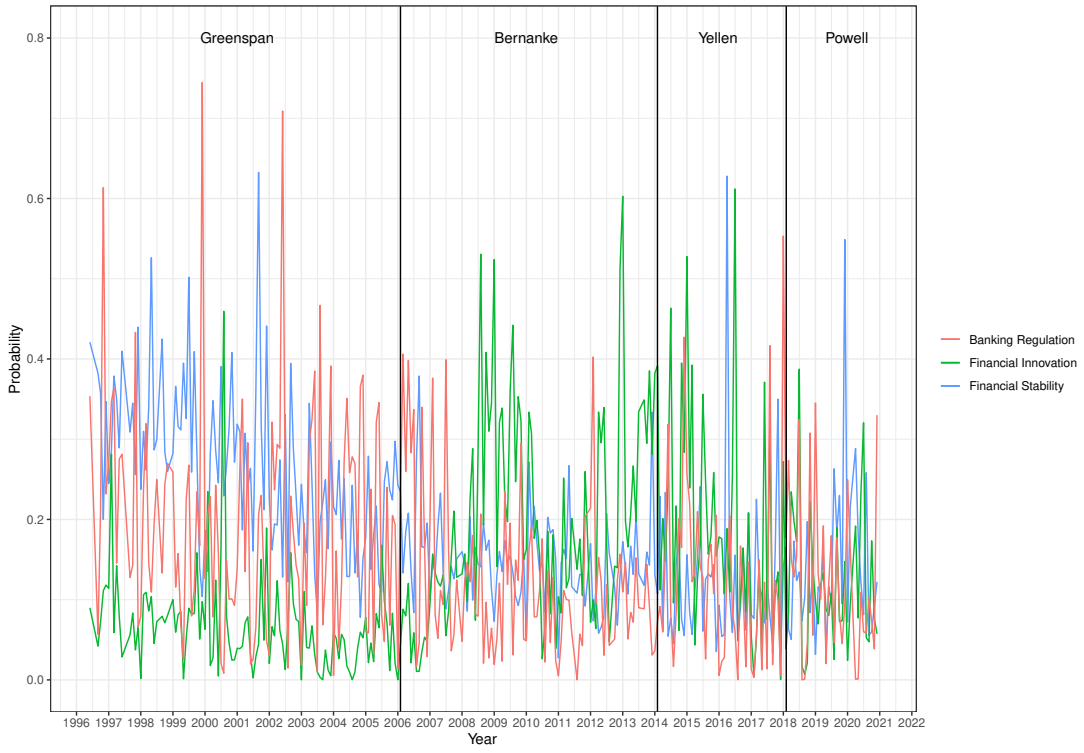


Figure 2.5: Evolution of non-keyword topic scores (monthly averaged) in the FED speeches.

(1996-2006), while the mandate of Bernanke coincided with the financial crisis and debated mostly the interest rate and financial stability. Inflation attracted the attention of Federal Reserve members during the tenure of Yellen and continues to dominate the overall topics. Recent development, related to the Covid-19 pandemic, shows a resurgence of interest rate and financial stability topics in the speeches.

2.5.2 FOMC transcripts

FOMC transcripts that occured after the public release decision (1993) were scraped from the Federal Reserve website, selecting all documents from 1994 until 2016, comprising 250 meeting and conference-call transcripts, preprocessed to remove numbers, symbols, stopwords, as well as lemmatization, in a similar fashion as for FED speeches. For the rest of the analysis, individual contributions within each meeting were gathered into a single document for each member, so to build informative documents for the KeyATM estimation.

Table 2.3 show high consistency of keywords topics with the selected keywords, while non-keywords are not as informative as in the governors' speeches, given that FOMC transcripts are by essence dialogues/debates between members around specific points, rather than structured texts given as addresses.

Inflation	Interest rate	Unemployment	nk topic 1	nk topic 2	nk topic 3
rate ✓	bank	unemployment ✓	think	chairman	chairman
inflation ✓	market	rate ✓	go	greenspan	bernanke
growth	credit	policy	say	president	president
price ✓	financial	will	will	will	will
forecast	interest ✓	labor ✓	one	move	thank
risk	rate ✓	chair	get	governor	think
year	mortgage ✓	committee	just	go	go
economy	asset	market	mr	yes	governor
policy	loan	economic	make	question	let
market	will	percent	us	one	question
high	capital	year	now	get	inflation [1]
percent	fund	time	market	issue	comment
will	business	yellen	point	know	meeting
increase	treasury	purchase	good	number	just
see	new	thank	way	meeting	like
low	large	job ✓	much	point	issue
point	balance	inflation [1]	thing	change	discussion
demand	institution	employment ✓	want	inventory	take
term	liquidity	september	really	significant	make
last	debt	statement	see	look	okay

Table 2.3: Top 20 words of KeyATM for FOMC transcripts (keywords have ticks and numbers in brackets indicates keywords of specified topics).

2.6 Individual preferences estimates

Topic scores, as results of KeyATM, were used to estimate governors' preferences for inflation and unemployment for both corpora[5]. The idea is to link the inflation topic score drawn from KeyATM with the prevailing inflation rate in the United States (lagged year-on-year monthly inflation rate). For each governor in both corpora, I estimated inflation preferences $\pi_i^\star$ using a rolling regression with grid search values ranging from 0 to 5 (increments by 0.1) and a vector of remaining keyword topic scores (unemployment and interest rate) as covariates, then I selected the value of $\pi_i^\star$ for each governor that maximizes the goodness of fit (R^2).

$$\alpha_{it} = \delta_1(\pi_{t-1} - \pi_i^\star) \times \mathbb{1}[\pi_{t-1} > \pi_i^\star] + \delta_2(\pi_{t-1} - \pi_i^\star) \times \mathbb{1}[\pi_{t-1} \leq \pi_i^\star] + \mathbf{X}_t\Lambda + \epsilon_{it} \tag{2.2}$$

[5]The analysis excluded governors who do not appear in both corpora, for the sake of comparability in terms of policy preferences.

The same methodology applies to unemployment preferences $U^{\star}$ which uses lagged monthly unemployment rates in the United States and a vector of remaining keyword topic scores (inflation and interest rate) as covariates. Results are given in Table 2.4 at governor and corpus levels.

Figure 2.6 offers an in-depth picture of partisanship among governors using the estimated inflation and unemployment preferences respectively from FED governors' speeches and FOMC transcripts. Preferences from FOMC transcripts seems to be homogeneous when compared to FED speeches, indicating similarities between governors that could be identified as small clusters, while preferences of FED speeches are scattered and cannot capture strong alignments. Particularly, the idea of consensus (Meade, 2005) appears to be prevailing at FOMC transcripts, independently from their voting and it is possible that the prevailing variability in FED speeches' preferences result from a kind of freedom endowed to governors' when giving their addresses.

FOMC members' preferences are mostly comprised between 2% and 3.5% for inflation rate, and between 5% and 7% for unemployment rate, while FED speeches have relatively lower unemployment preferences (between 4% and 6%) and noticeable departure from the 2%-3% range for inflation rate. FED corpus tend to have lower unemployment preferences compared to FOMC.

Overall, there is a significant downward trend in the inflation preferences, at the same time an upward trend in unemployment preferences. This may indicate growing worries about the general price levels that were found to be more important, at the governor level, than unemployment numbers. Governors with a law studies background tend to favor low unemployment preferences, in comparison with others from an economic background which have higher inflation preferences, and recent FOMC members (tenures in the 2010s) tend to favor low inflation levels (around 1%), in comparison with former members.

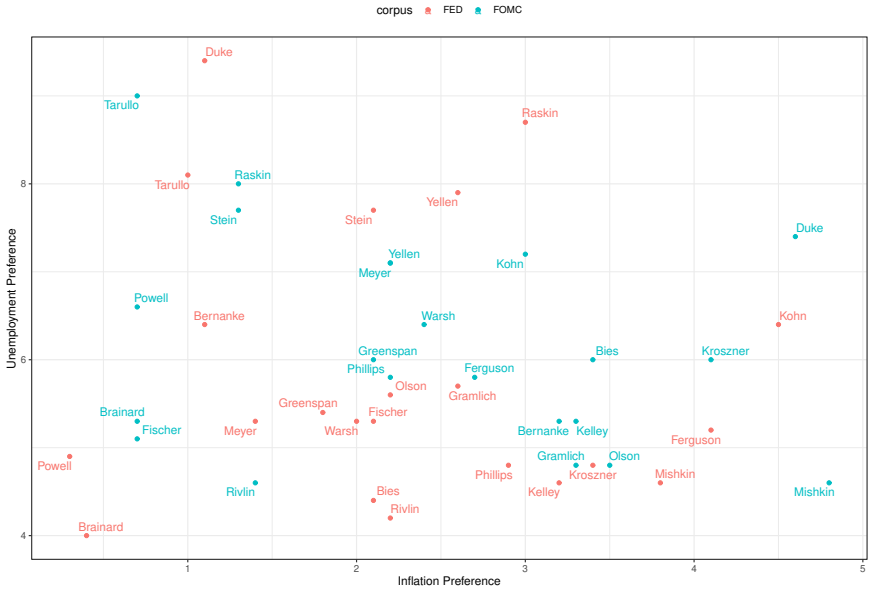


Figure 2.6: Biplot of estimated inflation and unemployment individual preferences.

Governor	$\pi^{\star}_{FED}$	$U^{\star}_{FED}$	$\pi^{\star}_{FOMC}$	$U^{\star}_{FOMC}$
Bernanke	1.1	6.4	3.2	5.3
Bies	2.1	4.4	3.4	6.0
Brainard	0.4	4.0	0.7	5.3
Duke	1.1	9.4	4.6	7.4
Ferguson	4.1	5.2	2.7	5.8
Fischer	2.1	5.3	0.7	5.1
Gramlich	2.6	5.7	3.3	4.8
Greenspan	1.8	5.4	2.1	6.0
Kelley	3.2	4.6	3.3	5.3
Kohn	4.5	6.4	3.0	7.2
Kroszner	3.4	4.8	4.1	6.0
Meyer	1.4	5.3	2.2	7.1
Mishkin	3.8	4.6	4.8	4.6
Olson	2.2	5.6	3.5	4.8
Phillips	2.9	4.8	2.2	5.8
Powell	0.3	4.9	0.7	6.6
Raskin	3.0	8.7	1.3	8.0
Rivlin	2.2	4.2	1.4	4.6
Stein	2.1	7.7	1.3	7.7
Tarullo	1.0	8.1	0.7	9.0
Warsh	2.0	5.3	2.4	6.4
Yellen	2.6	7.9	2.2	7.1

Table 2.4: Governors' inflation and unemployment preferences estimated using FED governors' speeches and FOMC transcripts.

2.7 Conclusion

This works investigated central bank policy signals under narrative economics with the help of a keyword-based topic models, so to map individual monetary objectives of FED governors from topic scores learned through FED speeches and FOMC transcripts. Three keyword topics, concerning monetary objectives set by the Congress in the United States, seem to be dominant in both corpora, while non-keyword topics were found to be informative in FED speeches and concerning financial stability, financial innovation and banking supervision. Individual preferences on inflation and unemployment rates were found homogeneous in FOMC transcripts rather than FED speeches, confirming a potential consensus among the members probably due to the debate-structured meetings, while FED speeches yielded scattered preferences that may result from a freedom given to governors when writing their addresses at specific occasions. Individual characteristics, as for the educational background, could explain wide differences in the preferences estimates, where governors with a background in law tend to favor low level of unemployment, in contrast to tolerated higher inflation preferences for economic-educated governors. Finally, broad differences in individual estimates demonstrate heterogeneity in the used communication sources but confirms a potential shared consensus during FOMC meetings.

Chapter 3

Multilinguality in central banking[1]

3.1 Unfolding language and policy preferences of swiss central bankers

3.1.1 Introduction

Narrative economics established itself as a growing field of research in economics (Shiller, 2017), benefiting from the continuous developments of technical solutions that enable powerful analyses of documents (Grimmer et al., 2022), mostly using topics models, as a popular class of machine learning algorithms dedicated to text data (Gentzkow et al., 2019a).

Such unsupervised models constituted the workhorse of text mining applications, aiming at uncovering latent groups of words, known as topics, from the triplet structure document-topic-word (Blei et al., 2003) that offers a better understanding of the corpus but remains exclusively conducted at the monolingual level, while a growing number of corpora are available in different languages need advanced mechanisms to ensure a inter-lingual translation when uncovering topics (Lucas et al., 2015). Recent advances in Natural Language Processing (NLP) offer ways to handle multilingual documents in the same corpus (Bianchi et al., 2021a) using the power of distributional representations (Mikolov et al., 2013; Dieng et al., 2019) which vectorize words and paragraphs to ensure a better topic representation through semantic search (Angelov, 2020; Grootendorst, 2022).

Monetary policy, as a dynamic subfield of economics, witnessed a growing number of applications featuring text data and aiming, particularly, at extracting sentiments from speeches and minutes to explore their impact on financial markets (Grimmer and Stewart, 2013) or to uncover policy preferences of monetary committees (Hansen et al., 2017; Shapiro and Wilson, 2019; Diaf, 2022) mostly using monolingual corpora, where the devised methodologies were built around Latent Dirichlet Allocations.

Central banks around the world have been leveraging channels to maintain a permanent informative channel link with the public and many of them adopted a multilingual communication strategy to reach a wider audience (Athanassiou, 2006; ECB, 2011), especially for multicultural countries or monetary unions. By doing so, central banks diffuse the same information in many languages to prevent any bias stemming from translation and avoid using a *coded language* (Muchlinski, 2011) strategy that increases uncertainty perception among the public. Common references in this area are

[1]Section 3.1 is a paper was co-authored with Florian Schütze and presented at the 6^{th} International Conference on Advanced Research Methods and Analytics (CARMA2024), Valencia, Spain. URL https://ocs.editorial.upv.es/index.php/CARMA/CARMA2024/paper/viewFile/17405/8910.
Section 3.2 is a drawn from the paper TopicShoal: Scaling partisanship using semantic search, co-authored with Ulrich Fritsche and presented at the 18^{th} Conference on Natural Language Processing (KONVENS 2022). URL https://aclanthology.org/2022.konvens-1.20

the European Central Bank (ECB), Deutsche Bundesbank and the Swiss National Bank (SNB) whose communication operates in three main languages provide good example on how multilinguality operates as an adopted communication strategy.

This work sheds light on historical developments of the SNB communication during the last 25 years, by analyzing the trilingual corpus of its members' speeches (1997-2022) using semantic search models, namely $Top2vec$ (Angelov, 2020), that proved to be useful in getting powerful topic representations for both monolingual and multilingual corpora. Particularly, the resulting semantic space permits to test further hypotheses about the rhetoric used by central bankers when addressing their audience, as well as quantifying other interests related to uncertainty and crisis, two commonly debated themes in modern macroeconomics.

Results suggest the SNB communication is centered around the objectives assigned to central bankers, as the number of topics (8) learned by $Top2vec$ is relatively low regarding the depth of the corpus (671 documents). Moreover, secondary objectives as for systemic risk and banking supervision and financial stability have been regularly discussed since the breakup of the 2008 financial crisis, while developments of the Euro-area remained relevant for the stability of the Swiss Franc, even after abandoning the Euro-peg. Finally, topics related to standard monetary practices, as for unemployment and inflation, were mostly reported in English language, in contrast to uncertainty which mainly occur in German speeches.

3.1.2 Central banking communication

The availability of central banking corpora revived the interest of researchers in investigating in-depth hypotheses within monetary policy committees and the impact of central bank narratives on financial markets, mostly with monolingual text-as-data techniques and sentiment analysis methods, with a focus on financial markets (Blinder et al., 2008; Hüning, 2020). This revival could be seen as a product of transparency efforts adopted by many central banks during the last 35 years, aiming at informing the public and moving from secrecy and misunderstanding (Goodfriend, 1986) to a more transparent and accountable image (Haldane and McMahon, 2018), where uncertainty is regarded a forward guidance tool at the hands of monetary policy committees Greenspan (2004). Researchers initially targeted the interaction between central banks and financial markets using sentiment analysis (Blinder et al., 2008) then studying the inner functioning of monetary policy committees (Baerg and Lowe, 2020) and central bankers' preferences toward monetary objectives (Shapiro and Wilson, 2019; Diaf, 2022) by combining plain bag-of-word topic models, as for the Latent Dirichlet Allocation and document scaling techniques which ignore grammatical information (Ash et al., 2024). Recent works used NLP techniques, as for word embedding (Mikolov et al., 2013), to produce a language model (Baumgaertner and Zahner, 2021) based on a relatively big corpus of collected speeches and minutes from different central banks around the world, exclusively in English language, while Diaf and Fritsche (2022) used a pre-trained transformer model to scale Bundesbank executive members, whose speeches were given in four different languages (Karim El-Ouaghlidi et al., 2019).

3.1.3 Topic models

From a text mining perspective, one could identify two separate fields of applications in central banking communication: topic modeling and document scaling, the former tries to yield relevant and interpretable groups of words known as topics (Hansen et al., 2017) while the latter aims to rank document and central bankers for the sake of comparability (Baerg and Lowe, 2020) based on different probabilistic (Slapin and Proksch, 2008) and non-probabilistic (Laver et al., 2003) bag-of-word assumptions.

Bag-of-word techniques remain largely biased, as they overlook grammatical structures that play a key role in extracting narratives (Ash et al., 2024), and need either post-hoc inference or prior information (Eshima et al., 2020), as a semi-supervised or supervised strategy, that may improve the initial results but still being heuristic models, whose results might be dependent on starting values or further hyperparametrization chosen by practitioners, knowing that they were dedicated to monolingual corpora and their application to multilingual documents remains problematic, as they need a reliable transfer mechanism that translates results from one language to another (Bianchi et al., 2021a). Lucas et al. (2015) proposed to combine machine translation with Structural Topic Models (STM) (Roberts et al., 2013) to study comparative politics across languages, while Bianchi et al. (2021b) devised a transfer learning mechanism, based on variational autoencoders, combining distributional representations and bag-of-word that were later generalized to multilingual applications (Bianchi et al., 2021b) yielding results that depend on initial topics usually learned in English and then transferred to other languages via transformers.

The proposed methodology, centered around Top2vec, belongs to the semantic search category of topic models, built on a multilingual pre-trained embedding model to ensure a robustness regarding the richness of corpus and the necessity to adopt a multifaceted approach (Ruder et al., 2019) in-lieu of simple translation procedures that may bifurcate similar cross-language signals contained in the corpus. The pre-trained multilingual models (Dieng et al., 2019) ensures having a sound, online transfer learning that enables mapping topic model results to different language without semantic and syntactic losses.

$Top2vec$, as a three-step semantic search model, used to uncover topics from the dual embedding space constituted of words and documents vectors learned via a dimensionality reduction technique (UMAP) (McInnes et al., 2018), then topics are extracted using a hierarchical classification technique (HDBSCAN) (McInnes et al., 2017) which identifies dense areas of documents, whose centroids are assumed to be a topic vector (Angelov, 2020). In a similar approach, Grootendorst (2022) proposed Bertopic where a clustered TF-IDF scheme (Term Frequency Inverse Document Frequency) is adopted within each topic to get more differentiated structures. Although the functioning of Bertopic is modular, where practitioners can choose between different algorithms at each step, an outlier class is often set to documents which could not be assigned to any learned topic(not well represented in the embedding space), thus generating an information loss when eliminating documents that might be informative to the build-up of topics.

3.1.4 Application

The corpus of SNB speeches was scraped from the SNB website, consisting of 669 different addresses given between 1997 and 2022 in three languages: English (374), French (90) and German (205).

Table 3.1 shows the resulting topics (8) learned via $Top2vec$ and using a multilingual sentence transformer[2] as pre-trained embedding model, to unify the framework and avoid setting as many models as languages that need a post-hoc translation mechanism (Lucas et al., 2015).

Topics, ordered by their importance, are closely related to the goals assigned to SNB [3] as to guarantee a price stability (topics 1 and 4), stability of financial markets (topics 3 and 7), banking system

[2] distiluse-base-multilingual-cased maps sentences and paragraphs to a 512 dimensional dense vector space and can be used for tasks like clustering or semantic search https://huggingface.co/sentence-transformers/distiluse-base-multilingual-cased-v1

[3] https://www.snb.ch/en/iabout/snb/id/snb_tasks

Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6	Topic 7	Topic 8
wahrungspolitik	bankensystem	wahrungspolitik	inflationsraten	euroraum	nationalbank	wahrungspolitik	gold
inflationsraten	bankensystems	finanzmarkten	inflationsrate	euro	geschaftsbanken	devisenmarkten	wahrungspolitik
inflationsrate	finanzkrise	finanzmarkte	inflation	eurogebiet	kantonalbanken	geldmarktzinsen	geldmarktzinsen
inflationary	geschaftsbanken	inflationsraten	inflationary	eurogebiets	bankensystem	geldmarkt	wahrungsfonds
inflation	finanzsystem	geldmarktzinsen	inflationnistes	eurozone	banknotes	devisenmarkt	banknote
inflationnistes	finanzsystems	finanzmarkt	wahrungspolitik	euros	bankensystems	finanzmarkten	banknotes
geldpolitik	zentralbanken	inflationsrate	economist	wahrungspolitik	zentralbanken	currencies	nationalbank
finanzpolitik	finanzmarkten	wahrungsfonds	economists	currencies	banknote	finanzmarkte	monetary
finanzkrise	zentralbank	currencies	inflationsziel	euroland	zentralbank	finanzmarkt	monetaire
economist	finanzmarkte	geldmarkt	economique	europaischen	banknoten	currency	geldpolitik
economists	bundesbank	finanzkrise	economiques	european	bundesbank	inflationsraten	kantonalbanken
geldpolitischen	bankensektor	devisenmarkten	finanzkrise	europas	wahrungspolitik	inflationsrate	monetaren
geldpolitische	finanzmarkt	finanzpolitik	volkswirtschaft	europaische	bankensektor	wahrungen	banknoten
inflationsziel	kantonalbanken	inflationary	economy	europe	bank	monetaire	geldpolitische
geldpolitisches	nationalbank	banknotes	economics	inflationsraten	interbank	inflationary	geldpolitisches
economique	grossbanken	inflation	macroeconomic	currency	grossbanken	monetary	geldmarkt
geldpolitisch	interbank	inflationnistes	recession	inflationsrate	banken	monetare	geldpolitischen
fiskalpolitik	banks	devisenmarkt	wirtschaftliche	inflationary	banque	inflation	bankensystem
geldpolitischer	bank	geldpolitik	economie	europeen	bancaires	monetaren	monetare
economiques	bancaires	economists	economies	europeenne	banking	inflationnistes	geschaftsbanken

Table 3.1: Top 20 words of the learned topics using $Top2vec$.

Term	Corr.	Term	Corr.
uncertainties	0.982	indebtedness	0.770
incertitude	0.971	decisively	0.769
incertitudes	0.946	ungefahr	0.758
uncertain	0.931	preisstabilitaet	0.744
unsicherheiten	0.888	glaubwurdigkeit	0.744
unsicherheit	0.887	inevitable	0.744
unabhangigkeit	0.806	instability	0.742
doubts	0.802	unerwunschten	0.740
unsecured	0.791	conviction	0.737
doubt	0.771	reliability	0.734

Table 3.2: Top 20 terms related to ”uncertainty”, and their correlation given by the cosine similarity of their respective vectors.

(topics 2 and 6) and the currency (topics 5 and 8). Topic 1 could be identified as the practice of monetary policy being closely related to fiscal policy (policy mix), while topic 4 specializes in inflation-targeting.

The multiple occurrence of many words in different topics indicate various contexts used by central bankers when giving their addresses, depending on the prevailing macroeconomic/financial situation and the ”hot topics” they tried to address them. For example, Topic 4 deals with inflation in a macroeconomic context, while Topic 1 still debates inflation but from an inflation targeting perspective, close to fiscal policy which could hint a latent ”policy mix” debate. The topic content uncovered from the speeches given during 25 years is constrained by a relatively fewer terms, if compared to 10 different topics learned via the same methodology using Bundesbank speeches during the period 2012-2017 (Diaf and Fritsche, 2022), which might indicate a potential ambiguity (Baerg, 2020) used by central bankers.

Table 3.2 and Table 3.3 offer specific pictures on the word usage, respectively related to uncertainty and crisis within the SNB. Although the central banking jargon is rich in technical terms related to macroeconomics and financial markets, the use of uncertainty (Table 3.2) seems to be limited to just its literal aspect (in three languages), while crisis-related terms have mostly a macroeconomic color linked to the economic status and financial/banking system (Table 3.3).

Term	Corr.	Term	Corr.
crise	0.983	konjunkturellen	0.692
krise	0.979	jeopardise	0.686
krisen	0.978	deflation	0.685
crises	0.963	deflationary	0.664
konjunktur	0.784	threatened	0.659
finanzkrise	0.783	risikopraemie	0.658
recession	0.738	droht	0.655
konjunkturlage	0.721	collapse	0.649
rezession	0.714	ausbruch	0.645
konjunkturelle	0.705	einbruch	0.642

Table 3.3: Top 20 terms related to ”crisis”, and their correlation given by the cosine similarity of their respective vectors.

	score_uncertainty	score_inflation	score_unemployment	score_interest rate	score_crisis	score_monetary policy
score_uncertainty	1					
score_inflation	0.45	1				
score_unemployment	0.28	0.5	1			
score_interest rate	0.12	0.23	-0.04	1		
score_crisis	0.44	0.55	0.52	0.42	1	
score_monetary policy	0.38	0.44	-0.1	0.11	0.11	1

Table 3.4: Correlations between different built-in indices (variables are cosine similarity scores between the terms and each document of the corpus.

Distances of each document to the word ”crisis”, in terms of cosine similarities of their respective vectors, were used to build a crisis-index to measure its prevalence within each document in the corpus. Although the documents are irregular in terms of issuance, a monthly-aggregated index was built and found to be correlated to business cycle fluctuations in Switzerland (Figure 3.1), where the use of a crisis-related rhetoric seems to color the SNB speeches during technical recessions, except the period 2011-2013, coinciding with the Euro-peg adopted by the SNB.

As for the crisis index, other indices were built to establish a link between specific narratives related to monetary policy aspects and the use of uncertainty and crisis terms which bear the attention of economists in particular and the audience in general. Table 3.4 shows the correlations between the indices reckoned for inflation, unemployment, interest rates as well as the term ”monetary policy”. Overall, correlations are weak and positive, probably due to the span of the corpus and a dynamic word usage, with the use of uncertainty being more dependent on inflation than interest rate or unemployment.

Built indices were used to check a potential language preferences when addressing monetary policy aspects via running a Correspondance Analysis (Benzécri, 1973) between scores greater than 0 and languages, resulting in two dimensional plot[4] as showed in Figure 3.2. English seems to be the go-to communication language when addressing standard monetary policy practices (inflation, interest rate and unemployment) probably targeting an international audience, while German and French languages were found to be preferred when expressing uncertainty, hinting a possible use for domestic addresses or speeches given to a local audience.

3.1.5 Conclusion

Central bank communication has long been a black box when it comes to understand monetary policy signals delivered through its communication channels, and this difficulty rises up when the strategy

[4]Inertia (information percentage) of Dimension 1 and 2 are 91% and 9% respectively.

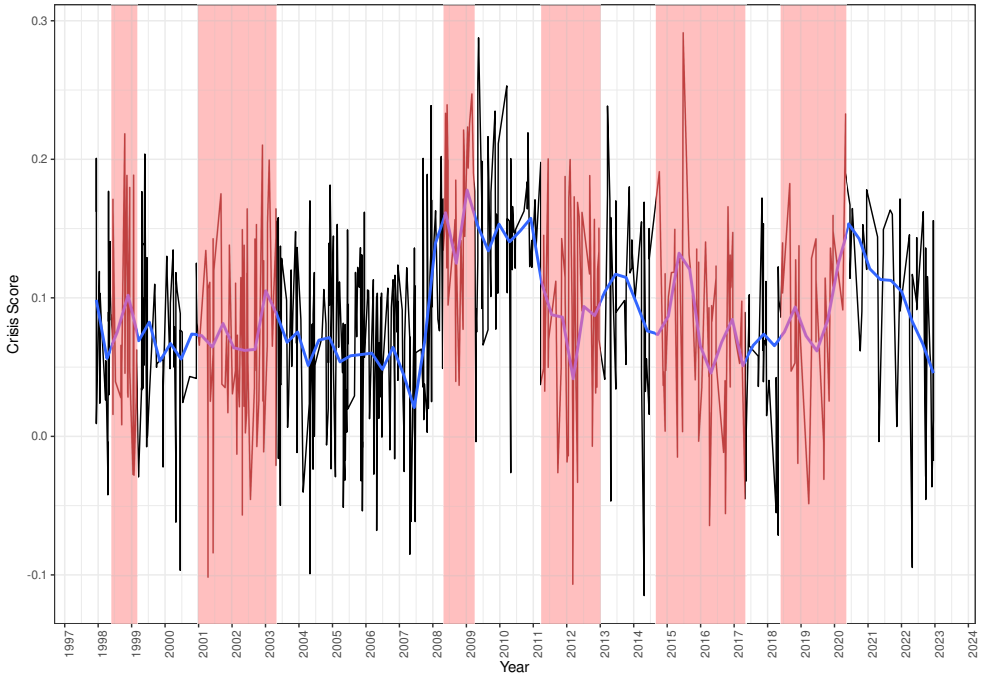


Figure 3.1: Crisis-scores per document and recession periods (in shaded pink). Blue curve represents the LOESS smoothed curve

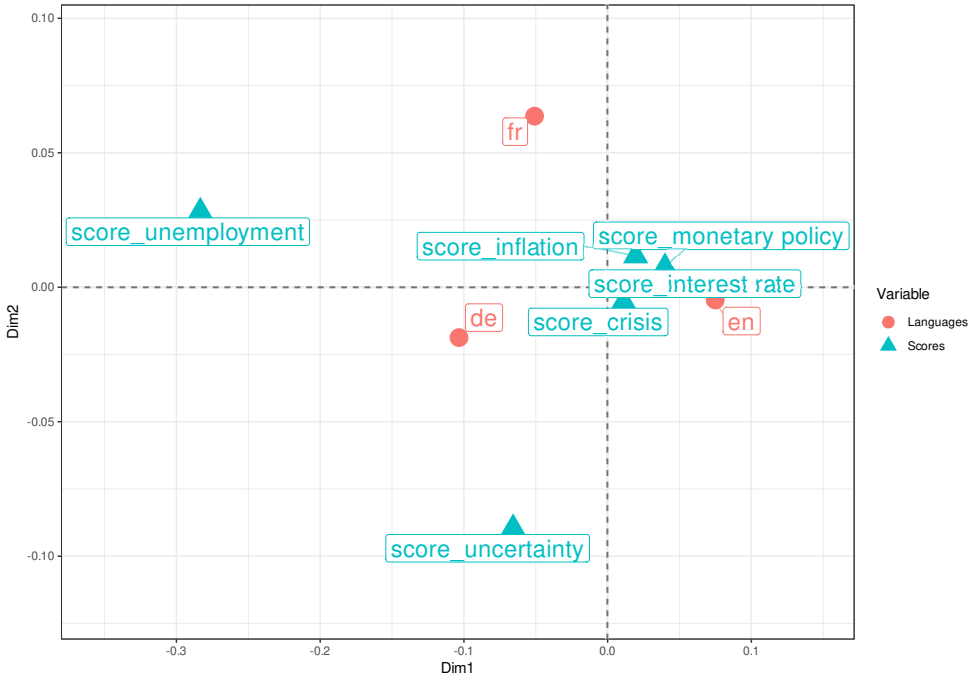


Figure 3.2: Correspondence analysis between the built-in indices (Scores) and documents' languages.

adopts multilingualism to achieve more transparency in its missions. Recent advances in natural language processing permit to unify the analysis framework of multilingual corpora, usually using the power of distributional representations and semantic search architectures. They can uncover meaningful word associations for robust topic structures, in-lieu of bag-of-word techniques that require translation mechanisms and biased human intervention. The multilingual corpus of SNB speeches during the last 25 years reveals a communication strategy essentially built around the goals assigned to SNB, with special interests on financial markets, banking system and currency stability, whose developments are important in steering the modern monetary policy practices. However, the corpus does not offer practical discussions or specific jargon that might color the topics, if compared to other corpora from Eurosystem central banks. The reduced number of topics (eight), compared to the depth and the span of the corpus, does not unfold granular structures that could be assimilated to sub-topics, denoting a moderately informative communication style often adopted to ensure financial markets and potentially colored with ambiguity, as a stylized fact of modern central banking communication. Aside from topic results, semantic search strategies helped building useful associations within the triplet document-topic-word and were found, when using terms related to crisis, to highlight the responsiveness of central bankers toward turmoil and business cycle fluctuations as well as a clear preference for the English language when dealing with standard monetary policy concerns.

3.2 Multilingual topic modeling on the Bundesbank

3.2.1 Introduction

Scaling is considered as the second component of the text mining family, after topic models. It mostly deals with classification techniques aiming at ranking documents on scales, using either parametric of non-parametric inferences.

Parametric scaling techniques were used mostly for political debates to uncover partisanship and ideological differences between parties or politicians (Slapin and Proksch, 2008; Lauderdale and Herzog, 2016), on the basis of probabilistic assumptions on the word counts. The output of such techniques is suboptimal, because terms are taken separately due to the bag-of-word hypothesis.

Diaf and Fritsche (2022) proposed TopicShoal as a two-stage scaling scheme, built on the basis of topics learned via Top2Vec (Angelov, 2020). Aside from the subtle efficiency in ranking parties or speakers on a unidimensional scale, this method is able to scale multilingual corpora when using multilingual pre-trained models. The following application is an excerpt of the TopicShoal paper, featuring the analysis of partisanship of the German Bundesbank executive members.

3.2.2 Top2Vec

First attempt to uncover artificial topics used single value decomposition (SVD) on the document-term matrix (Deerwester et al., 1990). Statistical learning was later used to extract topics as hidden probability distributions joining the observable pair document-word. Latent Dirichlet Allocation (Blei et al., 2003) was established as the workhorse for text-as-data applications, by automatically extract topics in a collection of documents.

This class of Bayesian mixed-membership models do not take into account word interactions and assumes word independence. In most cases, topic quality is not optimal and many researchers adopt post-hoc inferences to try to correct or guide the inference (Eshima et al., 2020).

On the basis of vectorization of words, known as word embedding (Mikolov et al., 2013), Angelov (2020) proposed a semantic search approach to genuinely uncover topics based on a dual document and word representation. $Top2Vec$ augments the usual distributional representation methods, as for Word2Vec, by adding a paragraph vector to the neural network (Angelov, 2020) to create a joint word and document representations forming a semantic space able to uncover associations that helps learning coherent topic vectors from dense areas of document using Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN) (McInnes et al., 2017), under the hypothesis that the number of dense areas of documents is equal to the number of topics. Hence, the number of topics is no longer a hyperparameter as for most algorithms.

$Top2Vec$ features a structure of independent, mostly low-correlated, topics because of the HDBSCAN application, ensuring a non-overlapping outcome often found in traditional topic models, hence enabling a robust Bayesian aggregation on independent topics, instead of a debate-structure that might have an intertwined topic prevalence.

3.2.3 Bayesian factor analysis

We use a modified version of the Bayesian aggregation used in Wordshoal (Lauderdale and Herzog, 2016) by setting the document positions as being drawn from a truncated normal distribution, instead of a normal distribution, as the document-topic coefficients are indeed distances mainly on the [0,1]

Topic	Top 10 words
1	eurosystems finanzpolitik eurosystem euroraums finanzkrisen eurozone bankensektors geldpolitik geldpolitischer finanzkrise
2	bargelds geldpolitik geldmarkt bankbilanzen currency monetaren bargeld monetaire wahrung eurosystem
3	bankensektors bankensektor innovationen innovations finanzbranche finanzsektors bankensystems innovation finanzsektor bankensystem
4	eurosystems eurosystem zahlungsverkehr euroraums eurozone zahlungsmittel euroraum kreditvergabe geldmarkt transaktionen
5	repercussions risikoteilung nachhaltig risques risks risiko nachhaltige krisenmaßnahmen risque risk
6	empirical data statistics analyses statistical trends informationen indicators analysen finanzsystems
7	digitalen verbraucher digitale digitalisation consumer digital consumers cyber technologien technologie
8	cyber security sicherheit threat sicherheiten sicherzustellen safeguarding vulnerable secure danger
9	blockchain bitcoins bitcoin geldmarkt currencies zentralbankgeld bankbilanzen bankensystem geldpolitik bankensystems
10	geldpolitik geldpolitischer geldmarkt renminbi geldpolitische geldpolitischen currencies currency zentralbankgeld staatsanleihen

Table 3.5: Top 10 words of the topics learned by $Top2Vec$ on the Bundesbank speeches corpus.

interval.

Let ψ_{ij} defines the score of i^{th} document in the j^{th} topic learned via $Top2Vec$. The Bayesian aggregation used in Wordshoal to infer a latent scale, represented by a vector of speakers' positions θ_i is as follows:

TopicShoal algorithm

1^{st} Stage: Apply $Top2Vec$ and extract the inferred topics:
ψ_{ij} defines the distance between the i^{th} document and the j^{th} topic (based on cosine distance).

2^{st} Stage: Each topic inferred is assumed to form a debate:
Inferring ideal points θ_i using the following factor analysis:

$$\psi_{ij} \sim \mathcal{N}(\alpha_j + \beta_j \theta_i, \tau_i)$$
$$\theta_i \sim \mathcal{N}_{trunc}(0, 1)$$
$$\alpha_j, \beta_j \sim \mathcal{N}(0, 0.25)$$
$$\tau_i \sim \mathcal{G}(1, 1)$$

---

where $\mathcal{N}_{trunc}$ denotes the truncated normal distribution as ψ_{ij} are represent document-topic distances. β_j is a topic polarization parameter.

Lauderdale and Herzog (2016) assumed debates being independent and serving as a basis to a multiple Wordfish scaling within each debate, that renders different word contribution for each scale. While this assumption allows a dynamic word contribution per debate, it ignores a potential topics' prevalence that might differentiate speakers or parties out of the debate dimension. Hence, building an Bayesian factor analysis on semantic topics makes it possible to track their prevalence in the unidimensional scale of positions, using the learned β_j. Put it differently: We assume topic usage to be distinctive for the positioning in the ideological dimension. Some topics have a high degree of distinctiveness (high absolute values of beta), whereas other topics are not distinctive. The sign of beta tells us, if the respective topic is leaning more to the "left side" or the "right side" of the scale (as it is normalized to have a mean of zero).

In other terms, TopicShoal ensures a debate transfer from a time perspective to a topic structure for a better interpretability of the ideal positions. This is motivated by the fact that debates are defined by their occurrence, but usually discuss the same topics or concerns.

3.2.4 Results

The dataset of Deutsche Bundesbank executive board members' speeches (Karim El-Ouaghlidi et al., 2019) is used to test the multilingual version of $TopicShoal$ with the help of a multilingual embedding model that ensures a topic-transferability between different languages used in the corpus.

	β_i
Topic 1	-0.16
Topic 2	0.44
Topic 3	-0.15
Topic 4	0.22
Topic 5	-0.78
Topic 6	-0.22
Topic 7	0.30
Topic 8	-0.82
Topic 9	-0.01
Topic 10	0.53

Table 3.6: Estimated topic intensity β_i using TopicShoal on Bundesbank executive board members' corpus.

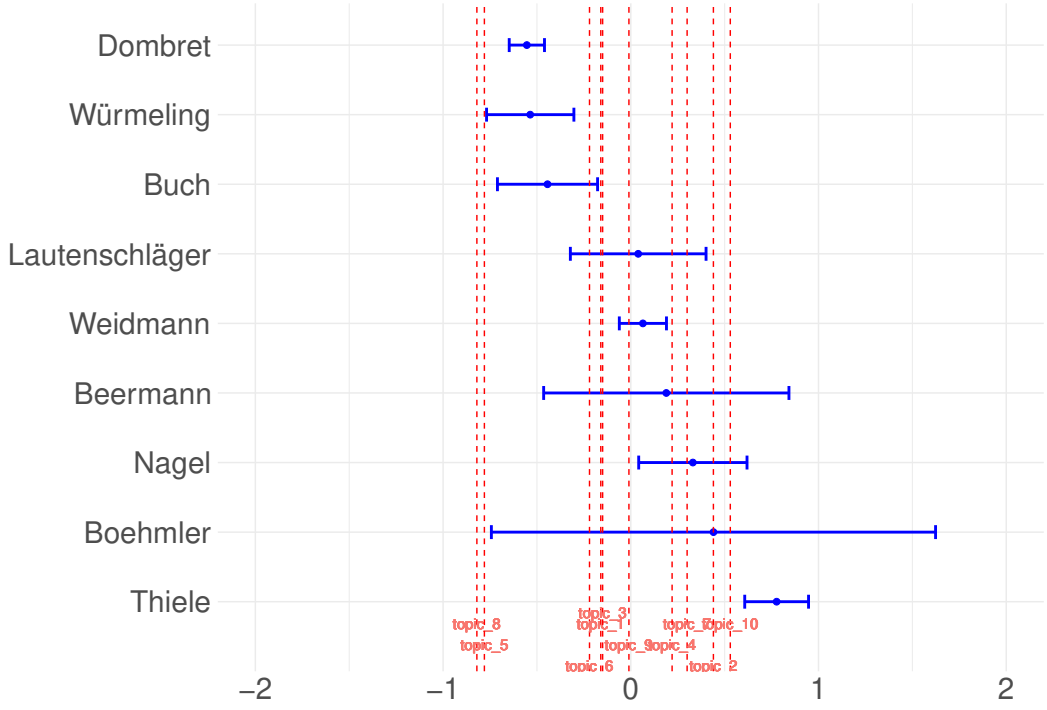


Figure 3.3: Estimated Bundesbank executive board members' ideal points using $TopicShoal$ with projected topic contributions.

The dataset comprises 791 speeches given by nine different executive board members during the period 2012-2017 in four different languages (english, french, german and italian) although english and german share 98% of the corpus. $TopicShoal$ is used to extract central bankers positions using multilingual embedding [5] (Reimers and Gurevych, 2019) given to $Top2Vec$ that uncovered 10 different topics related to various aspects of monetary policy practices, as for risks and vulnerabilities (topic 5), European concerns (topic 1 and 4), financial innovation (topic 3), security and digitalization (topic 7, 8 and 9) and monetary policy (topic 10) as displayed in Table 3.5.

The positional analysis, as mentioned in Figure 3.3, helps classifying members into small groups of similar interests, given the learned topics, where topics related to classical monetary policy (topics 2 and 10) are polarizing positive members' positions, while risks and crisis-related concerns are mostly linked to negative positions, as reported in Table 3.6. Positions with wide confidence intervals (Beermann and Böhmler) could be explained by the variety of speeches, members gave during the period, while firm positions with relatively small confidence intervals (Dombret, Weidmann and Thiele) indicate a potential specialization or theme preferences of the members.

[5] paraphrase-multilingual-MiniLM-L12-v2

Chapter 4

Ambiguity in central bank communication: a strategic vagueness or a pure randomness?[1]

Abstract

In recent years, there has been a significant amount of research on monetary policy that focuses on exploring communication channels used by central banks and using text-based covariates to improve the predictability of many macroeconomic variables. Although economists have extensively researched this area, the literature does not cover the specific characteristics unique to central bankers as for ambiguity, which is considered as a stylized fact and a common feature of modern central bank practices. Its use as a communication strategy could be seen as a forward guidance tool in the hands of decision makers. Previous approaches to study ambiguity have used Poisson-based document scaling techniques to infer ideal points and define an ambiguity measurement as the difference between the learned ideal point and the real one that should prevail if all hypotheses are verified. This paper employs an augmented scaling technique, CommunityFish, which is based on hierarchical clustering to obtain groups of words, or communities to measure ambiguity in the collection of Federal Reserve governors' speeches (1996-2020) and explaining its main drivers, along with other macroeconomic variables. Results indicate fluctuations of ambiguity scores per-speaker, with a pronounced trend linked to crises and inflation spikes. In probabilistic terms, ambiguity measurement confirms the deviation of ideal points from the classic Poisson paradigm. Communities were found to be useful independent covariates in explaining ambiguity fluctuations, which exhibit randomness and weak persistency, being mostly related to a crisis jargon.

4.1 Introduction

Ambiguity has a long history in human life, first scrutinized by philosophers (Titt, 2021), before being treated in the field of linguistics. Black (1937) considered ambiguity as the amount and complexity of information to which we are exposed, rather than a specific phenomenon, while some others still define it as the situation where a piece of information triggers multiple but a finite number of different interpretations or linguistic representations (Titt, 2021), similar to the definition given by Black (1937), which distinguishes between ambiguity and vagueness.

Most modern approaches to the study of ambiguity still apprehend it from a lexical perspective, coining it as the polysemy of words and phrases, and emphasizing the importance of the context in revealing the true meaning of messages.

[1]This paper was accepted at the 17^{th} International Conference on Statistical Analysis of Textual Data (JADT 2024), Brussels, Belgium. https://jadt-2024.sciencesconf.org/data/program/Ambiguity_in_Central_Banking_Communication.pdf

In monetary policy, ambiguity is considered, with consensus, as a stylized fact of modern monetary practice across central banks around the world (Baerg, 2020). Despite transparency efforts adopted since the late 1980's, modern central bank communication still suffers from an information deficit that renders signals unclear for the audience. This misunderstanding arises as an information ambiguity that can also affect speeches given by experts and practitioners as well.

In text mining, document scaling refers to a class of mostly statistical methods that aim to derive a coefficient that summarizes the information for each document in the corpus. Thus, if two documents have a similar word usage, or close probability distributions over words, they may have similar coefficients in the built scale, even if the vocabulary is not strictly identical.

The most dominant approach to document scaling considers the Poisson distribution as the paradigm or handling word frequencies (Slapin and Proksch, 2008). It assumes rare events, even though some words may be very rare and deviate from the Poisson paradigm, thus influencing the scale with their respective contributions. Lo et al. (2016) proposed a Monte Carlo-based inference based on the negative binomial distribution to better scale of documents, and to provide a measure of the deviation from the Poisson distribution known as the clarity parameter, interpreted as the inverse of ambiguity.

While the notion of ambiguity has been extensively studied from a polysemic perspective, this work considers ambiguity as communicative fact that emerges from the flow of information delivered by speakers to the audience, and thus as a hidden textual feature that could be approached using coefficients, or ideal points, that can represent each document in the corpus.

In this paper, I adopt an augmented document scaling method to assess ambiguity from a positional perspective, that is, the deviation of ideal points from their optimal values. For this purpose, I take into account the community structure as given by Diaf (2023), which consists in detecting the most significant groups of words in the corpus (communities) through hierarchical clustering (namely the Louvain algorithm (Blondel et al., 2008)) and using them as features for the scaling algorithm, instead of simple word counts. Thus, communities are more interpretable features than words, in addition to speeding up the learning process whose structure is based on intensive Monte Carlo simulations. Communities, as text variables, can therefore serve as covariates to explain variation in ambiguity, along with other metadata information contained in the corpus.

I argue that the measurement of ambiguity does not exclusively rely on document scaling but goes beyond it, as an explanatory exercise to identify its sources from textual features. Narrative signals emerging from the collection of documents are able to discern ambiguity either as a random noise, or as a structured phenomenon. Therefore, the corpus of the Federal Reserve governors' speeches (1996-2022) is used to test external central bank communication and check ambiguity persistence, at the speaker and period levels.

Results show that ambiguity scores are embedded with a high volatility per time and per speaker. Its variability is not purely random but could be explained by the occurrence of specific communities, as n-gram features, related to the economic/financial jargon. Recent serving governors tend to have high ambiguity averages compared to previous ones, while financial and banking jargon dominates ambiguity formation, along with certain terms directly related to the 2008 financial crisis. The high number of uncovered communities reinforces the hypothesis of a dynamic word usage used by central bankers (Diaf and Schütze, 2022) when considering long time periods.

4.2 Central bank communication

The availability of central bank corpora has renewed researchers' interest in investigating hypotheses within monetary policy committees and the impact of central bank narratives on financial markets. This has been done mostly using monolingual text-as-data techniques and sentiment analysis methods, with a focus on financial markets (Blinder et al., 2008). This revival could be seen as a product of transparency efforts adopted by many central banks during the last 35 years, aiming at informing the public and moving from secrecy and misunderstanding (Goodfriend, 1986) to a more transparent and accountable image (Haldane and McMahon, 2018), where uncertainty is considered as a forward guidance tool for monetary policy committees (Greenspan, 2004). Initially, researchers focused on the interaction between central banks and financial markets using sentiment analysis (Blinder et al., 2008), then studied the inner working of monetary policy committees (Baerg and Lowe, 2020) and central bankers' preferences towards monetary objectives (Shapiro and Wilson, 2019).

This was achieved by combining plain probabilistic topic models, such as the Latent Dirichlet Allocation (Blei et al., 2003) and document scaling techniques that ignore the semantic features due to the Bag-of-word assumption. Recent studies have used NLP techniques, such as word embedding (Mikolov et al., 2013), to develop a language model (Baumgaertner and Zahner, 2021) based on a large corpus of collected speeches and minutes from various central banks around the world, exclusively in English, while Diaf and Fritsche (2022) employed a pre-trained transformer model to scale the speeches of Bundesbank's executive members, which were given in four different languages, using a hybrid topic model-document scaling approach.

Although topic models are commonly used to analyze central bank communication, their results are mostly agnostic and require post-hoc adjustments. However, Diaf (2022) found that robust semi-supervised techniques can capture meaningful topics without the need of post-hoc adjustment.

4.3 Ambiguity in text mining

The concept of ambiguity is often unclear, particularly in text mining, as numerous approaches have been proposed to address various shortcomings, however, linguistic ambiguity remains the primary focus. These methods can be broadly classified into two main categories (Hasan et al., 2020): knowledge-based and statistical approaches, the former relies on structural information available in lexicons and dictionaries, while the latter uses word frequencies.

Fawareh et al. (2008) and Bhonde et al. (2010) proposed semantic approaches to address ambiguity in entity identification/recognition using fuzzy sets. Turenne (2013) considered ambiguity as uncertainty noise linked to polysemic words and used a nearest neighbor-strategy to identify similar structural content in a given corpus. Hasan et al. (2020) evaluated several measurements of synonymy and ambiguity using semantic representations.

For textual data, document scaling comprises a set of unsupervised algorithms that aim to rank documents, based on specific assumptions. Probabilistic scaling methods are commonly used by researchers in political sciences to learn a unidimensional scale for document positions. For example, Wordfish (Slapin and Proksch, 2008) assumes that word frequencies follow a Poisson distribution, which allows for the derivation of word and document-level positions, as well as a policy position, that can be used as a proxy for ideological stance.

Although document scaling techniques are widely used in political sciences, they remain subjected to the bias-prone expertise/judgment of practitioners as unsupervised measurement models. Goet

(2019) suggested to use supervised models to produce meaningful estimates of polarization, which could be seen as a semi-unsupervised approach.

Attempts to estimate ideological stances rely on the Poisson distribution which assumes equality between the mean and the variance of each word occurrence in every document. Additionally, document positions recovered via Wordfish are expected to be normally distributed $\mathcal{N}(0, 1)$ which, is not always verified in practice.

Diaf (2023) proposed an augmented version of Wordfish, called CommunityFish, based on a community detection algorithm, and aims to concentrate the scaling on meaningful n-grams extracted via Louvain algorithm. Hence, instead of considering all words in the corpus, the input given to Wordfish (the document-term-matrix) will only feature communities (n-grams of different lengths depending on the corpus structure).

Departures from the Poisson paradigm are frequently encountered when many words appear frequently or rarely in the corpus, making the use of the Poisson distribution biased. Lo et al. (2016) used a scaling procedure that relies on the negative binomial distribution instead of Poisson to improve positional estimates and uncover the clarity parameter, whose value determines the departure from the Poisson hypothesis. The inverse of the clarity parameter, known as ambiguity, provides information on the ideological position of a given document can be.

The model assumes the count of the word j in the document i at time t noted as , follows a Poisson distribution:

$$y_{ijt} \sim Poisson(\lambda_{ijt}) \tag{4.1}$$

$$log(y_{ijt}) = \alpha_{it} + \psi_j + \beta_j \times \omega_{it} \tag{4.2}$$

where α is a set of parties/actors fixed effects, ψ represents word fixed effects and β the word-importance when building the ideological positions.

Instead of a Poisson-generated model, Lo et al. (2016) suggested an alternative parametrization on the basis of negative binomial (NB) distribution, defined as the probability distribution associated with the number of successes in an independent sequence of Bernoulli trials before a specified number of failures occur. A random variable $X \sim NB(r, p)$ where p is probability of success r as a mean $\mu = \frac{r(1-p)}{p}$ and a variance $\sigma^2 = \frac{r(1-p)}{p^2} = \mu + \frac{1}{r}\mu^2$ is the predefined number of failures could be approached, asymptotically, as a Poisson distribution for big values r ($r > 50$)

$$Poisson(\lambda) = \lim_{r\to\infty} NB\left(r, \frac{\lambda}{r+\lambda}\right) \tag{4.3}$$

so that parametrization 4.1 could be rewritten as :

$$y_{ijt} \sim NB\left(r_{it}, \frac{\lambda_{ijt}}{r_{it}+\lambda_{ijt}}\right) \tag{4.4}$$

The model in equation 4.4 is estimated using Monte Carlo Markov Chains (MCMC)[2], constraining the first values of α_{it} and β_j to be null. The negative binomial scheme captures, via the variance component $\frac{1}{r_{it}}\mu^2$ the overdispersion in the document i at time t, relative to what one would expect

[2] Lo et al. (2016) with slice sampling and a burn-in of 1,000 iterations so to take into account 5,000 draws to calculate point estimates.

from the Poisson distribution, hence, $\theta_{it} = \frac{1}{r_{it}}$ is interpreted as the clarity parameter, indicating an increased ideological clarity for big values of r_{it}.

4.4 Application

The corpus of speeches given by Federal Reserve Governors (1996-2020) was scraped from the Federal Reserve website. It consists of 1,603 speeches given by 31 different governors (speakers), lemmatized to reduce the wording and to facilitate the application of the Louvain algorithm (Diaf, 2023) which yielded 91 independent communities whose structures are mostly dominated by bigrams. Among them, 42 were found to significantly explain ambiguity variability over time (Table 4.2).

Table 4.1 shows per-speaker average ambiguity scores and variances which are independent from the number of speeches given. The concentration of speeches around their means could be explained by relatively low variances. The means remain higher than the standard ratios adopted by Lo et al. (2016), indicating that word frequencies depart from the Poisson distribution due to the influence of outliers, which are apparent in Figure 4.1 as spikes. These findings support the hypothesis of a dynamic word usage, due to the variety of themes discussed by governors over the past 25 years (Diaf and Schütze, 2022). As a result, estimates of ambiguity are highly volatile (Figure 4.1) and the number of communities identified via hierarchical clustering, as a hard clustering scheme, is inflated.

Long-serving FED chairmen (Greenspan and Bernanke) have ambiguity scores close to the corpus average. In contrast, governors who make occasional addresses (Jefferson, Lindsey and Barr) have the lowest averages. Recent governors (Gramlich and Brainard) exhibit high averages. This difference could be explained by recent debates on issues related to the pandemic and environmental concerns, which depart from standard monetary policy practices.

Speaker	Avg. Ambiguity	Var. Ambiguity	Num. speeches	Speaker	Avg. Ambiguity	Var. Ambiguity	Num. speeches
Gramlich	0,367	0,092	59	Fischer	0,167	0,029	45
Duke	0,276	0,066	44	Olson	0,167	0,026	46
Brainard	0,263	0,122	107	Powell	0,164	0,031	99
Mishkin	0,235	0,028	27	Tarullo	0,157	0,023	54
Kroszner	0,216	0,056	44	Kohn	0,152	0,030	76
Meyer	0,216	0,084	73	Clarida	0,148	0,014	42
Yellen	0,214	0,053	66	Bies	0,147	0,012	77
Waller	0,212	0,061	19	Warsh	0,133	0,016	18
Bernanke	0,210	0,055	215	Phillips	0,118	0,011	12
Quarles	0,210	0,069	57	Jefferson	0,116	0,020	2
Bowman	0,209	0,057	43	Lindsey	0,110	0,005	3
Rivlin	0,197	0,050	17	Barr	0,108	0,021	3
Raskin	0,193	0,047	17	Madigan	0,106	——-	1
Stein	0,181	0,016	16	Kelley	0,096	0,001	12
Ferguson	0,180	0,050	123	Cook	0,076	0,002	2
Greenspan	0,180	0,051	184	Average	0,198	0,050	1603

Table 4.1: Estimated ambiguity average and variance per speaker.

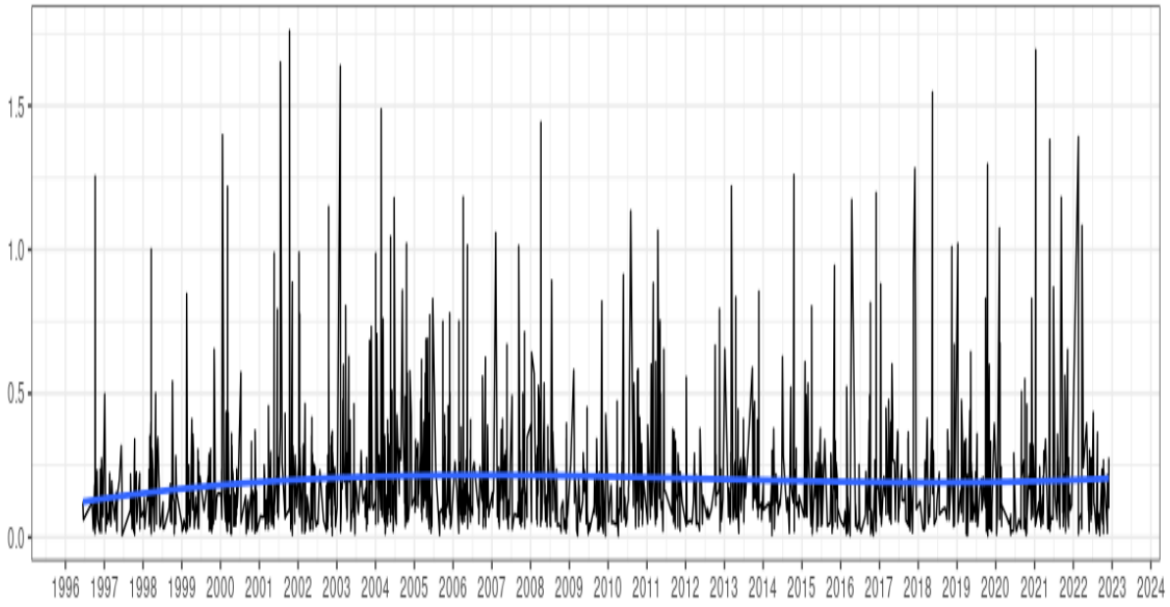


Figure 4.1: Estimated ambiguity scores over time (Line in blue is the Loess curve.

Table 4.2 presents 91 communities that significantly explain ambiguity scores, along with their respective occurrences in the corpus. These communities are related to financial and banking jargon, as well as economic status description, with noticeable bigrams linked to the 2008 financial crisis. Ambiguity scores, taken as monthly averages, do follow a weak autoregressive process of order 2, indicating a certain cyclicality in the governors' reported speeches. This is mostly due to coordinated communication efforts, which try to act as a single voice for main debated themes and issues.

4.5 Conclusion

Ambiguity is often associated with polysemy and has been studied in terms of different meanings a single word can have, depending on its context. However, some consider ambiguity to be an unclear strategy that adds noise to the discourse. In text mining, ambiguity is linked to the message, a speaker or a party tries to convey to the audience, where the words' probability distribution departs from the standard Poisson paradigm as words occurrences become very rare of frequent. Advances in document scaling now enable learning an ambiguity parameter, which measures the difference between the prevailing word distribution and the ideal one, but proved to be computationally expensive in terms of simulations. To address this issue, hierarchical clustering can be used to scale documents more efficiently and identify which n-grams play a significant role in ambiguity formation across the corpus. In monetary policy, ambiguity is considered as a stylized fact and a commonly used strategy by central bankers around the world, whose identification for the case of the United States' Federal Reserve showed a strong correlation with the jargon used to describe the economic status and terms related to the financial crisis. This persistence may stem from common debated topics of interest and a coordinated communication strategy. Furthermore, ambiguity scores indicate significant variations at the governor level, surpassing conventional thresholds. This is likely due to the diverse range of topics addressed by governors over the past three decades, which extend beyond the typical focus on inflation, unemployment, and interest rates in standard monetary policy practices. At the light of these developments, it cannot be assumed that ambiguity measurement, at least in the case of the Federal Reserve, is a pure random occurrence. Rather, it is a stylized fact of the communication strategy adopted for policy purposes, given the growing number of themes and issues that governors need to address during their speeches.

Comm.	Top Words
com_1	0.5, 1 , 1.5, 1.6, 1.7, 10 , 100, 11 , 12 , 13 , 14 , 15 , 16 , 17 , 18 , 19 , 2 , 2.2, 2.4, 2.5
com_2	1990 , 2001 , 91 , deep , great , recession , deal , britain , depression, moderation, pleasure
com_3	19th , 2000 , 2008 , 2009 , 20th , 21st , addition , already , although , another , bad , careful , case , century , certain , coming , critical , critically, crucial , cyclical
com_4	2003 , age , brookings , commercial , discussion , displaced , economics , eds , educated , finance , hispanic , imf , kb , kiley , macroeconomics, mb , michael , nber , non , older
com_5	absorb , academic , activity , adverse , african , american , analysis , anecdotal , baseline , carnegie , co , conference, continued , economic , economy's , efficiency, empirical , employment, evidence , evolving
com_6	absorbing , approach , associated , audit , balanced , based , best , broad , broker , card , challenges , communication, compliance , consensus , control , counter , counterparty , credit , debit , default
com_7	accelerating , achieve , actual , adjustable , anchor , annual , average , benefit , black , bound , breakeven , bring , broadly , capacity , close , college , commission , communications, communities , compensation
com_8	accessible, figures , version
com_9	accommodative, actions , appropriate , call , comment , conduct , conducting , conventional , corrective , decision , disclosure , dual , expansionary , extra , final , fiscal , general , highly , implement , instrument
com_10	account , balance , budget , checking , congressional, current , enforcement , fed's , law , okun's , population , retirement , savings , sheet , deficit , surpluses , office , sheets , accounts , circumstances
com_11	accounting , accurate , adequate , asymmetric , available , basle , capital , cash , charge , check , comprehensive , conservation , contingency , coso , countercyclical, coverage , cra , creating , data , detailed
com_12	achieving , agencies , assets , banking , broader , changing , complex , consolidated, crisis , denominated , depository , derivative , dollar , economy , education , educational , electronic , enough , entire , ex
com_13	across , adopt , advanced , affordable , anti , asian , basel , borrowing , care , civil , committee , committee's, computing , costs , countries , developed , developing , driving , east , economies
com_14	act , bliley , cares , dodd , frank , glass , gramm , improvement, leach , lombard , main , oxley , sarbanes , secrecy , steagall , street , substantial, wall , requires
com_15	active , around , bank , bank's , bankers , banks , board , board's , borrower's , bulletin , central , chartered , clearing , community , company , corporation, deposit , development, digital , encourage
com_16	ad , hoc
com_17	adam , smith
com_18	additional , adversely , allow , allows , also , among , anchored , ask , assess , became , become , becoming , behind , better , bit , briefly , can , clear , comparative, competition
com_19	adjustment , advance , advisory , attitudes , back , climate , closely , come , considerable, cut , date , dramatic , examination , forward , fundamental , going , interagency , issued , look , looking

com_20	agency , assessment , asset , auto , backed , beveridge , bond , border , city , collateralized, commodity , common , corporate , country , cre , cross , crude , crypto , debt , declined
com_21	aggregate , baby , boom , bottom , business , cambridge , census , confidence , consumer , consumption , cycle , dashed , demand , direct , distribute , domestic , econometric , excess , expenditures, favorable
com_22	almost , always , certainly, surely
com_23	andrew, levin
com_24	artificial , intelligence
com_25	athanasios, orphanides
com_26	bear , stearns
com_27	blanchard, olivier
com_28	blue, chip
com_29	bretton, woods
com_30	brian, sack
com_31	building, blocks
com_32	centrally, cleared
com_33	chief , executive
com_34	clarida, richard
com_35	congress, passed
com_36	constrained, discretion
com_37	creative , destruction
com_38	cutting, edge
com_39	d'amico , stefania
com_40	de , facto, novo
com_41	depend , importantly
com_42	dot , plot
com_43	double, digit
com_44	downward, put , upward , pressure
com_45	due , diligence
com_46	elected , representatives
com_47	else , equal
com_48	et, al
com_49	examples, include
com_50	expressed, views , concern
com_51	fannie, ginnie, mae
com_52	fiber, optic
com_53	francisco, san , diego
com_54	freddie, mac
com_55	friedman, milton
com_56	hong, kong
com_57	hyun, song, shin
com_58	jpmorgan, chase
com_59	kenneth, rogoff
com_60	lehman , brothers
com_61	los , angeles
com_62	lose , sight
com_63	machine , learning

com_64	michigan, surveys
com_65	mishkin , frederic
com_66	moral , hazard
com_67	nellie, liang
com_68	north , carolina, holland
com_69	often , referred
com_70	owner , occupied
com_71	practical, matter
com_72	prepayment, penalties
com_73	raw , materials
com_74	rely , heavily
com_75	repurchase, reverse , agreement , agreements
com_76	secular , stagnation
com_77	site , web , examinations
com_78	soft , landing
com_79	speech , view , charts , delivered
com_80	st , louis
com_81	stanley, fischer
com_82	states , united , kingdom
com_83	taper , tantrum
com_84	terrorist, attacks
com_85	thomas , laubach
com_86	trimmed, mean
com_87	uninsured, creditors
com_88	vast , majority
com_89	via , webcast
com_90	watch, live
com_91	william, english, wascher

Table 4.2: List of communities learned using $CommunityFish$ algorithm.

Chapter 5

Investigating uncertainty in central bank corpora[1]

Abstract

The study of narrative economics using text mining techniques grew recently with a dominance of unsupervised models built upon the bag-of-word assumption, that try to extract meaningful information for further inferences. Yet, word independence remains a constraining hypothesis when the narratives rely on an implicit word usage that belongs to a specific jargon. Recent advances in natural language processing offer powerful distributional representation tools to preserve the semantic and the syntactic meaning of the text, as a reliable alternative to frequency-based probabilistic topic models. Monetary policy, as an important application field of narrative economics, carries uncertainty in its decision-making process as a forward guidance tool, besides being a strategic aspect of its communication policy, and witnessed several attempts to construct uncertainty indices, using uncertainty-related word counts. This yielded questionable measurements that overlook key technical terms not encompassed in wordlists, hence making the computed indices biased and semantically agnostic vis-à-vis the proper jargon. This work proposes an in-depth assessment of uncertainty in a collection of international central bankers' speeches (1997-2022) and identifies its key drivers using semantic search models, namely $Top2vec$, to uncover nested semantic topic structures, at the national and international levels. These proxies of uncertainty sources proved to be robust in discerning uncertainty features associated to probability and risk, in-line with the Keynes-Knight debate about uncertainty in macroeconomics. Moreover, a robust uncertainty index could be built from the similarity between each document and the term $uncertainty$, either country-specific or internationally, which was found to follow major world financial and banking events of the last two decades.

5.1 Introduction

Narrative economics (Shiller, 2017) popularized the already existing interest in extracting information from textual data using a variety of techniques, often borrowed from machine learning, and use the results as covariates for further inferences and tests (Gentzkow et al., 2019a). Social scientists inherited guidance and practices from computer scientists (Grimmer et al., 2022), when it comes to deal with text data, and the dominant techniques are mostly based on the bag-of-word assumption, whose application, even if it succeeds in yielding acceptable results, remains suboptimal and constrained (Grimmer and Stewart, 2013) because it overlooks the semantic features stemming

[1]This paper was co-authored with Florian Schütze and appeared under the title: Uncovering uncertainty in narrative economics: A semantic search approach. In New Frontiers in Textual Data Analysis. Springer Nature Switzerland, 2024. URL https://www.springerprofessional.de/uncovering-uncertainty-in-narrative-economics-a-semantic-search-/50037724

from documents (Ash et al., 2021), as for word interactions and context information, which play a key role in defining the extracted narratives from the corpus. Advances in natural language processing investigated the importance of context words and the possibility to transform words into vectors encompassing the semantic and syntactic features, also known as distributional representation (Mikolov et al., 2013). This concept was later extended at the paragraph or document level (Dieng et al., 2019) and adopted by other architectures (Angelov, 2020) as a dual word-document embedding used to efficiently search for meaningful and coherent topic representations that shed more light on the traditional triplet structure document-topic-word.

Central banking, as an active economic field of interest, witnessed several contributions belonging to the text-as-data chapter. Mostly to gauge sentiments communicated by reports (Gentzkow et al., 2019a), scaling central bankers (Baerg and Lowe, 2020; Diaf, 2022; Shapiro and Wilson, 2019) to understand potential partisanship among monetary policy members or investigating transparency within these committees (Hansen et al., 2018). However, specific characteristics of central banking communication, as for ambiguity (Baerg, 2020) and consensus (Meade and Stasavage, 2008), make it difficult to discern intrinsic features as for uncertainty, which is considered as a forward guidance tool in modern central banking (Greenspan, 2004).

Earlier attempts to quantify uncertainty, at the general level, used wordlists (Loughran and McDonald, 2011) exclusively for the English language as for the Economic Policy Uncertainty index (EPU) (Baker et al., 2016) which counts the occurrence of uncertainty terms in a wide collection of newspapers and aggregate the results to release a monthly index. This scheme was later expanded to the case of Monetary Policy uncertainty index (MPU) (Husted et al., 2020) while Kerdjoudj and Curé (2015) proposed a Resource Description Framework (RDF) to uncover uncertainty from semantic web that relies on knowledge graphs.

The existing literature about uncertainty estimation in economic corpora employs word-count indices (Baker et al., 2016; Ahir et al., 2022) and although they were found to be informative and able to be used as covariates for further inferences (Tobback et al., 2018), they fail to determine sources of uncertainty and tie them to specific topics of interest that better explain the context used for such assertions. Moreover, the word-count approach used to compute uncertainty indices remains limited because the available wordlists are not as dynamic as the proper language used by central bankers known to be ambiguous (Baerg, 2020).

This paper proposes a semantic, topic-based approach to assess uncertainty in central banking communication using semantic search models, namely $Top2vec$ method, to identify relevant topic structures and terms, semantically related to uncertainty from the Bank of International Settlements (BIS) collection of central bankers' speeches (1997-2022) around the world (Hansson, 2021). While traditional bag-of-word methods use variational inference on word frequencies to agnostically learn global topic structures, their output do not fit corpora with a dynamic wording and embedded information that come as nested topics related to a specific task, as for uncertainty. By doing so, we augment previous attempts to create a language model for central banking (Baumgaertner and Zahner, 2021) by setting a domain-specific semantic space without any prior information that might alter the results (Papakyriakopoulos et al., 2020) as for the use of pre-trained embedding models.

We argue that semantic search strategies are better suited to uncover semantic uncertainty, as explained by Szarvas et al. (2012), and offer the advantage of not requiring further post-hoc inferences or prior information from external sources that may alter the results (Eshima et al., 2020). Lastly, they can detect uncertainty key drivers from topic representations and reveal their contents, in-line with the idea of $political\ framing$ (Entman, 1993) that considers the spe-

cific use of words and phrases when discussing an idea, conveys political messages (Vafa et al., 2020).

Applied to the BIS corpus (1997-2022), the methodology confirms the macroeconomic color of uncertainty debated by central bankers throughout the world, and unveiled a knightian approach that distinguishes uncertainty from risk, putting in front the role of communication strategies, at different central banks, as a determinant of uncertainty variation in their monetary policy practices. An international uncertainty index, built upon the content of each document, proved to follow major international developments in finance and banking that prevailed during the last two decades.

Our paper makes two distinct, significant contributions. In topic models, we demonstrate the efficiency of semantic search models and their ability to uncover nested informative topics when analyzing a large collection of documents, hence, being able to reveal specific word usage, as for uncertainty, and getting most correlated topics and words associated to it. In policy analysis, we introduce a bias-free, corpus-based uncertainty index built upon policy-context topics (Szarvas et al., 2012), which proved to be in-line with major economic and financial developments experienced during the last two decades.

The paper introduces uncertainty in modern central banking and the need to adopt a semantic approach to uncover its latency in texts as an improvement to the existing uncertainty indices. Final section discusses the results and tests conducted on the BIS corpus.

5.2 Methodology

5.2.1 Topic models

Bag-of-word (BOW) techniques have been used by researchers as the default choice for the text-as-data applications. Because of word independence, they cannot uncover semantic features embedded in the corpus (Ash et al., 2021) and remain weak when exploring complex corpora having local nested structures, even if a multitude of extensions were designed to tackle such shortcoming (Blei and Lafferty, 2006; Eshima et al., 2020) they remain suboptimal if compared to distributional representation models (Mikolov et al., 2013) based on the inclusion of context information when vectorizing words and documents in the corpus, allowing for a more semantic approach to text analysis (Le and Mikolov, 2014) that outperforms BOW-based solutions.

$Top2Vec$ (Angelov, 2020) augments the usual distributional representation methods (Mikolov et al., 2013), by adding a paragraph vector to the neural network, hence creating a joint word and document representation forming a semantic space able to uncover associations that help learning coherent topic vectors from dense areas of documents using Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN) (McInnes et al., 2017). This approach offers a better detection of documents and words related to a given task, either using a local embedding model (for instance $Doc2vec$ (Dieng et al., 2019)) or a pre-trained embedding model (Devlin et al., 2019) for the case of multilingual corpora. The underlying hypothesis of $Top2vec$ assumes the number of dense areas of documents is equal to the number of topics, hence, the latter is no longer a hyperparameter, as for most topic model algorithms. The semantic space yields non-overlapping low-correlated topics having an informative power, compared to probabilistic topic models, better represented by vectors that allow comparisons with documents and words. $BERTopic$ (Grootendorst, 2022) follows the same methodology as for $Top2vec$ but uses a clustered frequency transformation within each topic (clustered Term Frequency-Inverse Document Frequency, cTF-IDF), in addition to allowing the use of sentence transformers as pre-trained models, however it gathers small isolated clusters in a single topic of outliers, that could be still interpretable for the task. While the use of sentence transformers helps to increase the efficiency of topic models, it embeds results with an information bias linked

to the corpus upon which the models were trained on (Papakyriakopoulos et al., 2020), altering the semantic space that might not render a clear picture of the studied corpus. For instance, investigating uncertainty in the field of monetary policy is different from fiscal policy, although both fields belong to macroeconomics.

Methodology:

1. Apply $Top2Vec$ (Angelov, 2020) to all documents in the corpus and retrieve the topic representations.
2. Compute cosine similarity of each topic vector with the word vector of $uncertainty$ to get most related topics to uncertainty.
3. Extracting the most correlated terms (words) to uncertainty by filtering their respective vectors to $uncertainty$ word vector.

5.2.2 Central bank communication

As a result of text-as-data revolution, monetary policy witnessed a growing interest in collecting and analyzing large volumes of speeches, meeting transcripts and other textual materials that may help understanding the inner functioning and monetary policy committees around the world (Baerg, 2020). This was facilitated by previously-adopted transparency strategies (Gentzkow et al., 2019a) aiming at informing the public and breaking the secrecy that used to prevail until the 1980s (Goodfriend, 1986). Aside from classical sentiment analysis, text-as-data application in the field of central banking communication mostly used classic topic models (Hansen et al., 2018), as for the Latent Dirichlet Allocation (LDA) (Blei et al., 2003), or adopted NLP improvements as for word embedding (Baumgaertner and Zahner, 2021).

Investigating uncertainty in narrative economics was first done at the global economic level as for the EPU index (Baker et al., 2016; Ahir et al., 2022) and later for monetary policy with MPU index (Husted et al., 2020) using a simple word count of uncertainty-related words or selected terms from the Loughran-McDonald uncertainty wordlist (Loughran and McDonald, 2011) with no guarantee about an effective analysis of context words, being corpus-specific, are often ignored when building such indices. Annotation tools for keyword-based switching measurements (Keith et al., 2020), sentence-based text classification (Jean et al., 2016) and out-of-distribution classifier (Hu and Khan, 2021) were also proposed to provide further measurements of uncertainty.

Leveraging uncertainty is nowadays part of monetary policy practices over the world, and key feature of their communication strategies. As for the United States, it was shown that the usage of uncertainty in speeches rises ahead of crisis and declines right after crisis (Diaf and Fritsche, 2022), proving that central bankers attempt to influence the public and to steer market movements when delivering their speeches. However, assessing sources of uncertainty remains a pure philosophical subject still debated nowadays (Packard et al., 2021), although it finds its origin from earlier works of Keynes (Keynes, 1936) and Knight (Knight, 1921), with no attempts to unfold its origin or quantify its magnitude within central banking communication.

5.3 Data and results

The corpus of international central bankers' speeches was collected from the Bank of International Settlements (BIS) (Hansson, 2021) comprising 16,888 speeches given by members of 118 different central banks around the world in English language, from January 1997 to March 2022, with

Topic	Top 10 Words
1	australia australian graph australians melbourne rba ceda zealand that pretty
2	bsp philippine filipinos filipino advocacies bangko lahat sentral natin pilipinas
3	swedish riksbank the sweden subduing deregulations that subdue in is
4	indian india rbi viz fiis extant fii compulsions contextual emptions
5	think dont jct wouldnt thing isnt bild germans really something
6	union emu supranational european sovereignty rompuy pact parliaments political esm
7	south africa rand african sarb johannesburg africans the in of
8	canadians canadian saskatchewan canada alberta ontario interprovincial quebec vancouver thiessen
9	albania albanian instat equilibriums tirana lek deems boa fullani ardian
10	norwegian norway norges gausdal krone storting petroleum gpfg nok norsk
11	auditors ethics ethical audit compliance unethical culture boards auditing profession
12	climate carbon environmental climaterelated greening decarbonisation ngfs cop emissions green
13	esrb macroprudential systemic macro prudential microprudential warnings esfs esrc crdiv
14	academic renowned his him eminent goethe prestigious he professor excellent
15	keynesian bayesian nonlinearities literature phillips canonical stochastic noisy nonlinear equations
16	spanish foreseeably espana frob prolongation cnmv habitually spain ngeu banco
17	srm ssm edis srf discretions single brrd resolution union mutualised
18	mutualising kilter union disciplining europe german germany pact mutualisation mutualised
19	accord ama irb basel approaches npr ii lcbos supervisors qis
20	barbados caribbean caricom trinidad tobago barbadian jamaica oecs guyana barbadians

Table 5.1: Top 20 topics learned by $Top2Vec$.

contrasted numbers of publications by central bank.

Applying $Top2Vec$ on the corpus, without pre-processing steps, yields 174 different topics that could be separated into two distinct groups: general topics and country-related ones, as stated in Table 5.1, which displays the most debated 20 topics, ranked by their occurrence. Eleven of the top 20 topics are frequent contributions of major central banks, the remaining nine topics are general-purpose related to ethics, climate, academia, keynesianism and banking supervision (Basel Accords). $Top2vec$ is indeed a blend of three unsupervised methods that tries to cluster similar documents in one topic, resulting in many country-specific topics that gather speeches from the same central bank. Representations are clearly granular if compared to a plain LDA learned with 10 topics (Table 5.4) which appears to have repetitive, overlapping terms that do not permit to differentiate topics, due to a regular jargon appearing in the speeches and weighing on the learning process.

The semantic space formed by clustered documents and words vectors (Angelov, 2020) will be used to get the correlation (semantic similarity) of each speech in the corpus with the $uncertainty$ word vector. Results in Table 5.2 show uncertainty correlated to situations where information is vague, ambiguous, or misleading in-line with the uncertainty definition given by Szarvas et al. (2012). The most 20 correlated topics with the word $uncertainty$ are less country-specific than the overall topic structure displayed in Table 5.1. For instance, the most uncertain topic (14) has a correlation of 0.34 to the word uncertainty and is mainly linked to a macroeconomic modeling jargon, while the second one (topic 29) deals with cacophony and secrecy that used to prevail before transparency efforts at the end of 1980s, in what was described as monetary mystique (Goodfriend, 1986). Other topics show that uncertainty is closely related to macroeconomic analysis, particularly topic 126 highlights a descriptive statistical jargon (“medians”, “uptrend”, “decelerate”) and Topic 165 primarily relates to international energy markets, as a source of global business cycle-related fluctuations and hence, uncertainty.

Topic	Cor.	Top 10 Words
14	0.343	keynesian bayesian nonlinearities literature phillips canonical stochastic noisy nonlinear equations
29	0.322	communication communicate communicating cacophony postmeeting audiences secretive geraats transcripts delphic
33	0.287	mpc boe-pressoffice bankofengland uk pretty broadbent spare undershoot furlough england
55	0.286	monetary policy alertness smets theoretical literature rostagno detken activist misalignments
26	0.256	tltros reflationary tltro pepp recalibrated reinvestments app dfr measurably accommodative
35	0.254	mertens elb fomc kiley reifschneider longerrun liftoff roberts chung johannsen
60	0.251	softness pickup readings homebuilding damped crosscurrents payrolls nondefense softened nonfarm
79	0.250	definition mechanistically monetary euro ecb eurosystem orientation policy price assigns
85	0.243	mexican mexico noncore chilean peso bonanza sanchez nafta emes generalized
24	0.242	allotment nonstandard ltros liquidity refinancing dysfunctional dysfunctions eurosystem malfunctioning mro
126	0.227	medians sluggishness heightening moderately mpm uptrend tankan firmness materialize decelerate
2	0.202	swedish riksbank the sweden subduing deregulations that subdue in is
120	0.197	doubtless pickup damped payrolls damp damping quickened softness outsized nonfarm
39	0.194	snb swiss franc berne chf francs switzerland maechler discontinuation sfr
165	0.192	gas crude oil opec refineries barrels drilling shale barrel gasoline
50	0.183	confirms governing almunia outcome upside subdued area tilted unutilised medium
7	0.179	canadians canadian saskatchewan canada alberta ontario interprovincial quebec vancouver thiessen
66	0.179	turkish cbrt turkey cbt implementations tl mtp turkstat lira disinflation
9	0.178	norwegian norway norges gausdal krone storting petroleum gpfg nok norsk
31	0.178	tiff clouding longerrun mpr ontario recuperation wilkins poloz ottawa omicron

Table 5.2: Top 20 topics correlated with the term $uncertainty$.

"$uncertainty$" terms

Term	Corr.	Term	Corr.
uncertain	0.784	surrounds	0.516
uncertainties	0.779	tensions	0.514
surrounding	0.665	surrounded	0.508
heightened	0.580	aversion	0.507
geopolitical	0.579	cautious	0.507
knightian	0.538	future	0.500
downside	0.530	reaction	0.499
outlook	0.526	unknown	0.498
weigh	0.520	weighing	0.495
prospects	0.517	difficult	0.489

"$uncertainty\ and\ knightian$" terms

Term	Corr.	Term	Corr.
uncertain	0.64	unknowable	0.47
unknown	0.612	probabilistic	0.467
uncertainties	0.606	surrounds	0.459
probabilities	0.558	subjective	0.458
unknowns	0.541	events	0.453
knight	0.512	tail	0.453
probability	0.51	ambiguity	0.452
surrounding	0.493	variables	0.451
aversion	0.488	predict	0.436
unpredictable	0.472	heightened	0.436

Term	Corr.	Term	Corr.
risks	0.702	given	0.534
uncertain	0.625	extent	0.531
uncertainties	0.622	impact	0.531
associated	0.596	surrounding	0.527
heightened	0.585	probability	0.525
may	0.566	appetite	0.519
aversion	0.561	inherent	0.516
unexpected	0.549	volatility	0.515
thus	0.538	perceived	0.512
possibility	0.537	taking	0.511

Table 5.3: Top 20 terms related to $uncertainty$, $uncertainty\ and\ knightian$ and $risk$, and their correlations given by the cosine similarity.

In Table 5.3, most correlated terms to uncertainty are shown. These are primarily linked to a jargon of forecasters' ("outlook", "future", "prospects") and terms with negative connotation ("downside", "tensions", "aversion", "cautious", "unknown" and "difficult"). Both meet a latent description of unknown, unpredictable situation, not related to a probabilistic approach, as for keynesian uncertainty, but to a knightian one.

In discerning the knightian uncertainty types, Table 5.3 give most correlated words respectively with the terms uncertainty+knightian and uncertainty+risk that proved the robustness of the semantic space constructed via Top2vec. Words in TTable 5.3 have a macroeconomic color describing an unknown state, in contrast to words associated to $risk$, which are tied to financial markets and potential spillover impacting them.

Topic	Terms
1	financial , crisis , stability , system , central , risk , global , risks , banks , policy
2	inflation , growth , economic , policy , monetary , year , economy , rate , prices , rates
3	financial , bank , central , sector , banking , development , gentlemen , ladies , economic , us
4	banks , india , sector , financial , credit , per , banking , cent , bank , reserve
5	can , one , us , time , many , now , central , much , even , need
6	policy , monetary , inflation , central , rate , bank , interest , rates , price , stability
7	see , speeches , data , chart , uk , bank , paper , policy , m , available
8	rate , inflation , policy , interest , also , per , cent , prices , riksbank , monetary
9	payment , payments , services , bank , banks , system , new , central , financial , systems
10	bank , economic , japan , percent , economy , prices , rate , financial , firms , growth
11	debt , fiscal , crisis , banks , public , government , measures , economy , economic , sector
12	per , cent , prices , economy , growth , years , new , rate , demand , canada
13	banks , capital , risk , banking , bank , regulatory , basel , risks , supervisory , requirements
14	euro , area , european , ecb , union , countries , monetary , europe , economic , eu
15	federal , reserve , percent , u.s , economic , board , community , financial , fomc , new
16	exchange , countries , global , foreign , international , capital , economies , currency , domestic , trade
17	bank , financial , information , management , institutions , also , public , system , governance , central
18	market , markets , banks , liquidity , financial , credit , bank , risk , securities , assets
19	growth , economic , productivity , economy , labour , can , market , new , also , investment
20	financial , islamic , finance , development , industry , global , also , climate , hong , asia

Table 5.4: Results of a plain LDA showing the learned 10 topics and their most relevant terms.

For each of the 118 central banks in the data set, a country-based uncertainty index can be constructed from the topics uncovered by Top2vec, suited for international comparisons. In Figure 5.1, the correlation of speeches with uncertainty at the Federal Reserve System, as an example of a country-specific index, can be seen. It is tracked using a smoothing curve (LOESS), in-line with business cycles and financial developments in the United States. Uncertainty rose sharply during the 2007 pre-crisis period before receding a few months afterward. Spikes indicate critical events as for the European debt crisis (2010), the presidency of Donald Trump (2017) and the Covid-19 pandemics respectively.

A monthly aggregated international uncertainty index over all central banks, was built as the semantic similarity between each speech and the term uncertainty (Figure 5.2), giving a general overview on uncertainty content at world's central banks over time. Its smoothed curve is in-line with major international economic and finance episodes that occurred during the past twenty years, as for the subprime crisis, the European debt crisis, Brexit and recently the Covid-19 pandemic.

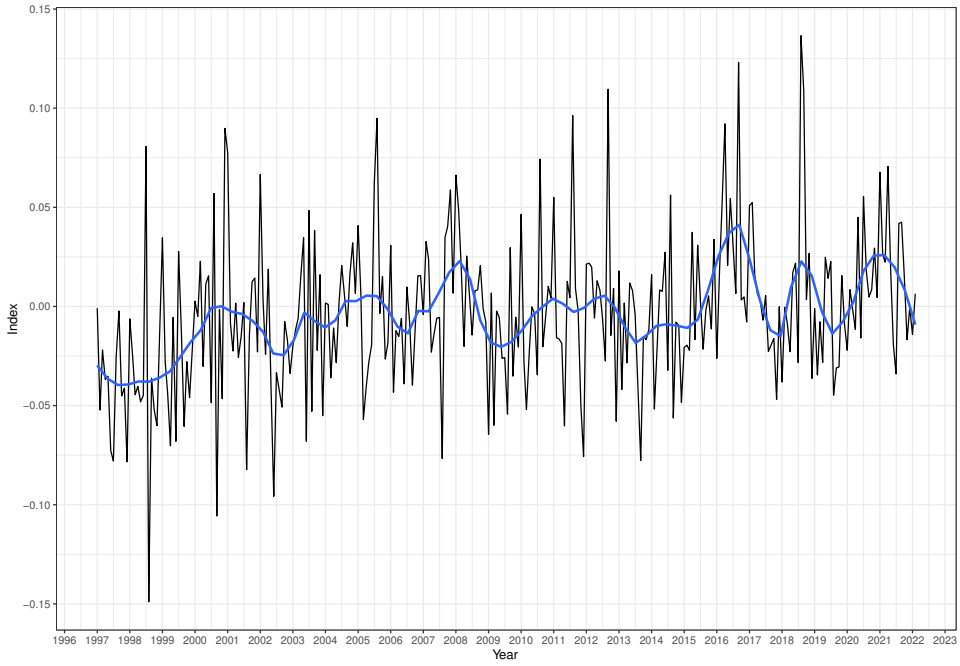


Figure 5.1: Aggregated monthly uncertainty index for the Federal Reserve System (Board of Governors). The blue line indicates LOESS smoothed curve.

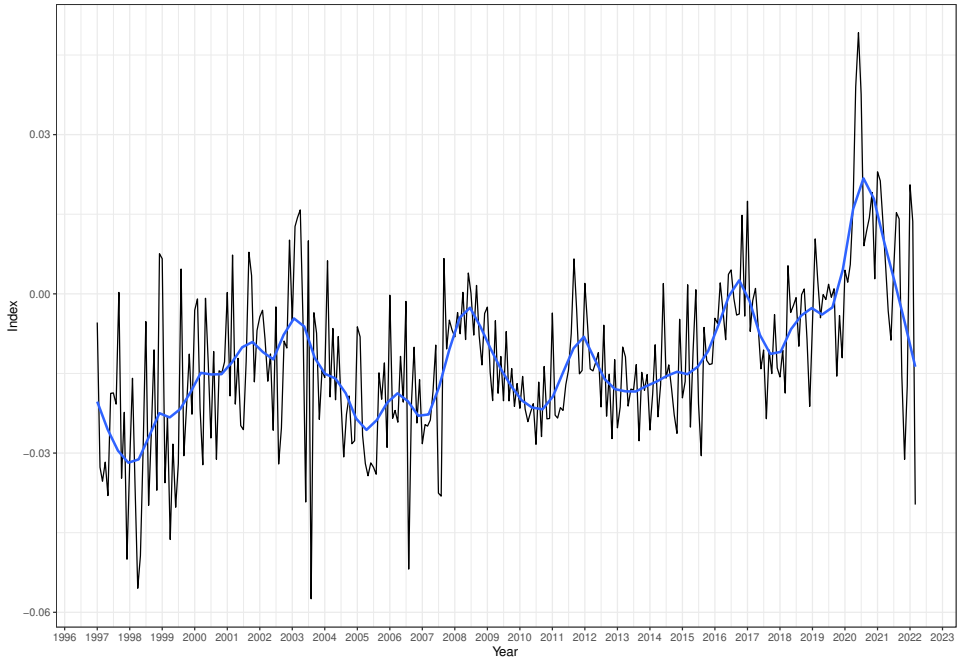


Figure 5.2: Aggregated monthly world uncertainty index on the whole corpus. The blue line indicates LOESS smoothed curve.

5.4 Conclusion

Uncertainty, as a feature of modern monetary policy practices and an adopted communication strategy by central bankers, remains difficult to pin down with usual automated text analysis tools based on word frequencies. Given the implicit characteristic of uncertainty, semantic search models proved able to extract meaningful topics, by learning a joint semantic space of documents and words that successfully identifies sources of uncertainty in a collection of international central bankers' speeches. A distinction between knightian and keynesian uncertainties clearly emerged from the corpus' rhetoric, separating probabilistic uncertainty from risk, the former being tied to a jargon of economic forecasters' jargon and the latter linked to financial markets. Although the corpus is dense, featuring 118 central banks around the world over 25 years, several topics were found. These match recent challenges as for macroprudential and banking supervision, climate concerns and macroeconomic forecasting, aside from the dominance of country-specific topics, due to regular contributions of many central banks, denoting a widespread active communication strategy seen as a continuation of transparency efforts adopted since the end of 1980s. Semantic similarities, as per-document uncertainty scores, proved to be a good proxy to measure uncertainty prevalence, at both the country and international levels. It shows plausible dynamics that follow major economic developments. The rhetoric used by modern central bankers seems to apprehend uncertainty ahead of crises and try to lower its importance in the aftermath, as a way to dissipate confusion and unknown outlooks among economic agents.

Chapter 6

Uncovering causality links in central bank corpora[1]

Abstract

Investigating narratives in specific text-as-data applications has been extensively implemented using hybrid approaches, combining several NLP solutions and statistical techniques. While this direction provides a substantial understanding of key semantic patterns, it remains a challenging task not applicable to all languages. Narratives, especially in social sciences, are often regarded as the study of causality between specific terms, to uncover embedded directed or indirected links for explanation and further hypotheses testing. This paper establishes $CausaMap$, as a semi-supervised method to learn causality between pre-defined terms of interests, built on dependency parsing and the use of tree-based structural equation models to quantify direct and indirect causal links. Thus, testing terms' causality in a directed network of parsed, lemmatized words is equivalent to uncovering informative inner structures with preserved semantic properties, without external information. Applied to two different corpora in English and German languages, $CausaMap$ yielded mixtures of directed acyclic graphs and bidirectional edges, forming a narrative picture linked to monetary objectives and macroeconomic goals reported in the Federal Reserve governors' speeches (1996-2024). The German political parties' manifestos (1990-2021) demonstrated a dominance of policy-oriented terms when studying economic and employment debates, where the relatively high number of uncovered connecting nodes explains the causality buildup of pre-defined terms as well as language-specific morphologies.

6.1 Introduction

Social scientists found in the growing availability of text data a new field of application, whose aim is to extract meaningful textual patterns to explain the rhetoric of given documents. During earlier stages, text mining techniques and machine learning were used to extract meaningful groups of words, namely topics, able to classify documents and discern their content (Deerwester et al., 1990; Blei et al., 2003). Because of the constraining word independence hypothesis, several attempts shifted this interest into the field of natural language processing (NLP) by employing various neural-based solutions to enhance the interpretability of results (Mikolov et al., 2013).

A particular interest was given to exploring a story-based structure called narratives (Ash et al., 2024), considered as an essential feature of social sciences studies as well as an engine in both individual and collective decision-making (Shiller, 2020). In this setting, causality provides context and meaning by

[1]This paper was presented at the 16^{th} IEEE International Conference on Knowledge Graphs, Limassol, Cyprus (November 2025) and published online https://ieeexplore.ieee.org/document/11391993/

successfully connecting separate pieces of information to create narratives (Takayanagi et al., 2024).

Hence, the study of narratives aims to provide measures that capture relationships between elements, for instance entities/characters, concepts/organizations or objects (Ash et al., 2024), which are not found in most text-as-data algorithms used by practitioners (Gentzkow et al., 2019b). An accent has been made on the semantic properties connecting these elements, favoring the use of knowledge graphs as suitable structures (de Kok et al., 2024), encompassing events with spatial and temporal coordinates, in-line with the definition of economic narratives as the causal accounts of past economic events (Andre et al., 2022).

For this aim, several tools have been proposed to extract narratives via complex architectures relying on NLP solutions as for pre-trained language models (de Kok et al., 2024). Ash et al. (2024) built $Relatio$, a narrative extraction technique, using semantic role labeling (SRL) as a building block. The goal is to identify the triplet $agent-action-patient$ in every plain sentence, conditional to the availability of a language-specific SRL. However, most of practitioners still extract $artificial$ narratives without clear causality measurements by combining topic models (Roos and Reccius, 2024), either with their own expertise or with prior information borrowed from other sources, as for word embeddings (Mikolov et al., 2013) and transformers (Devlin et al., 2019), making the analyses prone to biases.

While $Relatio$ proved to be useful when studying political corpora in the United States (Ash et al., 2024) and economic news corpora (Lange et al., 2022), it requires a robust semantic role labeling algorithm for each language, which is not always readily available. Furthermore, several additional pre-processing steps (named entity recognition and event detection) are crucial to perform such tasks, but not properly identified or time-consuming.

The use of dependency parsing has increasingly become popular among practitioners seeking more interpretable output. Due to the grammatically-based syntactic formalism connecting terms, a dependency structure could be defined as a directed graph of words (Kádár et al., 2021), used to improve several NLP architectures as for semantic role labeling and machine translation. This requires defining robust dependency parsers (Straka et al., 2016), usually trained with fast prediction-based neural network classifiers (Chen and Manning, 2014; Straka, 2018).

As part of the narrative economics wave (Shiller, 2020), central banking communication provides examples of interests in understanding monetary practices around the world and test several hypotheses regarding their functioning and the impact of their communication on economic agents. Such corpora are regarded as a complex output of an indirect rhetoric, embedded with some properties, as for ambiguity and consensus, making the analysis quite difficult in comparison with other corpora.

This acknowledgment makes it arduous to analyze similar corpora from a simple text mining perspective and requires, as for other political corpora, advanced strategies to test causality-driven hypotheses that go beyond the classical topic models adopted by many researchers (Roos and Reccius, 2024). Hence, the need to uncover hidden causal patterns calls advanced semantic transformations coupled with network representations, to give a better understanding of interactions between terms of interests.

Hence, this paper proposes a knowledge graph approach to mine causal links by revealing most important features predicting particular terms of interest. It fits the definition of $collective\ economic\ narrative$, characterized by containing a cause-effect relation (Roos and Reccius, 2024). The scheme uncovers semantic word associations, by identifying robust directed links between specific terms using a tree-based structural equation model. The latter is a method that been employed in biology

on protein–protein interaction (PPI) data (Grassi and Tarantino, 2023) to reveal process-specific information and can be applied to text data as well. Dependency parsing genuinely preserves the semantic properties of the corpus via the edges connecting terms in the resulting representation graph (Kádár et al., 2021), yielding a map of causality links regarding user-specified terms.

The result, namely $CausaMap$, is a semi-supervised knowledge graph summarization that discerns causal links of selected terms and displays a narrative picture encompassing the standard 4W aspect (who, what, when, where) used by many NLP solutions (Ash et al., 2024) through dependency parsing (Straka et al., 2016). Causal links between nodes are semantically grounded and learned via structural equation models (Grassi and Tarantino, 2023) to give general-level narrative understanding without a framing analysis (Entman, 1993), so to better capture the complexity of term-term associations without requiring information from borrowed corpora or other technologies as for Large Language Models (LLMs). The availability of dependency parsers in many languages allows the application of $CausaMap$ to all languages covered by the Universal Dependencies treebanks (Straka, 2018).

Tested on two corpora, this $narrative\ map$ proved to be genuine in revealing and quantifying causality links between specified terms of interest in central banking and political manifestos. Selected terms related to the macroeconomic jargon yielded a causal graph articulated around monetary objectives assigned to the Federal Reserve in the United States. It confirms the preponderance of macroeconomic analysis articulated around interest rate and inflation targeting. A policy-oriented rhetoric, in terms of political pledges, dominated the debate over the economic status and employment policy in the post-reunification German political manifestos.

Significant causal links, ranked by their intensities, might be used for other tasks as semi-supervised topic models (Eshima et al., 2023; Diaf, 2024) and acting as a transfer learning scheme for other term-specific applications.

The paper outlines the related text-as-data applications (Section 6.2) and details the causal-tree approach used by the structural equation model (Section 6.3), to explain the $CausaMap$ methodology (Section 6.4) before conducting causality experiments on two selected corpora in English and German (Section 6.5).

6.2 Text-as-data

The assumption of word independence (bag-of-word hypothesis) has been the dominant paradigm in modern text analysis. Most of the existing methods used to study text data overlooked the semantic structure linked to specific interests. This usually yielded lower output representations, usually interpreted with the authors' experience, or augmented with post-hoc adjustments to improve the quality of results (Eshima et al., 2023).

Topic models, as a subclass of machine learning algorithms, has been extensively used to extract informative patterns from corpora, using groups of words assumed to form semantic patterns, known as topics (Blei et al., 2003). Because of the probabilistic inference, topics are defined as probability distributions over words, rather than semantic features (Blei and Lafferty, 2006). This mostly generates agnostic output, known as one weakness of the bag-of-word assumption related to most text mining techniques.

The adoption of neural networks shifted the interests of the community from text mining to NLP (Mikolov et al., 2013) and many architectures have been adopted as a de facto choice or

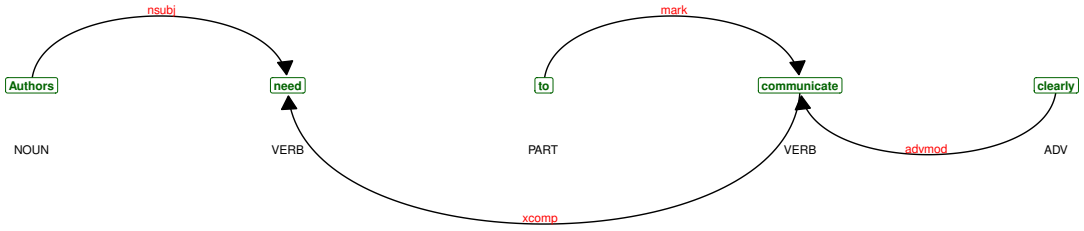


Figure 6.1: Example of a dependency parsing output (tokenization, part of speech tagging and dependency relations) from a simple sentence, using UDPipe (Straka et al., 2016).

pre-processing steps when analyzing texts. This is due to their ability to encompass semantic patterns and overcome the bag-of-word hypothesis.

Recent AI-based pre-trained models showed improvements in many application fields but still considered biased because trained on heterogeneous sources, causing intrinsic limitations (Brown et al., 2020). A particular trend sought to investigate causal capabilities of large language models (Takayanagi et al., 2024) including non-English datasets. It found some language models falsely recognize non-causal sequences as causal sequences, while still facing challenges in efficiently leveraging in-context learning and domain adaptation.

Particularly, the growing interest in more interpretable results led researchers to adopt network analysis and related tools to investigate causality as directed acyclic graphs (DAG) structures (Pearl, 2009), often used as roadmaps to illustrate causal narratives. Network structures, as products of several hierarchical clustering schemes including communities (Blondel et al., 2008; Traag et al., 2019) and cliques (Derényi et al., 2005), have been employed in several text mining applications (Diaf, 2023, 2024) due to their informative features.

Advanced NLP solutions often use dependency parsing (Eisner, 2000) as an intergrated task and researchers tried to establish a graph-based approach (McDonald et al., 2005), as exemplified in Figure 6.1, encompassing several transformation steps (tokenization, part of speech tagging) to retrieve inner terms' interactions. This endows the solutions with granular associations that help achieving a higher accuracy in several tasks as for topic models and translation. For instance, the following links from Figure 6.1: $author \rightarrow need$, $communicate \rightarrow need$ and $clearly \rightarrow communicate$ were retrieved from dependency parsing and are assimilated as grammatically-based, causal links. The link $to \rightarrow communicate$ could be withdrawn from the analysis as the term "to" occurs as a stopword.

6.3 Causal trees

The presence of complex data structures need advanced statistical techniques for causality measurement, which was mostly conducted using simple regressions. The complex linkage between dependent and independent variables could be translated into a set of equations requiring a simultaneous estimation (Duncan, 1975). In this perspective, Grassi and Tarantino (2023) developed a variant of structural equation model, named $SEMTree$, to identify functional modules in the protein–protein interaction (PPI) networks, by tracking changes in molecular activity or phenotypic signatures. It employs a dual graph and data-driven algorithms to recover causal trees based on structural equation modeling Lou et al. (2024), used as a proxy of non-linear regressions.

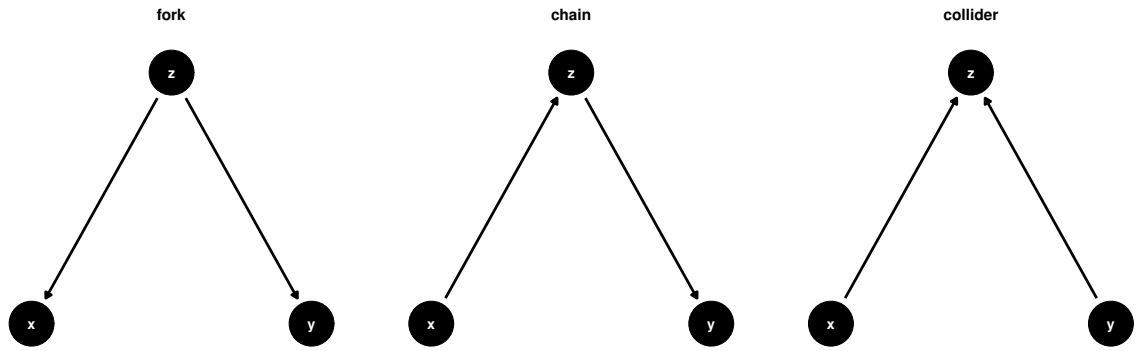


Figure 6.2: Synthetic examples of three causal patterns (McGowan et al., 2024), depending on the directions and the links between nodes.

Figure 6.2 shows three distinct types of causal patterns usually found in practice (McGowan et al., 2024), depending on the causality directions and links between the nodes. When z simultaneously causes x and y, the network is called $fork$ and z is the confounder. In the $chain$ pattern, z acts as a $mediator$ between x and y, in what is described as a $chain$ network; while z is seen as a $collider$ if it is caused by both x and y.

SEMTree recovers a directed or indirected Tree $T = (V, E)$ with p nodes (V) and $p - 1$ edges (E), having a unique path between any two nodes $(j, k) \in V$.

A directed acyclic graph (DAG) is usually used to understand the causality link between nodes. Hence, investigating the hypothesis that the variable Y_k causes the variable Y_j turns into uncovering a DAG $G = (V, E)$ with V as the set of nodes (variables) and E as the set of edges (connections). A DAG could be represented as a structural equation model with two components: Y_j and U_k, respectively an observed variable (a pre-defined term) and an unobserved error term. Assuming $pa(j)$ as the parent set of Y_j, β_{jk} is the weight of $Y_k \rightarrow Y_j$ and independent error terms $U_j \sim N(0, \sigma_j)$ for all $j \in V$, we can write:

$$Y_j = \sum_{k \in pa(j)} \beta_{jk} Y_k + U_j \tag{6.1}$$

This graph method, known as Steiner Tree (ST), aims to recover a tree from a full graph. It connects the nodes of interests to additional nodes in the $most\ compact$ way possible (Grassi and Tarantino, 2023) based on a very fast solution provide provided by the Kou algorithm (Kou et al., 1981); while the edge weights are, by default, based on the pairwise correlation $1 - |cor(j, k)|$ between the nodes j and k, whose significance is greater than a given threshold. ST method was found to have a great ability in retrieving sub-networks of interest, when compared to other schemes (Grassi and Tarantino, 2023).

Other learning algorithms have been adopted by SEMTree, as for directed trees (CAT) (Jakobsen et al., 2022), also known as arborescence or directed rooted tree, with the goal to uncover a connected DAG in which all nodes have a unique parent, except the root node. The CAT variant of SEMTree is seen as a reduced complexity model and constrained, if compared to ST, because CAT arborescence structure omits potential nodes whose presence in the graph is informative in identifying non-direct causal links.

The ST variant, at the contrary of CAT, permits to uncover connecting nodes. These are terms not pre-

defined, whose presence help enriching the narrative map via their additional causal linkage. From their location in the yielded graph, they can explain different causal patterns (fork, chain or collider) depending on their context ($pa(j)$ and Y_k in equation 6.1) and guiding the interpretation of the map, especially if key terms are omitted from the pre-defined list.

6.4 Methodology

Dependency parsing, employing the associated language parser (Straka et al., 2016), is used to get grammatically-linked terms of the corpus, in the form of bigrams of lemmatized terms, where stopwords are excluded from the analysis. The resulting bigram network creates a directed graph g as well as a feature co-occurrence matrix fcm, counting the occurrences of bigram links retrieved by the parser.

The resulting sparse fcm matrix is then transformed into a more Gaussian feature-matrix[2] (Zhao et al., 2012) to facilitate the computation of the structural equation model (Grassi and Tarantino, 2023), given a set of pre-defined terms T to test their causal linkages.

This semi-supervised architecture guarantees a preserved semantic embedding space, to be given as input graph without loss of information or the need of biased external source. Hence, the causality between the pre-defined terms is automatically semantic and does not require any further transformation, prior information in terms of language models or word embeddings. In other words, causality edges reflect the natural structure of the corpus' own narrative rhetoric, not altered by pre-trained corpora that might reinforce artificial non-causal links (Takayanagi et al., 2024).

For instance, if the terms $inflation$ and $targeting$ are chosen as pre-defined terms in monetary policy corpora, we need to assess the direct causal links $inflation \rightarrow targeting$ as well as $targeting \rightarrow inflation$, in addition to potential indirect links $inflation \rightarrow$C$\rightarrow targeting$ and $targeting \rightarrow$C'$\rightarrow inflation$, where C and C' are sets of connecting nodes denoting the indirect causal paths that could exist between $inflation$ and $targeting$.

Algorithm: CausaMap

---

1. Dependency parsing: Run $dependency\ parsing$ algorithm $UDPipe$ (Straka, 2018) over all sentences of the corpus and extract joint lemmatized terms as bigrams.
2. Network analysis Use the lemmatized bigrams to create a directed graph g and the associated feature co-occurence matrix fcm
3. Tree-based structural equation model: Transform the resulting fcm using Gaussianization (Zhao et al., 2012) to help relax the assumption of normality and use the output, together with g and a set of defined terms T to learn a causality graph cg using SEMTree (Grassi and Tarantino, 2023).

6.5 Application

Two corpora in different languages were chosen to test $CausaMap$ with selected pre-defined terms related to monetary policy in the United States (1996-2024) and the political debate over the economic status and the labor market in the German political manifestos (1990-2021). The goal is to test the efficiency of $CausaMap$ in retrieving robust causal graphs and assess their strength, when using a set of related terms given as priors.

[2]using HUGE (Huge-Dimensional Undirected Graph Estimation) method

6.5.1 Federal Reserve speeches

The corpus of the Federal Reserve (Fed) governors' speeches was scraped from the official website[3] and comprises 1,699 speeches given by different governors from 1996 to 2024. Speeches were first parsed to get the lemma-based dependencies using a pre-trained model for the English language from UDPipe (Straka, 2018). It resulted a network of bigram dependencies and its related Gaussian co-occurrence matrix, to be used as inputs for the SEMTree estimation, given prior terms showed in the subsequent graphs in green circles. Causal links are estimated using SEMTree, on the basis of significant[4] pairwise correlations between every two nodes.

Figure 6.3 shows the causality map of fourteen (14) selected terms (green nodes) related to classical macroeconomic jargon used by central bankers. The aim is to study causal links between given terms and their intensities represented by causality coefficients (β_k learned from equation 6.1). The bidirectional causality between $inflation$ and $targeting$ constitutes one pillar policy used in the central banking, confirmed by the highest causal coefficient (1.24) in the graph, while a neighboring bicausal link is less pronounced ($inflation$ and $price$). The lemmatized term $price$ comprises the often occurred term $prices$ in the corpus; an acts as a collider for the terms $interest$, $forecast$ and $inflation$, while causing the term $firm$ which is a collider of three terms ($rate$, $price$ and $analysis$). $Firm$ has a pivotal role connecting two subnetworks as shown in Figure 6.3, to its right are technical terms used by central bankers, while the left subnetwork is more policy-oriented and contains terms related to classic monetary policy practice.

Three connecting nodes, shown in white circles in Figure 6.3, were retrieved ($need$, $firm$ and $policy$). They act as bridging nodes between the pre-defined terms. For instance, the position of the node $need$ is pivotal in Figure 6.3, and could be seen as the parent root node causing $growth$, job and having a mutual causality with $rate$.

The application of Relatio (Ash et al., 2024) on the same corpus yielded the graph shown in Figure 6.4 where the narratives are built upon the agent-action-patient triplet, taking into account the top 50 terms linked to each pre-defined terms chosen in CausaMap (green circles in Figure 6.3). Although Relatio output is more informative in terms of connecting nodes, it lacks causality measurements and is embedded with detailed, complex linkage that make further causality estimation via structural equation models expensive. For instance, Relatio map is heavily built around the term $inflation$, while other terms (growth, price, rate) play a second role in the graph.

6.5.2 German parties' manifestos

Manifestos of German political parties are available at the $Manifesto\ Project$ (Lehmann et al., 2024), where political programs during the period 1990 to 2021 were used as a corpus of election interests and orientation after the German reunification. The corpus underwent the same procedure described in the previous subsection, at the exception of using a German pre-trained model for dependency parsing. The main interest is to extract narratives related to the economic status and labor market in Germany, assumed to be regular topics that continue to fuel debates and trigger reforms to maintain an overall competitiveness, amid changing national and international environments.

Figure 6.5 shows the causality map of 13 pre-selected terms and their interactions linked to the economic conjuncture ($Konjunktur$). Due to rhetoric used by political parties, the term $brauchen$ (to need) highlights the importance of several aspects raised by politicians as for the European Union (EU), investments ($Investition$), small and medium-sized companies ($Unternehmen$), education

[3] http://www.federalreserve.gov/newsevents/speeches.htm

[4] Signficance threshold is taken as α=0.05

($Bildung$) and security ($Sicherheit$). This community is central to the graph shown in Figure 6.5 and related to another community dealing with life aspects ($Leben$), digitalization ($Digitalisierung$) as well as liberalism ($Demokratie$ and Marktwirtschaft). The term security ($Sicherheit$) seems to be pivotal as it acts as a collider of two main DAG components ($Leben$, $Wohlstand$) in Figure 6.5. Promises to restore growth and enhancing competitiveness form a separate DAG, where the term more ($mehr$) is an natural intensifier for growth ($Wachstum$) and competititveness ($Wettbewerb$).

Seven terms related to the labor market were selected to test $CausaMap$, as shown in Figure 6.6. The narratives display the need ($brauchen$) of education ($Bildung$), population ($Mensch$) as well as skilled workers ($Personal$). It required eight connecting nodes to uncover the causal links between seven pre-defined terms. Such findings are mostly found when setting a limited number of pre-defined terms, or when the selected ones do not form strong directed causal links, requiring additional connecting nodes to achieve a causal edge to specific terms.

The nodes $Integration$ and unemployment ($Arbeitslosigkeit$) are somewhat insulated from other pre-defined terms, as they need respectively two and three connecting nodes to get linked to the overall graph. Integration bears a social ($Sozial$) dimension (causal link 0.91), while education ($Bildung$) is seen as a vector of development (causal link 0.78).

Both applications in German language do not yield bicausal representations ($inflation$ with $targeting$; $interest$ with $rate$ in the Federal Reserve corpus) as compared to the English language. Although the scope of the applications is totally different, this finding could be explained by cross-linguistic differences in the languages (Vikner, 2019), which are responsible of generating a higher number of connecting nodes, necessary to link pre-defined terms. Bicausal representations stress out the occurrence of both terms in the same context and their higher informative power, suggesting a specific use as keywords to retrieve meaningful topic representations (Eshima et al., 2023), or as simple wordlists.

6.6 Conclusion

The study of narratives is seen as the continuation of the revolution that first started with the adoption of statistical methods for textual data. Since then, a constant improvement was brought by the incorporation of machine learning techniques and neural network architectures to give birth to what is known as natural language processing. In parallel, the interest of linguistics and other researchers has been on uncovering causal patterns that better explain interactions between the units of analysis. This was often investigated using hybrid approaches, whose application is constrained to specific languages. This work brought $CausaMap$ as a semantically-based, mapping algorithm that investigates causality between given terms of interest. It leverages the semantic power of dependency parsing and structural equation models to uncover narrative patterns as causal links between selected terms. This proved able to track causality between selected terms, unfolding connecting nodes and assessing the intensity of causal links. Applied on two different corpora in English and German, $CausaMap$ confirmed the macroeconomic essence of central bankers' speeches in the United States and determined key causal structures in terms of independent directed acyclic graphs, oriented toward macroeconomic analysis and monetary policy practice. On the political manifestos issued by German parties, CausaMap uncovered distinct links between social pledges and the labor market, while the debate on the economic status revealed additional nested nodes offering extra information on causality formation. The causality between pre-defined terms depends on the rhetoric used in the documents and the narrative map offers insights on the build-up of causal links, whether direct or indirect. Connecting nodes, in the context of Steiner Trees, have the key role of bringing extra information for causality formation and guide indirect causality paths between selected terms.

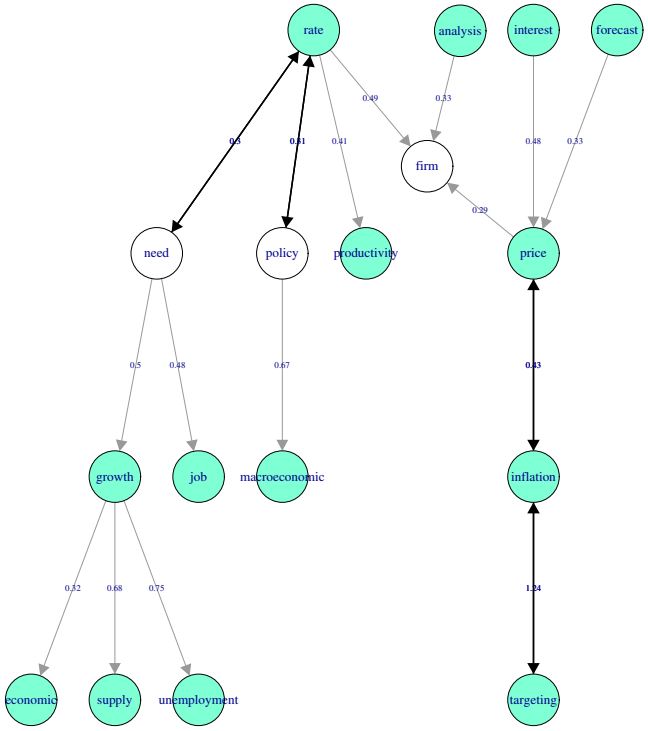


Figure 6.3: Causality of some macroeconomic terms in the Federal Reserve governors' speeches corpus (1996-2024). Pre-defined words are colored in green, while other connecting nodes are in white.

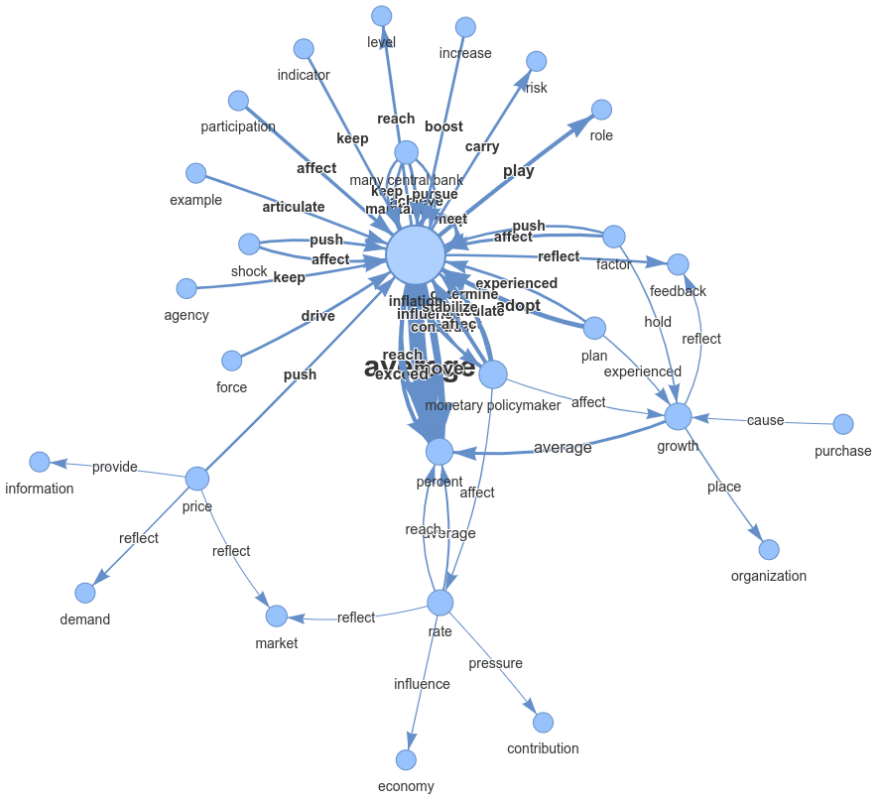


Figure 6.4: Output of *Relatio* using the same pre-defined terms as in Figure 6.3

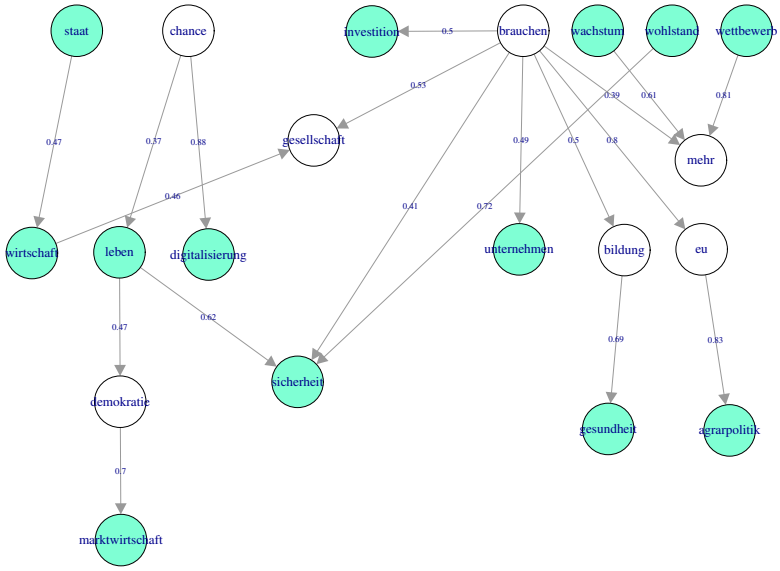


Figure 6.5: Causality of economic-related words in the German parties' manifestos (1990-2021). Pre-defined words are colored in green, while other connecting nodes are in white.

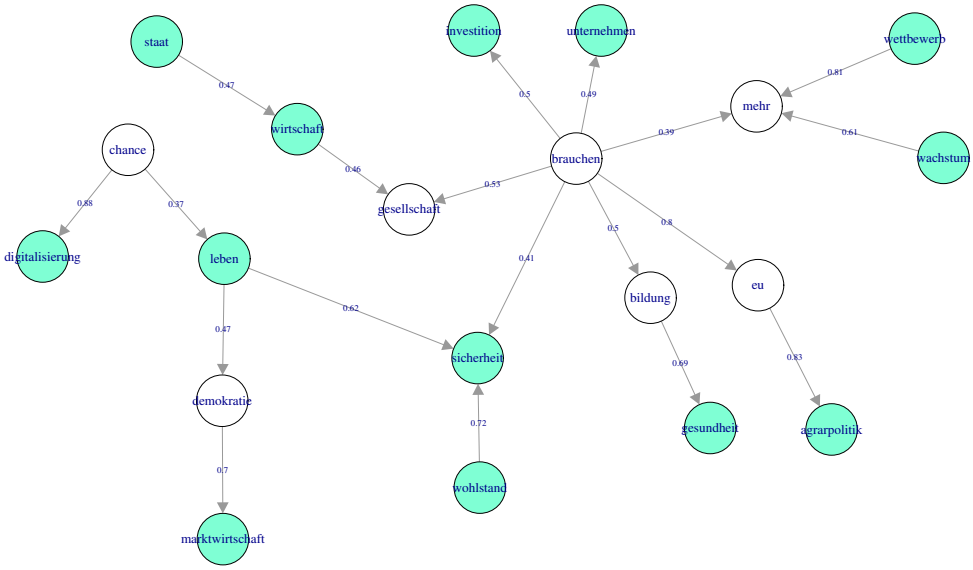


Figure 6.6: Causality of labor market-related words in the German parties' manifestos (1990-2021). Pre-defined words are colored in green, while other connecting nodes are in white.

Chapter 7

LLMs in central bank communication[1]

Abstract

The diagnostic of neural networks, particularly Large Language Models (LLMs), came to scrutiny via the use of random matrix theory to unlock per-layer inner correlations as proxies of fitting issues. Elements from statistical mechanics were used over the layers' eigenvalues to compute several power-law based metrics, which proved successful in assessing the overall LLMs performance without the need to access training and test data. Central banking communication, as a popular text-as-data application domain, witnessed several contributions, in terms of built LLMs. While this helped executing tasks ranging from sentiment analysis to policy preferences, limitations in the available corpora and the adopted transfer learning schemes still impede the accuracy of such models. This works provides a diagnostic of two released LLMs for central banking communication and investigates the inter-layer statistical properties to assess the $bona\ fide$ of such models. On two selected LLMs trained over central bank corpora, PL-metrics were found to have different behaviors depending on the task complexity. Findings confirm the relative importance of the size scale in the PL-metrics, at the contrary of what was found in most LLMs. This highlights the issue of training LLMs on a relatively small subset of data and limitations of building big architectures on small classification tasks, often requiring a substantial expertise.

7.1 The use of LLMs in central bank communication: too deep to fail ?

7.1.1 Introduction

Neural networks transformed the field of machine learning and permitted to execute many tasks and reach a higher accuracy, in comparison to statistical learning models used by practitioners. Deep Neural Networks (DNNs) have been the go-to solution for many fields of applications, as their performance was incredibly higher than most machine learning algorithms used by practitioners.

The text-as-data field knew several shifts during the last three decades, correlatively with the development of computing capabilities. Earlier text mining approaches were mostly based on word counts, while machine learning algorithms tried to uncover topics as hidden patterns in corpora. Natural Language Processing (NLP) concentrates most of today's applications and benefits from several advanced architectures, as for Large Language Models (LLMs), to achieve complex tasks. The particular use of memory processes, as for attention mechanisms, enhanced the overall performances of LLMs, which are currently the go-to solution for NLP tasks.

[1]Section 7.1 is a follow-up of a general application that was presented at the 8^{th} International Conference on Natural Language and Speech Processing (ICNLSP-2025), Odense, Denmark. https://aclanthology.org/2022.konvens-1.20

Besides the extensive use of text mining in social sciences, the field of central banking communication witnessed a growing interest of researchers trying to test hypotheses linked to central bankers and monetary committees. First attempts used sentiment analysis (Blinder et al., 2008; Gentzkow et al., 2019b) to check the responsiveness of financial markets to central bankers' addresses. Later, practitioners tried to uncover meaningful monetary-oriented themes, or topics, from large corpora of speeches and transcripts using topic models. While the later subclass of machine learning (Blei et al., 2003) is praised for its intuitive topic construction, it remains an agnostic application because of the bag-of-word hypothesis. This led to adopt in-depth semantic search schemes (Angelov, 2020; Diaf and Schütze, 2024) and the use a multitude of NLP architectures as for LLMs, to test partisanship and sentiments of central bankers.

The limited availability of central bank corpora, in comparison to other fields of research, pushed practitioners to rely on feature engineering, as for pre-trained models and transfer learning. Moreover, monetary policy corpora show distinct stylized facts, as for ambiguity and consensus (Meade, 2005; Baerg, 2020), making the text analysis more difficult than traditional corpora. This denotes a special attention that should be given to central banking texts, in addition of specific cultural features as for multilinguality (ECB, 2011).

Baerg and Binder (2024) explored the emotion/sentiment nexus and found LLMs outperform traditional text mining tools when uncovering emotions in central bank corpora, but warned against a massive use prone to bring misleading estimations.

If LLMs are regarded today as a solution to several NLP tasks, their diagnostic is often ignored by practitioners. The lack of training and test data used at the development stage prevents users from conducting tests and further robustness checks.

Recent advances in statistical learning suggested the use of Random Matrix Theory (Martin and Mahoney, 2021) to provide a diagnostic of DNNs based on power-law (PL) metrics without the need to access training data. Such schemes proved to be reliable when studying a large collection of LLMs, as they uncovered several training issues due to the use of attention mechanisms (Martin et al., 2021). Identifying layers with underfit/overfit helps fine-tuning the model without adopting complex, yet expensive hyperparametrization and feature engineering strategies.

This paper follows the strategy adopted by Martin et al. (2021) to test DNNs and check the robustness of available LLMs for central bank communication, namely $CentralBankRoBERTa$ (Pfeifer and Marohl, 2023) and $ECBERT\text{-}base$ (Gorodnichenko et al., 2023). The latter is a deep learning model developed to detect emotions embedded in Federal Open Market Committee (FOMC) press conferences, while the former is a dual classifier for economic agents and sentiment analysis.

Results show contrasting performances of PL-metrics (weighted α and log α norm) used to assess LLMs, while the scale size appears to have a big amplitude that affects PL-metrics compared to the information correlation. For CentralBankRoBERTa, there are no big differences between the simple α metric and the PL-metrics, however, the α remains a lower bound and not highly variable as found in most LLMs (Martin et al., 2021). On the other hand, ECBERT-base demonstrates noticeable distortions that could be interpreted as several scale collapses, stemming from data and the use of attention mechanisms.

The paper outlines the use of Random Matrix Theory in DNNs analysis (subsection 7.1.2), then details particularities of central banking corpora as well as the guidance researchers adopted so far (subsection 7.1.3). Subsection 7.1.4 features the analysis of two LLMs built upon central banking

corpora, using the $WeightWatcher$ tool (Martin et al., 2021), so to investigate bottlenecks and fitting issues often encountered in DNNs.

7.1.2 Random matrix theory

Machine learning algorithms have been dominated by the traditional train/test paradigm. The advent of neural network as heavy data consumers render this approach quiet difficult to implement for several reasons.

Train and test data have been the de facto tools to assess machine learning models in general, and neural networks in particular. In the absence of such data, elements from Random Matrix Theory were applied on weight matrices of neural networks (Martin and Mahoney, 2021). It resulted several norms and metrics, whose statistical properties match DNNs accuracy, without accessing data used to train the models (Martin et al., 2021). In other terms, this strategy permits to discover whether a layer learned too much from the noise (overfitting) or alternatively hasn't learned enough from the signal (underfitting), assuming data stem from two components: signal of noise.

The WeightWatcher tool (Martin et al., 2021) investigates the weight matrix W of each DNN layer, by analyzing its spectral properties. While every element of the weight matrix[2] W_{ij} is assumed to follow a normal distribution $\mathcal{N}(0, \sigma^2)$, the empirical covariance (Wishart) matrix $\boldsymbol{X} = W^\intercal W$ is taken as the basis for quality assessment of state-of-the-art DNNs, by extracting its eigenvalues spectrum. Alternatively, the correlation matrix does fit to the MP assumptions and is used to extract eigenvalues.

The Marchenko-Pastur distribution (Marchenko and Pastur, 1967) considers the spectrum of eigenvalues bounded between λ_- and λ_+ as relevant to the noise randomness. Its probability density $f(\lambda)$ is given for a $T \times N$ matrix and a noise level σ^2 as:

$$f(\lambda) = \begin{cases} \frac{N}{T} \frac{\sqrt{(\lambda_+ - \lambda)(\lambda - \lambda_-)}}{2\pi\sigma^2} & \text{if } \lambda \in [\lambda_-, \lambda_+] \,, \\ 0 & \text{if } \lambda \notin [\lambda_-, \lambda_+] \,. \end{cases}$$

where $\lambda_- = \sigma^2(1 - \sqrt{\frac{T}{N}})^2$ and $\lambda_+ = \sigma^2(1 + \sqrt{\frac{T}{N}})^2$
Because many matrices hold strongly correlated elements, the MP distribution is used to empirically evaluate a noisy spectrum of eigenvalues, that could be separated from other eigenvalues representing the signal.

Martin and Mahoney (2021) found most matrices in DNNs exhibit extreme eigenvalues, suggesting rather drawing elements from a power-law distribution, as for Pareto. This concept, known as Heavy-Tailed Self-Regularization (HT-SR) theory, is linked to situations where separating the noise from the signal becomes difficult to achieve, as eigenvalues are in this case better modeled via heavy-tailed distributions (Malevergne and Sornette, 2004), rather than a simple MP distribution.

For this aim, Martin and Mahoney (2021) estimated a truncated power-law fit (Clauset et al., 2009) over the MP curve, yielding the exponent α from the equation density-eigenvalues: $\rho(\lambda) = \lambda^{-\alpha}$ for $\lambda \in [\lambda_-, \lambda_+]$. The amplitude of the PL-exponent α is considered as the $information\ correlation$ index within each matrix, denoting the strength of the element-wise existing correlations. Moreover, the α exponent is a result of a power-law link that could be assimilated to a complexity index or a $fractal\ dimension$ (Mandelbrot, 1982).

[2]Layers having multiple weight matrices will have a concatenated weight matrix Martin et al. (2021).

Based on the eigenvalues spectrum λ_i of each correlation matrix $\boldsymbol{X}$, several metrics were used as for:

- Frobenius norm : $\|W\|_F^2 = \|X\|_F = \sum_{i=1}^{M} \lambda_i^2$
- Spectral norm : $\|W\|_\infty = \|X\|_\infty = \lambda_{max}$
- Weighted α : $\hat{\alpha} = \alpha Log \lambda_{max}$
- α norm (Shatten-norm) : $\|W\|_{2\alpha}^{2\alpha} = \|X\|_\alpha^\alpha = \sum_{i=1}^{M} \lambda_i^\alpha$

where λ_i is the i^{th} eigenvalue of $\boldsymbol{X}$, λ_{max} is the maximum eigenvalue and α is the fitted power-law exponent, usually truncated because it needs defining specific lower and upper bounds, respectively λ_- and λ_+. For instance, Figure 7.1 reports simulations yielding random-like eigenvalues fitted with a scale-invariant Marchenko-Pastur curve between λ_- $\simeq$0.18 and λ_+ $\simeq$1.37 and spikes (signal) associated with $\lambda_i > \lambda_+$. The PL-fit yields a value of 0.42 for α.

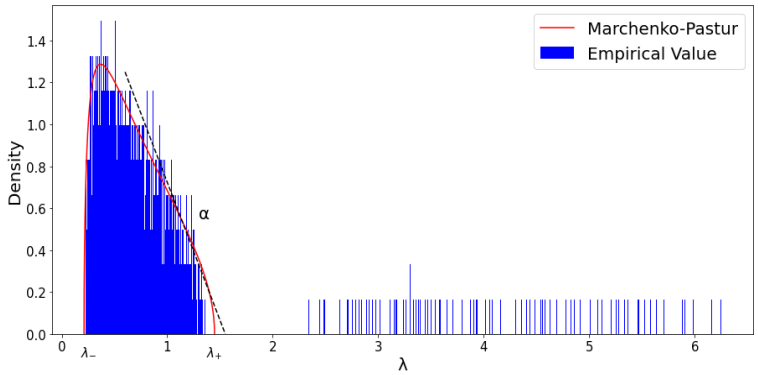


Figure 7.1: Marchenko-Pastur distribution simulated 1,000 times on the correlation matrix of an initial random matrix with $\frac{T}{N}$=5 and $\sigma^2 = \frac{2}{3}$. α is the PL-exponent of the Marchenko-Pastur fit over the interval [λ_-,λ_+].

The plain α metric is a scale-invariant, weak estimation of the information correlation, as it ignores the scale size (λ_{max}) within each layer. The latter remains an important determinant of HT-SR because DNNs are known to be non-linear, where LLMs particularly feature attention layers with large matrices. For small values of α, the scale size λ_{max} was found to be a good proxy for estimating the difference between the noise and the signal, while for higher α values (HT-SR), the signal gets mixed with the noise and λ_{max} is non-informative.

A clear distinction between norm-metrics and PL-based metric was given when studying the performance of several DNNs models (Martin et al., 2021). They concluded that PL-based metrics, aside from being good proxies for accuracy measurements, remain robust in detecting potential bottlenecks, or scale collapse than norm-based metrics. Hence, PL exponent remains a robust empirical metric to asses well-trained DNNs and quantify the layer-wise correlation flow (Martin et al., 2021).

In practice, α was found to match an ideal DNN fit when equaling 2. This means the DNN model performs well as it facilitates the propagation of information/features across layers (Martin et al., 2021), because it learns from both data signal and noise. Higher values in the interval]2,6] are proxies of underfit situations (not learning enough from the signal), while lower values below 2 are synonyms of overfit (learning too much from the noise). Large values of $\alpha > 6$ are associated with a

pure randomness.

Because the size of DNNs layers changes according to adopted architectures, Martin et al. (2021) proposed to weight the α with the scale size to produce the weighted α metric. It was found that for small α values, the weighted α approximates well the α Shatten-norm (Martin et al., 2021); the latter weighs the α exponent for all eigenvalues within the layer.

Martin et al. (2021) reported that weighted α and log α norm correlate at high level for well trained models and the size scale could be informally linked to situations where sentence parts are at a greater distance in a given text. This means the scale size is related, in the case of LLMs, to the language semantic structure.

The PL-exponent metrics help identifying layers where training underwent issues, that translate later into bottlenecks in form of overfit and underfit. Particularly in LLMs, high variability in the series of PL exponents is called $scale\ collapse$, mostly linked to transformers (Vaswani et al., 2017; Lefaudeux et al., 2022). As memory-based blocks of layers, transformers feature a complex inner structure usually yielding larger weight matrices.

7.1.3 Central bank communication

The transparency efforts, led by major central banks around the world, spread information supports as a communication strategy, whose aim is to reduce the secrecy surrounding central bank practices (Goodfriend, 1986).

First contributions to this field of research focused on sentiment analysis and financial markets (Gentzkow et al., 2019b), before witnessing a growing interest in applying novel solutions in machine learning, mainly to for topic detection and further econometric inferences.

Topic models became the standard unsupervised strategy, as for the Latent Dirichlet Allocation (Blei et al., 2003), used on central banks' minutes and transcripts (Hansen et al., 2018), while other attempts sought a guided approach based on keywords (Diaf, 2022) to better interpret the results and avoid post-hoc adjustments (Eshima et al., 2020).

The advent of neural networks introduced researchers to NLP as a reliable technology with a better interpretability of the results. Semantic search strategies (Angelov, 2020) offers a genuine blend of three unsupervised learning methods for topic detection, which was later expanded by Grootendorst (2022) for Bertopic. Examples of application sought the estimation of latent traits of central bank communication, as for uncertainty (Diaf and Schütze, 2024).

The introduction of LLMs was hailed as a revolution in NLP and was used to execute several tasks ranging from sentiment analysis, agent classification (Pfeifer and Marohl, 2023) to policy-oriented preferences (Gorodnichenko et al., 2023).

The accuracy given by these models do not often match the applications conducted by researchers, especially on newly released documents that were not part of the training data. The proved complexity of central bank corpora (Meade, 2005; Baerg, 2020) is another hinder that differentiates such documents from usual text data.

The absence of testing strategies for developed LLMs remains problematic when pronounce over the good faith of such technologies. Several schemes come into play when developing LLMs as for

feature engineering, data augmentation and transfer learning. The latter bears a particular importance because it distinguishes $normal$ corpora from specific ones.

The WeightWatcher tool developed by Martin et al. (2021) summarizes the theoretical foundation tackled in Section 2 and offers a strong basis to assess any trained DNN, without requiring train and test data. This pertains to LLMs and their inner functioning, featuring complex and imbricated structures as for attention mechanisms (Vaswani et al., 2017; Lefaudeux et al., 2022).

7.1.4 Application

Selected LLMs related to central bank corpora are investigated using PL-metrics adopted by Martin et al. (2021). They were previously found to be robust when assessing hundreds of LLMs, outperforming simple algebraic norms (Frobenius and spectral norms).

The weighted α and log α norm are compound metrics computed from a truncated PL-fit of the eigenvalues and the size scale. The purpose lies on investigating the inter-layer dynamic flow using above two metrics and uncover potential variability known as scale collapse (Martin et al., 2021), which is assumed to reveal dysfunctions in the learning process, mostly due to the used architectures of the type of training data used.

CentralBankRoBERTa

This model offers two distinct classification tasks often used in the area of central bank communication. A binary sentiment classifier for positive/negative tonality and an economic agent classifier featuring four labels: households, firms, the financial sector and the government.

Figure 7.2 displays the model's per-layer metrics which are relatively stable. The plain α metric is mostly comprised between 3 and 5, denoting weak to moderate heavy-tailed eigenvalue distributions, meaning layers are here prone to underfitting issues. Few peaks exceeding 6 denote the randomness of some layers, while the last layer is associated with the highest value of α, mostly due to a normalization scheme.

The weighted α and the log α norm are highly correlated and both exceeding the α metric. This confirm the importance of the scale size (λ_{max}) in this model, with a significant magnitude that shifts the PL-metrics above the α.

Due to the relative homogeneity of the corpus, incorporating just central bank speeches collected by the Bank of International Settlements (BIS), the metrics are not volatile and do not exhibit a scale collapse. The size scale, measured by the maximum eigenvalue, does not alter the distance between the metrics, except for the last layer. The output of the model, a standard multiclassification task, is probably not as deep as for other tasks.

The inter-layer metric stability reinforces the hypothesis of a good training performance of the model, due to the absence of a scale collapse and high variability.

ECBERT-base

This model was developed on the basis of the existing $ModernBERT$ (Warner et al., 2024), to estimate policy preferences (hawkish/dovish) in selected corpora. This classic preference exercise in monetary policy is not easy to handle due to the ambiguous behavior or central bankers.

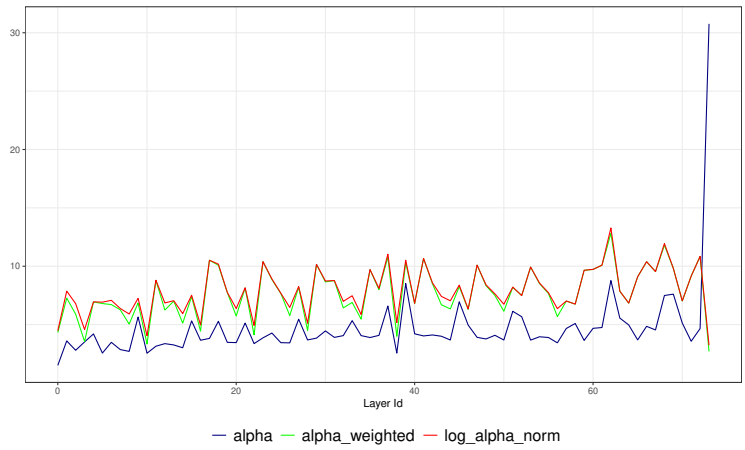


Figure 7.2: PL metrics estimated from CentralBankRoBERTa model

It results a high variability in the PL-metrics estimated and displayes in Figure 7.4. The α series is mostly comprised under the threshold of 5, however last layers exhibit a random behavior (α¿6) associated with a reduced amplitude of the size scale (λ_{max}). The latter played a key role in separating the weighted α and the log α norm from the α metric at earlier stage of the model. This could be interpreted by the relative complexity of the classification task, requiring a substantial expertise in separting preferenes, but also the transfer learning scheme stemming from the pre-retrained version of ModernBERT-base.

The scale size has a significant magnitude related to the heterogeneity of the inputs, general-purpose texts used to train the model, at the contrast of the relatively homogeneous central banking speeches used in the CentralBankRoBERTa model.

High variability of the PL-metrics could indicate a poor training quality, known as scale collapse.

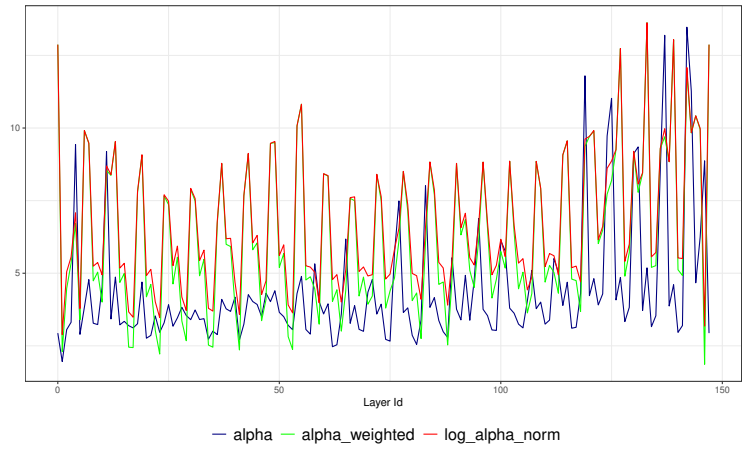


Figure 7.3: PL metrics estimated from ECBERT-base.

7.1.5 Conclusion

Machine learning witnesses a constant revolution in terms of algorithms and strategies adopted by practitioners, seeking to enhance models and their accuracy. The exercise of testing these models shifted from the traditional train/test metrics to explore elements from random matrix theory, with the purpose of studying asymptotic behavior of the models' complexity. This new strategy is particularly

robust for neural networks, whose architecture features many layers as single components of the information flow traveling between the input and output layers. This proved useful to test LLMs without the need to access train and test data, particularly for specific applications where tasks are difficult to handle. The field of central bank communication is a notorious example of linguistic complexity and ambiguous behavior of speakers, adding difficulties to classification tasks. The use of advanced per-layer metrics permitted to check potentially good trained LLMs for a simple agent classification task, while advanced policy preference estimation remains difficult to handle and may suffer from transfer learning schemes that alter the overall accuracy. The homogeneity of data proved to be a crucial element when training LLMs, because it ensures a relative stability of the layers' behavior and maximizes the information transmission between layers.

7.2 Investigating uncertainty using LLMs

7.2.1 Introduction

Innovations in the field of natural language processing (NLP) played a key role in maximizing information extraction from text documents and enhancing the performance of several tasks. The advent of LLMs accelerated dramatically the information extraction and enhanced the analysis of text data, outperforming most of NLP models used so far. Such models are built upon memory-based architectures called transformers (Vaswani et al., 2017), able to capture long-range correlations in training data.

Although hailed by users for their high accuracy and the speed of execution, LLMs remain black-box models whose inner architecture was trained on a huge variety of documents, making the output prone to judgment errors or interpretation. Their use is not free from errors and is subjected many contraints regarding the hyperparameter tuning and the data quality.

In empirical research, the incorporation of LLMs was hailed by many economists but remains a big challenge. Practioners interact with LLMs using prompts and do ignore the documents used during the training step, while the output may not be unique and its variability is linked to hallucinations.

Monetary policy, as a subfield of macroeconomics, witnessed an unprecedented interest of researchers trying to use the available documents to support their hypotheses and augment their inferences. Earlier approaches relied on sentiment analysis and text mining techniques whose interpretation is weak, in comparison to today's models.

Uncertainty plays a key role in the monetary policy practice and has been seen as a forward guidance tool at the hands of central bankers (Greenspan, 2004). First attempts to gauge uncertainty in text documents used basic word count methods, as for the economic policy uncertainty (Baker et al., 2016). Diaf and Schütze (2024) employed a semantic search approach (Angelov, 2020) to better apprehend contextual information and distinguish uncertainty from risk.

While semantic search models require a sizeable amount of documents to achieve robust results, they offer better results than dictionaries and other word counts. The linkages in the triplet document-topic-word are more informative than standard topic models (Blei et al., 2003), known to be agnostic unsupervised machine learning tasks.

Recent attempts to use LLMs in monetary policy targeted stock market dynamics (Kwon et al., 2024), emotions (Baerg and Binder, 2024) and inflation expectations (Zarifhonarvar, 2026), to name a few. Results suggest a cautious use of LLMs for tasks known to be complex and requiring in-depth analysis. Uncertainty is one of these tasks and needs a special attention from practitioners.

This work puts into practice LLMs to estimate uncertainty levels and tonality of the Federal Reserve minutes during the last 25 years. The same hyperparameters and prompting was used on meta-llama-3.1-8b-instruct and gpt-oss-120b models to compare the outputs and validate them with the monetary policy uncertainty index (MPU) (Husted et al., 2020) as well as the monetary policy surprise dataset (Bauer and Swanson, 2023).

Results show noticeable departures in the uncertainty levels estimated with the two selected LLMs, while tonality measurements remain close. The LLMs output could be used as leading indicators to predict many variables and indices, however, questions surrounding the prompting, hyperparameters and the inner architecture make the use of LLMs difficult for reproducibility tasks.

7.2.2 Uncertainty in text analysis

Uncertainty is seen as an hidden variable that cannot be captured via simple techniques. It does need detailed procedures in order to come up with sound estimates that meet reality.

Earlier attempts to gauge uncertainty used simple word counts (Loughran and McDonald, 2011) for a general-purpose estimation of how a document is uncertain. This strategy was used in economics and monetary policy to issue uncertainty indices, following the occurrence of specific terms in selected corpora (Baker et al., 2016; Husted et al., 2017).

While researchers still use these indices for investigative purposes, they lack semantic depth because uncertainty concerns, or alternatively confidence, could be mentioned by terms or expressions not encompassed by the word lists used by researchers.

Another concern is this line of research is the type of uncertainty researchers try to track. Diaf and Schütze (2024) used semantic search models to leverage contextual information to better assess the uncertainty in international central banking documents. This helps separating uncertainty from risk, as mentioned in Chapter 5.

On uncertainty drivers, earlier applications were not convincing regarding the formation of uncertainty and its interpretation. As models with reasoning capabilities, LLMs have been used in a wide range of economic applications as for stock market dynamics (Kwon et al., 2024), emotions (Baerg and Binder, 2024) and inflation expectations (Zarifhonarvar, 2026), to name a few. Results suggest a cautious use of LLMs for complex tasks requiring an in-depth analysis (Baerg and Binder, 2024), and warning the error-prone, customized prompting. These shortcomings stem from the often ignored behavioral patterns and cognitive processes (Echterhoff et al., 2024).

7.2.3 Uncertainty measurements

237 minutes released by the Federal Reserve covering the period 2000-2025 were downloaded [3] and used as input to uncover a scale-based indices for uncertainty ([1,10]) and a tonality ([-10,10]).

Prompting used in the uncertainty estimation:

$system\ prompt$ = 'You are a macroeconomic expert in estimating uncertainty and tonality in a given text.
Evaluate the uncertainty using a scale of 1 to +10 (10 being the most uncertain and 1 being the less uncertain).
Evaluate the tonality using a scale of -10 to +10 (10 being the most positive and -10 being the most negative). '

$meta - llama - 3.1 - 8b - instruct$ and $gpt - oss - 120b$ models were used to get results, on the basis of an identic zero-shot prompting with the same hyperparameters [4]. The experiment did not use any example for prompting (few-shot learning), in order to avoid a partial guidance that may impede results.

Over the last 25 years, minutes dealt with a variety of topics as for banking stability and financial innovation, whose importance was changing over time. Few-shot strategies might provide good anchors for uncertainty estimates, but would also distort comparison between documents deemed

[3] github.com/vtasca/fed-statement-scraping/blob/master/communications.csv
[4] maxTokens = 500, temperature = 0.3, topP = 0.9, topK = 20

	Surprise	Orthogonalized.Surprise	uncertainty_llama	uncertainty_openai	tonality_llama	tonality_openai
Surprise	1.00	-	-	-	-	-
Orthogonalized.Surprise	0.91	1.00	-	-	-	-
uncertainty_llama	0.05	0.01	1.00	-	-	-
uncertainty_openai	0.26	0.21	0.34	1.00	-	-
tonality_llama	0.16	0.08	-0.58	-0.33	1.00	-
tonality_openai	-0.18	-0.20	-0.32	-0.41	0.56	1.00

Table 7.1: Correlation matrix of uncertainty and tonality outputs with monetary policy surprises estimates.

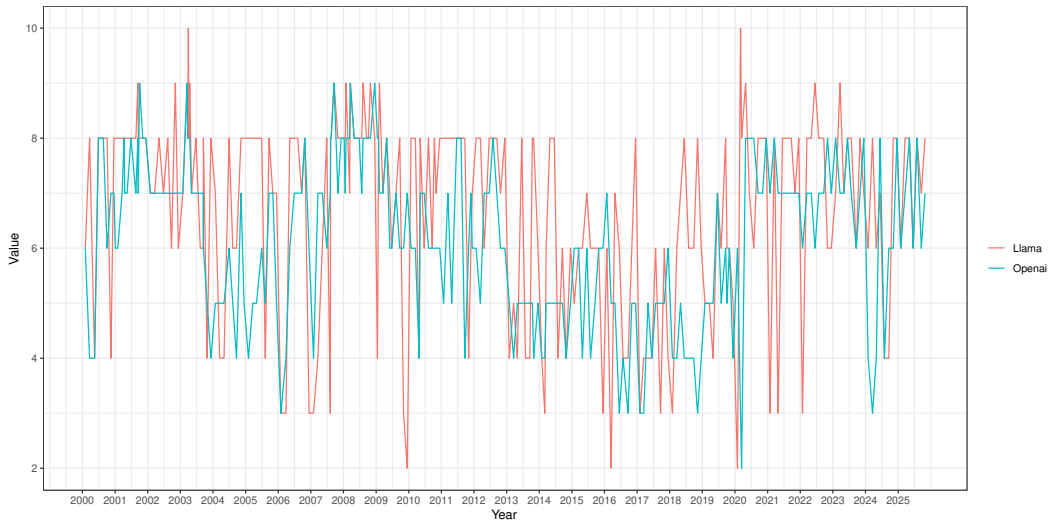


Figure 7.4: Uncertainty estimates from the Federal Reserve minutes (2000-2025)

to be close in terms of debated topics, but different in uncertainty scales. Uncertainty acts as a latent variable, whose estimation cannot be done straightforward but carries a perplexity, making its assessment different across experts.

Figure 7.4 displays the uncertainty covariates from the used LLMs with a noticeable divergence by periods. The link remains (0.34) as given in Table 7.1. A certain convergence, however, is seen in the tonality covariates (correlation 0.56) especially during the last three years of the sample (2022-2025) as given in Figure 7.5.

To investigate the predictive power of LLMs, I compared the resulting uncertainty and tonality covariates with monetary policy surprises dataset (Bauer and Swanson, 2023). Table 7.1 does not highlight any noticable correlation with Surprise and Orthogonal Surprise variables and the LLMs covariates are not informative to predict the surprises, in terms of causality.

7.2.4 Uncovering uncertainty factors

For the sake of illustration, a Federal Reserve minute released during the pandemic was used as an input to a prompt seeking uncertainty drivers and their interactions. The aim was to reveal a hidden narrative map on critical components given in the minute.

Figure 7.7 shows the graph-like LLMs output regarding uncertainty factors and makes sense of the rhetoric used by the Fed. The pandemic plays a key role in the narrative, as its node (pandemic health shock) is central and negatively impact its neighboring factors (labor market deterioration, fiscal stimulus and public health restrictions).

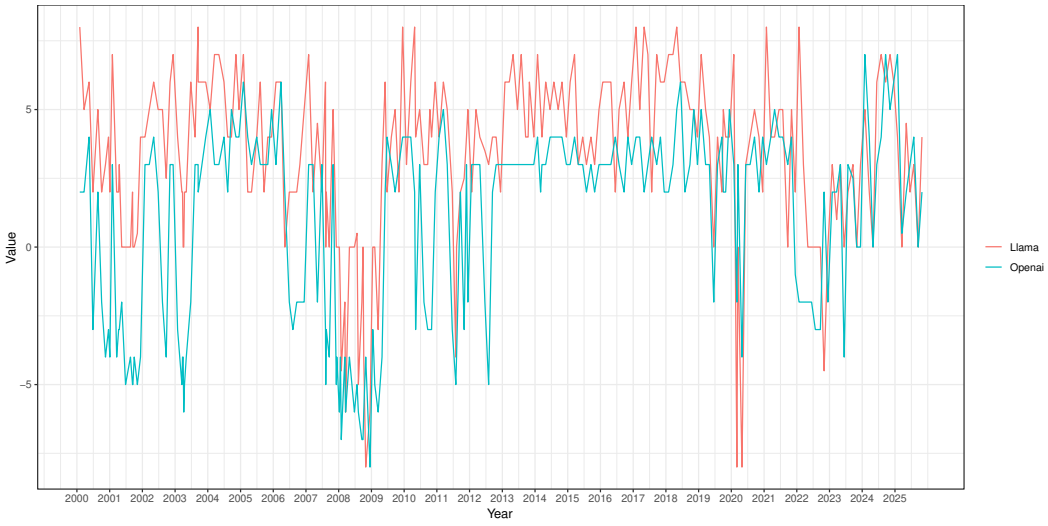


Figure 7.5: Tonality estimates from the Federal Reserve minutes (2000-2025)

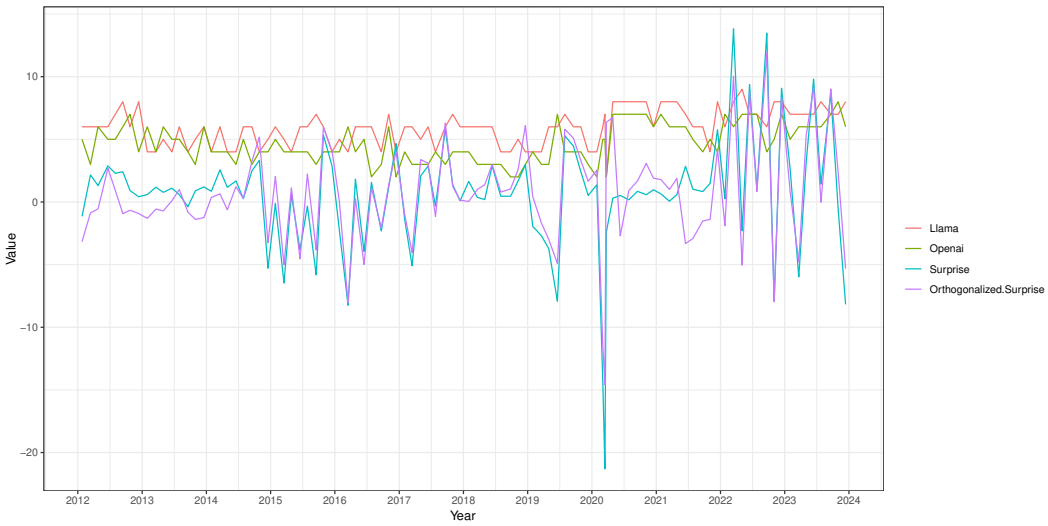


Figure 7.6: Uncertainty estimates from the Federal Reserve minutes (2000-2025) and monetary policy surprises.

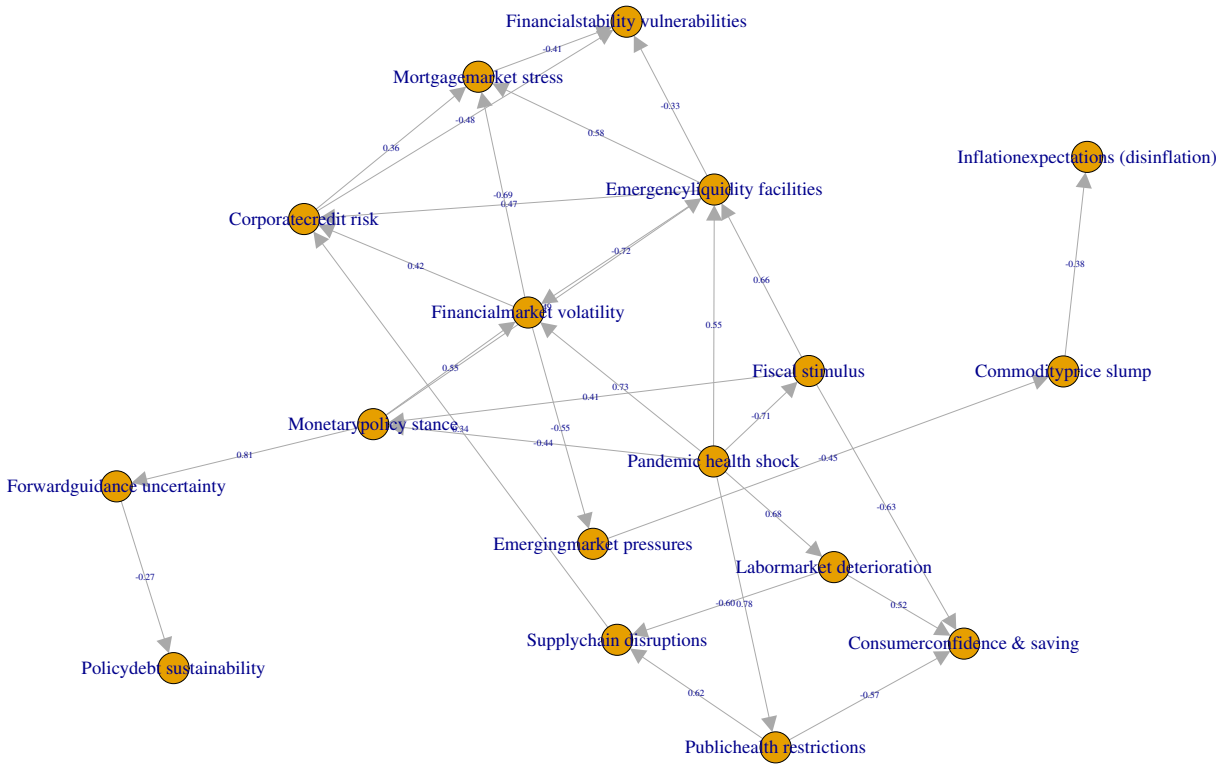


Figure 7.7: Uncertainty drivers (using OpenAI) from the Federal Reserve minute released on April 29^{th}, 2020

Policy-based uncertainty factors are found on the left side of the graph, consisting of policy-debt sustainability, forward-guidance uncertainty and monetary policy stance. On the right side, price-related nodes describe inflation expectations, commodity prices and consumer confidence.

The central part of the graph display intensive correlations between factors related to financial markets at top. Financial volatility has a pivotal role, along emergency liquidity facility, given the high number of vertices and their magnitudes.

Result displayed in Figure 7.7 may slightly change, in terms of nodes and magnitudes, depending on the LLM output. Tests on OpenAI ($gpt - oss - 120b$) showed promising results, in comparison to model$meta - llama - 3.1 - 8b - instruct$. The latter's results suffer from inconsistencies and biases, in comparison to OpenAI[5]. Setting the temperature parameter to 1 helps getting granular results without falling into the hallucination trap, while the dynamic rhetoric used in the addresses needs a rich vocabulary (topP und topK). Lower temperature values close to 0 yield poor results in terms of interpretability. Although reproducibility remains of a great concern, LLMs proved to be a reliable tool to uncover narratives, exceeding classic topic models in terms of quality and inference.

Prompting used to get latent uncertainty drivers

$system\ prompt$ = 'You are a macroeconomic expert in estimating uncertainty in a given text. Evaluate the uncertainty drivers using a network, where each driver is a node and each vertex is the correlation between nodes. Final output are two dataframes: one for edges and one for vertices. '

[5]both model were run using the following hyperparameters: maxTokens = 3000, temperature = 1, topP = 0.8, topK = 40

Chapter 8

Policy mix in an oil exporting country: effectiveness of countercyclical measures in mitigating external shocks[1]

Abstract

Gauging the impact of oil price variations on small, oil-exporting countries has been heavily investigated under the umbrella of monetary policy interventions, using a standard general equilibrium framework. For some countries, the monetary policy coordinates with fiscal policy to deliver a better response to external oil shocks, in an attempt to make the economic activity resilient to a potential backlash. This paper investigates the policy mix effectiveness in a small open economy, namely Algeria, and its ability to mitigate a negative oil price shock, using a dual monetary/fiscal DSGE framework. The model maps several frictions found in single-commodity economies as for a managed exchange rate regime, the existence of a foreign exchange market accessible to households and a sovereign wealth fund. Simulations show countercyclical fiscal measures (increase in government spending) coupled with monetary interventions have no expansionary effects on output, but still necessary to maintain a resilient economic activity especially for the non-oil sector. Under the sticky prices assumption, households tend to lower their consumption level and use their foreign currency savings as buffer against currency depreciation. This results in alleviating potential pressures on the supply side and preventing possible inflation spikes. Findings confirm the effectiveness of a monetary policy based on targeting export products, which better handles the negative terms of trade shock via a slight exchange rate depreciation. However, the fiscal dominance in the policy-mix leads to the accumulation of public debt, which might require fiscal consolidation during protracted periods of declining oil prices.

8.1 Introduction

DSGE models are usually linked to developed economies, where their advanced micro-level structure is featured and modeled as a block of equations (Smets and Wouters, 2003). However, applying these models to developing countries could be biased and must take into account several stylized facts, or frictions, that alter the models' general framework.

Emerging countries, at the contrary, show different characteristics. For instance, public investment remains an important pillar of the fiscal policy, which must be distinguished from public consumption (Berg et al., 2013). Monetary policy is often conducted as a hybrid mixture of inflation targeting and

[1] This chapter features an augmented text-based application built on a previously published paper at the Dynare Working Papers (CEPREMAP) https://www.dynare.org/wp-repo/dynarewp083.pdf

managed exchange rate regime, thus requiring two distinctive monetary policy instruments: interest rate and foreign exchange interventions (Benes et al., 2015). Households are often heterogeneous in their income distribution and access to financial markets, requiring to define a portion of the population being constrained to just use the earned wages (Mankiw, 2000).

Oil-exporting economies, on the other hand, exhibit different characteristics than emerging ones (Algozhina, 2016), calling for a separate specification of the production sector (oil and non-oil) and defining an accumulation fund that collects oil taxes (example of a sovereign wealth fund). In addition to foreign exchange interventions usually adopted by the central bank, Frankel (2011) suggested monetary authorities to adopt product price targeting (PPT), as an alternative to consumer price index (CPI), to make the monetary policy automatically countercyclical during periods of volatile terms of trade. Algozhina (2016) tested the effectiveness of these two anchors (PPT and CPI) jointly with fiscal instruments, on the basis of a defined welfare measure equaling the sum of three variances (input, inflation and real exchange rate) (De Paoli, 2009) and concluded that a procyclical fiscal policy coupled with CPI-based monetary targeting deliver the best response for the case of Kazakhstan.

The aforementioned frictions pertain to the case of Algeria, a small open, oil-exporting economy with a heavy reliance on hydrocarbon exports and a less developed financial market. The government plays a pivotal role in the economic landscape, especially during times of lower oil prices, where a prompt response is needed to prevent widespread negative impact on the real sector and a deterioration of its external position. Financial constraints to foreign currencies, because the local currency is not convertible, gave birth to a parallel market, allowing households to exchange, with a high premium, the local currency with main foreign currencies (IMF, 2018). Inflation rate has not reached, during last decade, high levels and was not impacted by an exchange rate pass-through; where price tensions have a domestic origin (fresh food products) (IMF, 2024).

The literature does not demonstrate any attempt to gauge an external oil shock, from a policy mix perspective. The interest of researchers was primarily given to assessing monetary policy practice (Boucekkine et al., 2021), although fiscal instruments have been heavily used to implement economic programs as well as financial planning (Chibi et al., 2019); while public spending remained uncorrelated with fiscal revenues.

Moreover, oil-exporting countries have different objectives regarding the macroeconomic policy and put stress on various aggregates in order to achieve stabilization. The loss function proposed by De Paoli (2009) considers equal weights for three variances, however, some countries put efforts on the currency stabilization to offset potential spillover. In this sense, text mining could be used to estimate an adequate loss function, whose weights translate the importance of their related aggregates from a textual perspective.
This work aims to address this issue by considering both monetary and fiscal policies as tools to face a mild decline in international oil prices, which accounts for a negative shock affecting the country's terms of trade (Frankel, 2011). It extends the model built by Algozhina (2016) and tailors its framework to match the specification of the Algerian economy, with the purpose to evaluate the outcomes of such actions on key macroeconomic aggregates.

We propose to put forward the interaction of a PPT-anchored, managed exchange rate regime with foreign exchange interventions (monetary instruments), in conjunction with a countercyclical fiscal policy built upon increasing government spending. To this end, we consider a DSGE model with sticky prices, rule-of-thumb consumers (non-ricardian) and optimizer consumers having access to foreign exchange markets (ricardians). It also features a non-oil sector, a single oil-producing

company jointly owned by the state and foreign investors; and the rest of the world as one entity.

Results confirm the effectiveness of the aforementioned policy-mix in making the economic activity resilient to negative oil shocks and supporting the real sector activity; but highlights a fiscal dominance with looming risks over the public debt. Notably, no expansionary effects on the output are noticed, however households' consumption and investment drop due to non-oil labor market fluctuations, despite subdued inflation and a mild exchange rate depreciation. Ricardian households use their foreign exchange savings to face uncertainties at the beginning of the shock, before building up their assets again. The efficiency of such policy may be constrained by prolonged lower oil prices and the viability of the external position. In parallel, public debt carries the burden of the countercyclical policies and may require fiscal consolidation over the medium run.

The paper outlines the related literature (Section 2). The theoretical, per-agent build-up of the model is expanded upon in Section 3, before conducting estimation and calibration (Section 4) and finally discuss the simulation results (Section 5).

8.2 Related literature

An abundant literature deals with the study of exogeneous adverse shocks and their impact through DSGE models, although the scope and the aims are different, depending on the economic structure and the modeling assumptions.

Algozhina (2016) simulated three fiscal policies and two monetary regimes for the case of Kazakhstan and founded a certain effectiveness of a procyclical fiscal stance coupled with inflation targeting without foreign exchange interventions. Results found the PPT rule as the main cause of high variations in output and exchange rate. Bańkowski et al. (2021) built a pre-pandemic and post-pandemic assessments of various policy-mix instruments for the Euro area and concluded that fiscal and monetary policies reinforce each other.

For the case of Algeria, we noticed several contributions, using DSGE frameworks, aimed at assessing a negative oil shock and its potential impact on the domestic economy, with a particular interest given to monetary policy actions and its instruments.

Dib (2010) cited three main transmission mechanisms during the 2008 global crisis that affected the Algerian economy. Increasing costs of firms' borrowing to finance their investments, decline in oil prices, and the lower world interest rates were all found to pose significant risks to Algeria, requiring expansionary monetary and government spending policies to mitigate such impact (Dib, 2010).

Allegret and Benkhodja (2012) studied the monetary policy as a macroeconomic stabilization policy after an oil shock and investigated the dynamic effects of four different shocks (oil price, real exchange rate, international interest rate and foreign inflation) along the appropriate monetary policy rule. A multi-sector DSGE model with nominal and real rigidities was built and showed core inflation targeting was the best monetary rule to stabilize both output and inflation.

Chibi et al. (2019) found, through a VAR model, an evidence of non-ricardian fiscal policy, that dominates monetary policy. It confirms a negative correlation between fiscal balances and government liability, that was prevailing during last two decades.

Boucekkine et al. (2021) used cointegration estimation to study the long-term stability of money supply (1979-2019) and found monetary aggregates M1 and M2 to have a stable demand, at the

contrary of fiat currency which was unstable.

8.3 Description of the model

The proposed model (Algozhina, 2016) comprises several frictions as for incomplete asset market, adjustment costs, collateral constraint and Calvo price setting. In the presence of a foreign entity (rest of the world), the domestic economy is not considered as a borrower, thus, the domestic and foreign discount factors are equal.

Two producers are considered in the model: an oil firm, owned by the government and foreigners, and non-oil firms owned by nationals. Non-oil firms are monopolistically competitive and have their prices fixed à la Calvo (Calvo, 1983), and their profits given to households. Governement's oil share and revenues are transferred to the SWF, whose returns are assumed to be transferred to the government budget. Foreign Direct Investments (FDI) mainly affect the oil production and responds to international oil prices.

Households are two types: ricardian and non-ricardian. Only ricardians can borrow from abroad and have restictions on their non-oil collateral. They also hold local government bond, own non-oil firms, rent capital to these firms and decide about investments. Non-ricardian households use all their revenues during each period. Labor market is assumed competitive without workers' unions and wage bargaining.

The CPI-Taylor rule includes lagged interest rate, the CPI inflation and the GDP (Sarno and Taylor, 2001). The PPT-Taylor rule incorporates oil price inflation and domestic inflation, weighted by the oil sector and non-oil GDP respectively (Sarno and Taylor, 2001). Rules for public consumption and public investment include fiscal debt, oil revenues and oil sector output to capture the fiscal policy position (countercyclical). Public investment is productive in the accumulation of public capital and also considered as an additional input in the Cobb-Douglas function for non-oil production, besides labor and physical capital.

8.3.1 Households

The domestic economy has a continuum of households on an interval [0,1] where a fraction μ represents non-ricardians unable to access financial markets and consume all their available revenues during each period (non-optimizers). Ricardian households $(1-\mu)$ are prospective optimizers that hold government bonds, borrow from abroad, invest in non-oil capital, rent capital to non-oil firms and receive profits from them. Labor market is assumed to be competitive and revenues are equal for all households (both household types have the same working hours).

The index S indicates a variable associated to ricardian households and N refers to non-ricardian ones. Ricardians optimize the utility function (Schmitt-Grohe and Uribe, 2002):

$$E_0 \sum_{t=0}^{\infty} \beta^t \frac{[C_t^S - \phi^{-1} N_t]^{1-\sigma} - 1}{1-\sigma}, \phi > 1, \sigma > 1 \tag{8.1}$$

under the budget constraint:

$$C_t^S + I_t + b_t + R_{t-1}^* \frac{RER_t}{RER_{t-1}} \frac{b_{t-1}^*}{\pi_t^{*r}} + T_t^S = W_t N_t + R_t^{kno} K_{t-1}^{kno} + R_{t-1} \frac{b_{t-1}}{\pi_t} + b_t^* + \Pi_t \tag{8.2}$$

where: $b_t = \frac{B_t}{P_t}$ is the real purchases of government bonds, RER_t is the real exchange rate (price of a basket of foreign goods in terms of a basket of domestic goods), $b_t^* = RER_t \frac{B_t^*}{P_t^*}$ is the real borrowing from abroad expressed in terms of domestic goods, R_{t-1} and R_{t-1}^* are respectively the domestic and foreign interest rates in nominal terms, T_t^S are real taxes collected, W_t real wages, R_t^{kno} is the cost of non-oil physical capital rental, $\pi_t = \frac{P_t}{P_{t-1}}$ is the inflation and Π_t is the real profit of monopolistic non-oil firms.

Non-oil capital movement incorporates adjustment costs fo inflation (Berg et al., 2013)

$$K_t^{no} = (1-\delta)K_t^{no} + \left[1 - \frac{\kappa}{2}\left(\frac{I_t}{I_{t-1}} - 1\right)^2\right] I_t, \kappa > 0 \tag{8.3}$$

The collateral constraint related to foreign liabilities to a future value of capital (Faia and Iliopulos, 2011) is expressed as:

$$R_t^* b_t^* = E_t \left\{ \Omega \frac{Q_{t+1}\pi_{t+1}^*}{RER_{t+1}/RER_t} K_t^{no} \right\} \tag{8.4}$$

where Q_t is a real shadow value of capital (Tobin's Q) and Ω is the upper bound of the leverage ratio. While this specification fits the case of a heavily-indebted country (Algozhina, 2016), the application will assign a lower upper bound value, given the relative funding capacity of Algeria.

Hence, households' optimization is stated as a maximization of the utility (1) with respect to consumption C_t^S, investment I_t^S, capital K_t^{no}, government bonds holdings b_t, foreign borrowing b_t^* and hours worked N_t, subjected to the budget constraint (2), capital accumulation equation (3) and collateral constraint (4).

The non-optimizing household has the same preferences as the optimizer but chooses only consumption and labor under the budget constraint:

$$C_t^N + T_t^N = W_t N_t \tag{8.5}$$

Each type of household ($i \in S, N$) has a composite CES consumption preference over domestic and foreign goods with $\eta > 0$ as an elasticity of substitution between goods:

$$C_t(i) = \left[\gamma^{\frac{1}{\eta}} C_{H,t}^{\frac{\eta-1}{\eta}}(i) + (1-\gamma)^{\frac{1}{\eta}} C_{F,t}^{\frac{\eta-1}{\eta}}(i)\right]^{\frac{\eta}{\eta-1}}$$

where γ is a home-bias parameter and (1-γ) indicated degree of openness. Minimizing a household's consumption expenditures gives the following CPI index:

$$P_t^{1-\eta} = \gamma P_{h,t}^{1-\eta} + (1-\gamma) P_{f,t}^{1-\eta} \text{ ou bien } 1 = \gamma p_{h,t}^{1-\eta} + (1-\gamma) RER_t^{1-\eta} \tag{8.6}$$

where $p_{h,t}$ is the price of domestic goods to composite consumption and RER_t is the relative price of foreign goods to composite consumption.
$C_t = \mu C_t^N + (1-\mu) C_t^S$ is the equation of the aggregate consumption. As for private consumption, investment is the CES basket with the same home-bias parameter γ and CPI price index for simplicity.

8.3.2 Firms

As for Galí et al. (2007), we assume the existence of monopolistically competitive non-oil firms producing differentiated intermediate goods, and a perfectly competitive non-oil firm producing a final domestic good, whose producer has a constant returns technology:

$$Y_t^{no} = \left(\int_0^1 X_t(j)^{\frac{\epsilon-1}{\epsilon}} dj \right)^{\frac{\epsilon}{\epsilon-1}}$$

where $X_t(j)$ is the input amount of intermediate good j and $\epsilon > 1$ is the elasticity of substitution between differentiated intermediate goods. It maximizes profit taking as given the domestic final goods price P_t^h and intermediate goods prices $P_t^h(j)$, so the optimal demand allocation could be written as:

$$X_t(j) = \left(\frac{P_t^h(j)}{P_t^h} \right)^{-\epsilon} Y_t^{no} \tag{8.7}$$

Each company producing intermediary goods has the identical Cobb-Douglas production function that incorporates private non-oil capital, labor and public capital:

$$Y_t^{no}(j) = u^{no} K_{t-1}^{no}(j)^{\alpha} N_t(j)^{1-\alpha} K_{G,t-1}^{\psi} \tag{8.8}$$

where the level of technology u^{no} remains constant and the usage of public capital are common to all firms.

Intermediate good producers solve their problem in two stages. First minimizing costs according to the production function (8) gives the real marginal costs common to all non-oil firms (real wage and rental cost of capital are assumed given) as:

$$MC_t = \frac{W_t^{1-\alpha} (R_t^{kno})^{\alpha}}{u^{no} K_{G,t-1}^{\psi} (1-\alpha)^{1-\alpha} \alpha^{\alpha}} \tag{8.9}$$

Second, intermediate non-oil producers choose the price P_t^{hop} to maximize their discounted real profits:

$$\sum_{m=0}^{\infty} \theta^m E_t \left\{ D_{t,t+m} Y_{t+m}^{no}(j) \left(\frac{P_t^{hop}}{P_{t+m}^h} - MC_{t+m} \right) \right\} \tag{8.10}$$

where $D_{t,t+m} = \beta^m E_t \left(\frac{U_{C_{t+m}^S}}{U_{C_t^S}} \right)$ is a stochastic discount factor stemming from the optimizing household's problem, subject to the demand constraint according to (7):

$$Y_{t+m}^{no}(j) = \left(\frac{P_t^{hop}}{P_{t+m}^h} \right)^{-\epsilon} Y_{t+m}^{no}$$

One can explain the price stickiness (Calvo, 1983) as a fraction $(1-\theta)$ of non-oil firms makes price adjustment at each period, at the contrary of a fraction θ that keeps their prices unchanged. The evolution of the domestic price index is given by:

$$(P_t^h)^{1-\epsilon} = \theta (P_{t-1}^h)^{1-\epsilon} + (1-\theta)(P_t^{hop})^{1-\epsilon}$$

The first order condition of this price setting decision (10) is:

$$\sum_{m=0}^{\infty} \theta^m E_t \left\{ D_{t,t+m} Y_{t+m}^{no}(j) \left(\frac{P_t^{hop}}{P_{t+m}^h} - \frac{\epsilon}{\epsilon-1} MC_{t+m} \right) \right\} = 0 \tag{8.11}$$

where $\frac{\epsilon}{\epsilon-1}$ is a frictionless price markup.

The production function of the oil firm is only dependent of capital input, so to match the hypothesis of a capital-intensive sector and canceling any complication coming from labor mobility between the oil and non-oil sector:

$$Y_t^o = (K_{t-1}^o)^{\alpha_0} \tag{8.12}$$

Oil capital accumulated by FDI which responds to world oil prices [2]:

$$K_t^o = (1-\delta)K_t^o + FDI_t^* \tag{8.13}$$

$$\widehat{FDI_t^*} = \rho_{FDI}\widehat{FDI_{t-1}^*} + (1-\rho_{FDI})\widehat{P_t^{o*}} \tag{8.14}$$

World oil price are assumed to follow an AR(1) process and has an exogenous shock referred as terms of trade shock:

$$\widehat{P_t^{o*}} = \rho_o\widehat{P_{t-1}^{o*}} + \epsilon_t^o \tag{8.15}$$

The oil firms receives its profits Π_t^{o*} net of royalties levied on production quantity at a rate τ_o:

$$\Pi_t^{o*} = (1-\rho^o)P_t^{o*}Y_t^o \tag{8.16}$$

The oil sector is owned by the foreigners and the government, the latter receives ι^{div} which denotes dividend share of oil profits

8.3.3 Fiscal policy

Government collects lump-sum taxes T_t and oil revenues OR_t to transfer them to the SWF. It issues treasury bonds to fund government purchases which include public consumption G_t^C and public investment G_t^I. The government's budget constraint is written as:

$$(1-\mu)b_t + T_t + \underbrace{(R_{t-1}^*\frac{1}{\pi_t^*} - \rho_{swf})SWF_{t-1}^*RER_t}_{OR_t} = p_t^g(G_t^C + G_t^I) + (1-\mu)R_{t-1}\frac{b_{t-1}}{\pi_t} \tag{8.17}$$

where $T_t = (1-\mu)T_t^S + \mu T_t^N$ and p_t^g is the price of government purchases related to the composite consumption with its own bias parameter γ_2:

$$p_t^g = \left[\gamma_2 p_{h,t}^{1-\eta} + (1-\gamma_2)RER_t^{1-\eta}\right]^{\frac{1}{1-\eta}} \tag{8.18}$$

Public investment is productive, so that the law of public capital motion is given by:

$$K_t^G = (1-\delta_g)K_{t-1}^G + G_t^I \tag{8.19}$$

Oil revenues, collected in foreign currency, consist of dividends and the share of government in the oil sector's profit.

$$T_t^{o*} = \tau^o P_t^{o*}Y_t^o + \iota^{div}\Pi_t^{o*} \tag{8.20}$$

which build up the accumulation of the SWF, following:

$$SWF_t^* = \rho_{swf}SWF_{t-1}^* + T_t^{o*} \tag{8.21}$$

[2] Hereafter, hat sign $\widehat{\ }$ denotes the deviation of the variables from their steady states.

Two fiscal instruments, public investment and public consumption, have the following laws with their responses regarding oil production (ϑ_{GI} and ϑ_{GC}, associated to fiscal cycle:

$$\widehat{G_t^I} = \rho_{GI}\widehat{G_{t-1}^I} + (1-\rho_{GI})[\vartheta_{GI}\widehat{Y_t^o} - \gamma_{GI}\widehat{b}_{t-1} + \gamma_{OR}^{GI}\widehat{OR_t}] \tag{8.22}$$

$$\widehat{G_t^C} = \rho_{GI}\widehat{G_{t-1}^C} + (1-\rho_{GC})[\vartheta_{GC}\widehat{Y_t^o} - \gamma_{GC}\widehat{b}_{t-1} + \gamma_{OR}^{GC}\widehat{OR_t}] \tag{8.23}$$

Because fiscal debt is part of the government's budget constraint, taxes require a separate equation that includes fiscal debt, public spending (similar to Galí et al. (2007)) and taxes from oil revenues.

$$\widehat{T}_t = \phi_b\widehat{b}_{t-1} + \phi_I\widehat{G_t^I} + \phi_C\widehat{G_t^C} - \phi_{OR}\widehat{OR_t} \tag{8.24}$$

8.3.4 Monetary policy

Nominal interest rate is a function of its lagged value, CPI inflation and the aggregate production, following a Taylor rule for inflation targeting:

$$\widehat{R}_t = \rho\widehat{R}_{t-1} + (1-\rho)[\phi_\pi\pi_t + \phi_y\widehat{Y}_t] \tag{8.25}$$

For the PPT Taylor rule, PPT inflation is a weighted mean of oil prices' inflation (in real terms) $\pi_t^o = \Delta\widehat{P}_t^{o*} + \Delta\widehat{RER}_t$ and domestic inflation $\pi_t^h = \pi_t - \frac{1-\gamma}{\gamma}\Delta\widehat{RER}_t$, where the weights are simply the GDP shares of oil sector s_o and non-oil sector $(1-s_o)$ respectively.

$$\widehat{R}_t = \rho\widehat{R}_{t-1} + (1-\rho)\left[\phi_\pi[s_o\pi_t^o + (1-s_o)\pi_t^h] + \phi_y\widehat{Y}_t\right] \tag{8.26}$$

Interventions in the foreign exchange market are buying/selling foreign currencies by the central bank, which accumulates reserves (Benes et al., 2015), following the exchange rate and its depreciation rate.

$$\widehat{fxr}_t^* = \rho_{fxr}\widehat{fxr}_{t-1}^* + (1-\rho_{fxr})\left[\alpha_1\widehat{RER}_t + \alpha_2\Delta\widehat{RER}_t\right], \alpha_1 < 0, \alpha_2 < 0 \tag{8.27}$$

where $fxr_t^* = RER_t\dfrac{FXR_t^*}{P_t^*}$ are the real exchange reserve expressed in domestic products.
Three clearing conditions were introduced in the model:
The domestic non-oil goods market:

$$p_t^h Y_t^{no} = \gamma\left[C_t + (1-\mu)I_t\right] + \gamma_2 p_t^g(G_t^C + G_t^I) \tag{8.28}$$

Real GDP from supply and demand sides:

$$p_t^h Y_t^{no} + Y_t^o P_t^{o*} RER_t = C_t + (1-\mu)I_t + p_t^g(G_t^C + G_t^I) + NX_t \tag{8.29}$$

Balance of payments condition equalizing its current and financial accounts.

$$NX_t + (R_{t-1}^*\frac{1}{\pi} - \rho_{swf})SWF_{t-1}^* RER_t - (1-iota^{div})RER_t\Pi_t^{o*} =$$

$$(1-\mu)\left(R_{t-1}^*\frac{RER_t}{RER_{t-1}}\frac{b_{t-1}^*}{\pi_t^*} - b_t^*\right) - RER_t FDI_t^*$$

8.3.5 The rest of the world

Three exogenous equations define the rest of the world, as a large economy:

$$\widehat{Y}_t^* = \rho_{Y^*}\widehat{Y}_{t-1}^* + \epsilon_t^{Y^*} \tag{8.30}$$

$$\widehat{R}_t^* = \phi_\pi^*\pi_t^* + \phi_y^*\widehat{Y}_t^* \tag{8.31}$$

$$\pi_t^* = \beta^* E_t\pi_{t+1}^* + \lambda^*\left(\sigma + \frac{\phi^* + \alpha^*}{1-\alpha^*}\right)\widehat{Y}_t^* \tag{8.32}$$

Overall, 27 variables, represented as log-deviations from the steady state, constitute a system of 27 equations: inflation π_t, aggregated households consumption $\widehat{C}_t$, hours worked $\widehat{N}_t$, domestic interest rate $\widehat{R}_t$, net exports $\widehat{NX}_t$, foreign reserve $\widehat{fxr}_t^*$, foreign interest rate $\widehat{R}_t^*$, l'inflation de l'étranger π_t^*, la production $\widehat{Y}_t^*$, external debt $\widehat{b}_t^*$, oil sector capital $\widehat{K}_t^o$, non-oil sector capital $\widehat{K}_t^{no}$, public sector capital $\widehat{KG}_t$, real exchange rate $\widehat{RER}_t$, la fiscal debt $\widehat{b}_t$, collected taxes $\widehat{T}_t$, public consumption $\widehat{GC}_t$, public investment $\widehat{GI}_t$, private investment $\widehat{I}_t$, oil production $\widehat{Y}_t^o$, non-oil production $\widehat{Y}_t^{no}$, aggregated production $\widehat{Y}_t$, domestic prices $\widehat{p}_t^h$, government purchasing prices $\widehat{p}_t^g$, sovereign wealth fund assets $\widehat{SWF}_t$, foreign direct investments $\widehat{FDI}_t$ and international oil prices $\widehat{P}_t^{o*}$

8.4 Estimation and calibration

The model was calibrated according to original methodology (Algozhina, 2016) as shown in Table 8.1, while some parameters where tuned to match the case of Algeria as for the oil royalty rate (τ^o), the persistence parameter in world oil price (ρ_o) and the ratios (nx_y, $bstar_y$, swf_y, fdi_y, s_o, gI_y, gC_y) (of Algeria, 2024).

We simulate, using Dynare (Adjemian et al., 2024), a policy-mix intervention to alleviate the effect of a persistent 1% drop in international oil prices (Pstar-o), by considering a countercyclical fiscal policy coupled with a PPT-anchored managed exchange rate regime, under sticky price hypothesis. This negative external shock could be seen as a negative terms of trade shock (Frankel, 2011), that needs to be addressed via exchange rate depreciation.

The parameters ϕ_π and ϕ_y, respectively inflation and output response in the PPT Taylor rule (Equation 26), are first learned using grid search strategy to minimize the loss measure expressed as the sum of three variances: inflation, output and real exchange rate (De Paoli, 2009). Values ranging from 0 to 2, with 0.1 increments, were used to run the grid search and yielded ϕ_π=1.7 and ϕ_y = 0.3. The output response (ϕ_y) is not far from the standard value (0.125) found by many researchers (Algozhina, 2016; Galí, 2015).

The use of text-based loss function, from the word frequencies shown in Table 8.2, yields similar results (ϕ_π=1.6 and ϕ_y = 0.4), while the overall loss is slightly higher than the standard weighting of De Paoli (2009), as shown in Table 8.3.

8.5 Results

A simulation of a 10% drop in international oil prices, coupled with a countercyclical fiscal policy at a managed exchange rate regime with sticky prices yields the impulse responses displayed in Figure

Parameter	Value	Definition
β	0.978	discount factor
γ	0.68	home-bias in consumption and investment
γ_2	0.90	home-bias in government purchases
Ω	0.54	the upper bound of leverage ratio
μ	0.50	the fraction of rule-of-thume (non-ricardian) households
α	0.30	non-oil output elasticity to private capital
ψ	0.16	non-oil output elasticity to public capital
α^o	0.70	oil output elasticity to private capital
ϕ	1.45	wage elasticity to hours worked
σ	2	the inverse of intertemporal elasticity of substitution (C_t)
σ	0.025	the depreciation rate of private capital (oil and non-oil)
σ^g	0.020	the depreciation rate of public capital
θ	0.9	price stickiness index
ϵ	9	elasticity of substitition for intermediate goods
κ	20	investment adjustment costs parameter
ϕ_y	2	output response in the Taylor rule
ϕ_π	1.8	inflation response in the Taylor rule
α_1	-0.18	exchange rate response in the intervention rule
α_2	-0.57	exchange rate change response in the intervention rule
τ^o	0.27	oil royalty rate
ι^{div}	0.05	the dividend share of oil profit accrued to the government
γ_{GC}=γ_{GI}	0.3	the response of public consumption/investment to fiscal debt
ϑ_{GI}	0.54	the response of public investment to output
ϑ_{GC}	0.3	the response of public consumption to output
γ_{OR}^{GC}	0.2	the response of public consumption to oil revenues
γ_{OR}^{GI}	0.1	the response of public investment to oil revenues
φ_b	0.4	the response of lump-sum taxes to fiscal debt
φ_{OR}	-0.3	the response of lump-sum taxes to oil revenues
φ_C	1	the response of lump-sum taxes to public consumption
φ_I	0.2	the response of lump-sum taxes to public investment
ρ_{GC}	0.53	persistence in public consumption
ρ_{GI}	0	persistence in public investment
$1-\rho_{FDI}$	0.8	FDI response to the world oil prices
ρ_{swf}	0.775	persistence in SWF process
ρ	0.65	interest rate smoothing in the Taylor rule
ρ_{fxr}	0.7	persistence in the foreign exchange reserves of a central bank
ρ_o	0.80	persistence in the world oil prices
σ_0	0.15	standard deviation of the world oil prices
nx_y	0.2	ratio of net export to output
$bstar_y$	0.05	ratio of foreign bond to output
swf_y	0.30	ratio of SWF to output
fdi_y	0.05	ratio of FDI to output
s_o	0.4	share of oil in total output
gI_y	0.35	ratio of public investment to output
gC_y	0.18	ratio of public consumption to output

Table 8.1: Calibration of the DSGE model's parameters.

Parameter	Frequency (%)	Weight in the loss function
Output	0.947	1.56
Exchange rate	0.591	0.97
Inflation	0.288	0.47

Table 8.2: Calibration of the DSGE loss function. Weights are estimated from their respective frequencies in the IMF Article IV releases (2002-2018).

	Loss function (De Paoli)	Text-based loss function
Loss	0.0017	0.0018
$Var(\widehat{\pi}_t)$	0.0014	0.0015
$Var(\widehat{Y}_t)$	0.0002	0.0002
$Var(\widehat{RER}_t)$	0.0001	0.0001

Table 8.3: Normal and text-based DSGE loss functions (as the sum of estimated variances) .

8.1.

Receding oil prices instantly impacts, at lower levels, many key aggregates in the model. However, most deviations from the steady states are mean-reverting with noticeable exceptions for the foreign currency held by households. The oil sector seems to be resilient to such impact, by maintaining its output (yo) and investments (ko) almost unchanged. Noticeable IRFs (yno, gi, gc, b and n) get stabilized after five periods, corresponding to the half life of the oil shock (infl), while fiscal-related variables show important variations then monetary ones.

Increasing government spending, consumption (gc) and investment (gi), has relatively no expansionary effects on the total output (y) but helps alleviating the oil price shock on the non-oil activity (yno) and keeps oil production (yo) practically unchanged (partly due its capital-intensive production function) by supporting investments in capital (slight long-run diminution of kg at 0.1% below its steady state).

From a household perspective, a negative oil shock shifts consumption instantly to an estimated 6% level lower than its steady-state before stabilizing at 1%, mainly driven by the constrained consumption of non-ricardian households. The investment behavior (kno) grows at a steady but lower rate before peaking at +0.2% in the long run. This could be explained by the relative importance of non-ricardian households in the simulation (μ=0.5), which take some time in adjusting their behavior in context of raising prices and currency depreciation; while ricardian households build their foreign assets (bf) to offset the impact during the first three periods, before stabilizing them by taking the opportunity of lower inflation rates (infl) and a minor depreciation of the exchange rate (appreciation of the real exchange rate rer, peaking at 3% after 4 periods) due to monetary policy interventions in a managed exchange rate regime.

Monetary policy interventions anchored on PPT, coupled with sticky prices, prevent soaring the domestic inflation via the exchange rate pass-through, assuming that public investments and consumptions are met by a substantial supply capacity, to not create persistent inflationary tensions in the domestic market. Achieving price stability, in correlation with a stable overall output, bolsters the real sector by maintaining an investment/consumption dynamic that may not hinder intra-sectoral changes. A drop in the number of hours worked (n) during the first 5 periods may be interpreted as difficulties, local businesses face when government limits its action. Such negative effect on unemployment provides another explanation of the drop in the household consumption (c) and

non-oil output (yno). Government subsidies[3], as part of government consumption (gc), could be seen as a catalyst in maintaining the purchasing power of non-ricardian households unchanged, by limiting a potential deterioration over the long run.

This favorable domestic policy generates a sizable amount of public debt (b), which accumulates to a 3% deviation units over the medium run, in correlation to government investments (gi) whose steady deviation reaches 0.8% after ten periods. The accumulation of public debt is explained by measures adopted toward non-oil sector, namely the small and medium sized entreprises (SMEs), as well as subsidies to households (IMF, 2024). Hence, supporting the economic activity during prolonged negative oil outlooks can exacerbate the fiscal debt and may hinder further implementation of countercyclical policies if the negative oil shock is expected to prevail over the long term. For instance, a 10% drop in oil prices shift instantly net exports to a negative deviation level of 15% and returns to equilibrium after ten periods.

Monetary policy interventions in PPT regime, particularly the use of foreign reserves, impacts the sovereign wealth fund (fxr) which is slightly affected by the adopted policy-mix (3% deviation lower than the steady state), in-line with a delayed deterioration of the public debt. This development indicates a policy preference toward keeping a robust external position to face potential sudden stops affecting the balance of payments (euro-dollar fluctuations or a prolonged drop in oil prices), at the expenses of domestic equilibria. The exchange rate depreciation (rer) is delayed and have a narrow fluctuation range, due to the PPT anchor targeting export commodities (here oil prices), at the contrast of CPI anchor which responds to import prices, not to export prices (Frankel, 2011).

The existence of a foreign currency market accessible to households helps ricardian households to face a sudden drop in oil prices by using their savings regarding the amount by which the exchange rate is expected to depreciate in the future. Impulse response function of households foreign holdings (bf) seems to be slightly delayed, but correlated with the real exhchange rate (rer) and domestic prices (infl). This hedging strategy could be seen as a buffer against any shock that may affect the real sector, in addition of being a proxy of domestic savings. By slightly reducing their consumption and under a managed exchange rate regime, non-ricardian households (rule-of-thumb consumers) deliver a consumption smoothing (Galí et al., 2007) that eases potential pressures on the supply side (bottlenecks and inflation spikes over the short run), while benefiting from the government's subsidy policy (social safety nets), aiming at protecting the purchasing power of lower income classes.

8.6 Conclusion

The study of governments' responses to adverse oil shocks has been extensively examined under a single perspective separating fiscal and monetary policy. The lessons learned from the 2008 financial crisis highlighted the necessity to adopt a policy-mix that provides a comprehensive response to support the real sector. Despite huge revenues earned from favorable international prices, oil producers have been facing macroeconomic imbalances following drops in oil prices. Algeria, as a single-commodity exporter, experienced several challenges during episodes of lower oil prices, necessitating a robust and effective response in terms of monetary and fiscal policies to support the economic activity. This paper simulated a negative oil price shock and studied its impact on main macroeconomic aggregates using a DSGE framework. It incorporated a countercyclical fiscal policy associated to a PPT (product price targeting) managed exchange regime with foreign exchange interventions and a substantial share of households having access to a foreign exchange market, due to the non-convertibility of the local currency. Findings confirm the positive impact of this

[3]Subventions could be either direct (social transfers) or indirect (price caps or staple food subsidies).

policy-mix in ensuring the sustainability of the economic activity by supporting the non-oil sector, canceling inflation spikes over the medium run and alleviating losses in the labor market. Under a managed exchange rate regime, the central bank can limit and delay the depreciation of the exchange rate without deteriorating its external position. Hence, canceling a potential inflation pass-through that may lead to a surge in domestic inflation. However, fostering government spending triggers a growing fiscal deficit that cannot be sustainable if oil prices are expected to remain at lower levels over the long run. The presence of ricardian households, endowed with access to foreign exchange market, emulates the existence of a forex market usually linked to the underground economy and responding to fluctuations of the nominal exchange rate. Foreign reserves detained by households are a proxy of private savings, which could be used to support consumption in uncertain times of high inflation. Finally, simulations underscore closer coordination between fiscal and monetary measures when facing a negative oil shock, to be implemented over the short run. In prolonged episodes of lower oil prices, government may be required to set up fiscal consolidation schemes to reduce the accumulated public debt, whose effectiveness hinges on the preservation of external and internal equilibria.

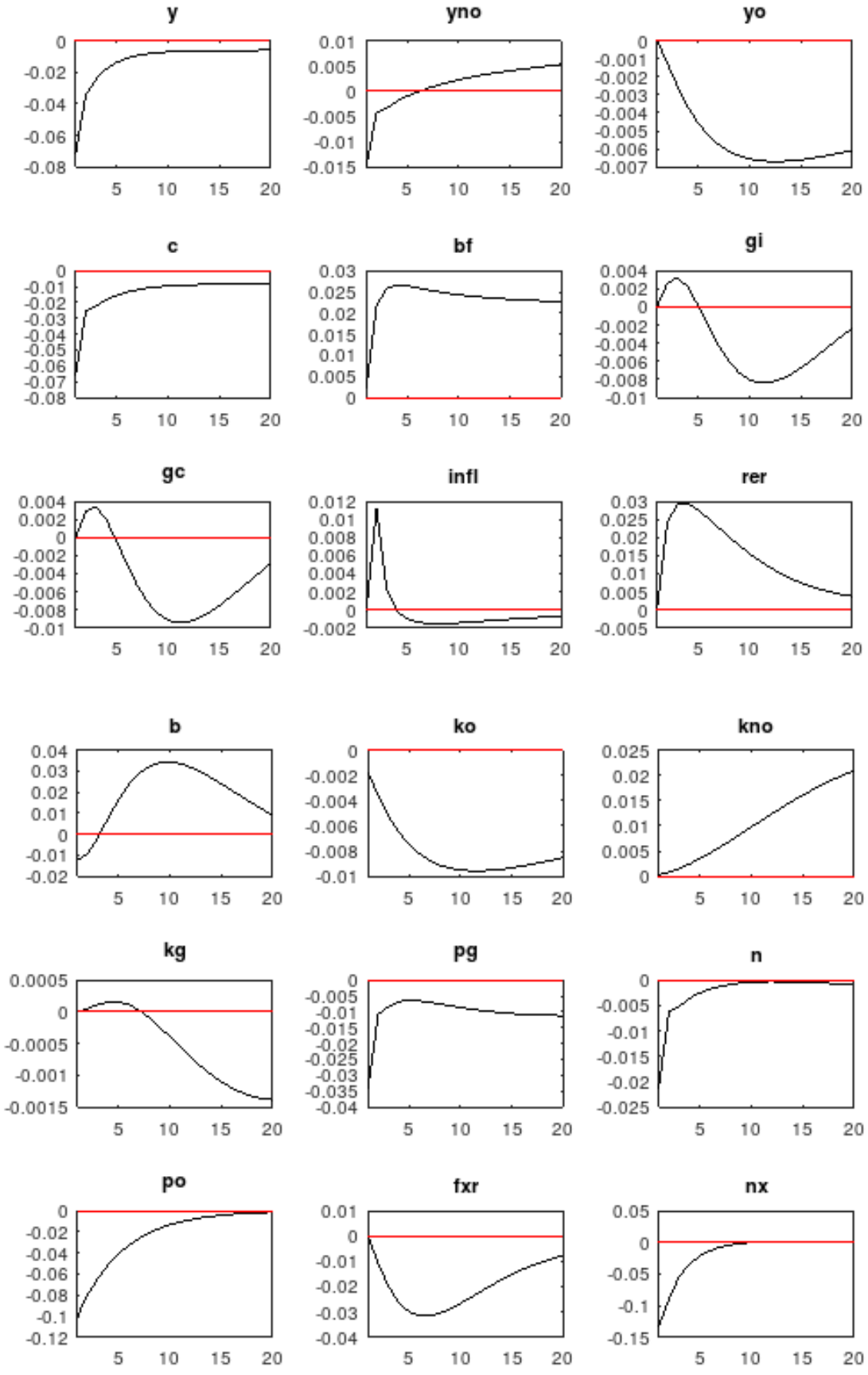


Figure 8.1: Impulse response functions following a 10% drop in international oil prices.

Chapter 9

Conclusion

The digitization wave put the humanity on a fast-track full of innovation and continuous empowerment. The transition from paper-based documents to digital files called the urgent need to develop text analytics, to take advantage of an overwhelming size of available corpora. Mainstream economics, known to have an abundant literature, knew the importance of extracting textual information from written supports and found in earlier text mining methods, a complement to existing econometric inferences. Particularly, the transition of world-leading central banks from secrecy to transparency began with the release of a substantial amount of documents, ready-to-use for the community.

Existing statistical methods and the then word counts were first applied to central bank corpora to extract basic results, assumed to explain many facts found in central banking practices. The adoption of machine learning by economists and communication researchers helped leveraging more facts about central bankers around the world and testing several hypotheses related to the inner functioning of monetary committees. With the continuous improvements, in terms of interpretability and results, earlier results proved to be agnostic and less informative, compared to advanced methods belonging to NLP.

Throughout this thesis, the readers get acquainted with common stylized facts found in central bank communication and ways to explore them using advanced text mining techniques and NLP methods. The complexity of communication adopted by many central banks requires selecting methods that go in-line with the corpus' characteristics.

Central bankers' preferences (Chapter 2) in terms of inflation and interest rates was probably the earliest issue researchers wanted to reveal. A combination of advanced topic models and iterative regressions was needed to estimate preferences at the Federal Reserve and FOMC, based on existing speeches and meeting transcripts. Findings confirmed corpus-based, speaker differences that hints an established institutional communication strategy, rather than a per-speaker communication strategy.

A particular feature in the European Union, is the adoption of a multilingual communication policy at several institutions. Central banks in Germany and Switzerland do use a multilingual strategy for external communication. This proved to be informative when scaling the board members in Germany (Chapter 3) according to their assigned portfolios, and identifying specific topics in Switzerland that turned out to be aligned with local business cycle fluctuations.

One fact inherited from the post-war monetary practices is ambiguity. The absence of clarity in the central bank discourse prevented researchers from understanding mechanisms behind monetary policy decisions around the world. From one perspective, ambiguity is seen as a communication strategy, leveraged by adamant speakers avoiding confusion and false signals. Such behavior could

be scaled using computationally-intensive scaling (Chapter 4), built upon probabilistic assumptions, rather than a biased human judgment.

The recession, that hit many countries in the world during the 2007-2009 period, stressed out the importance of quantifying uncertainty, as a synthetic index for policy purposes. While earlier attempts relied on dictionary and wordlists to gauge uncertainty, Chapter 5 proved the usefulness of semantic search models in discerning the uncertainty context and testing its variations in many corpora.

While topic models remains the most adopted technique used so far to study central bank corpora, researchers recognize their static structure is unable to uncover causal patterns in selected terms. This prevented uncovering causality evidences whose interpretability exceeds mostly-agnostic topic model results. Chapter 6 was an attempt to reveal causal links in given corpora, particularly at the U.S. Federal Reserve. The aim was to support earlier chapters with sound linguistic and realistic causal patterns that explain the monetary policy rhetoric, with the help of a hybrid network analysis-decision tree approach built on language morphologies.

The current hype around LLMs opened new research paths and application fields for practitioners interested in the text-as-data field. While many see LLMs as solutions, they mostly remain products whose users totally ignore the inner architecture and the trained corpora. Issues pertaining to the blind use of LLMs were detailed in Chapter 7, with the adoption of advanced techniques from Random Matrix Theory to study the inner functioning of LLMs as complex physical systems. The findings stress out the importance of testing LLMs before putting them into practice especially for high-level tasks as for scaling preferences and institutional ranking. An experiment on uncertainty and tonality measurement confirmed difficulties to get stable results that match the ground truth, however, promising insights could be retrieved thanks to reasoning capabilities featured by some LLMs.

Mainstream monetary economics has been heavily relying on dynamic econometric modeling to evaluation policy impacts, what is known as DSGE modeling. Such models comprise a set of equations related to each agent behavior and constraint, plus mathematical schemes to make the estimation possible. Chapter 8 genuinely used word frequencies to calibrate the loss function of a given monetary authority, following text-based preferences found on inflation, exchange rate and output. Such procedures permit a sound parameter calibration that follows the rhetoric used by the central bank when tackling these themes.

This work navigated the dodgy path of central bank communication with different techniques borrowed from statistics, computer science and linguistics. Testing several hypotheses about central bankers revealed facts that were either ignored by scholars, or simply confined in a theoretical debate. Findings confirm the severe difficulty of tackling monetary policy corpora and the delicate task of selecting the appropriate methods that maximize information extraction.

This highlights the role of reproducibility in research and transparency in results, often ignored by users who only consider applying few techniques, without testing their weaknesses and strengths. It is up to practitioners to tailor the methodology to test their hypotheses, without leveraging post-hoc adjustments and biased human-based scheme. By doing so, they guarantee a wide reproducibility of their results and genuine improvements.

Appendix A

Bibliography

Adjemian, S., Juillard, M., Karamé, F., Mutschler, W., Pfeifer, J., Ratto, M., Rion, N., and Villemot, S. Dynare: Reference Manual, Version 6. Dynare Working Papers 80, CEPREMAP, February 2024. URL https://ideas.repec.org/p/cpm/dynare/080.html.

Ahir, H., Bloom, N., and Furceri, D. The world uncertainty index. Working Paper 29763, National Bureau of Economic Research, February 2022.

Algozhina, A. Monetary Policy Rule, Exchange Rate Regime, and Fiscal Policy Cyclicality in a Developing Oil Economy. Dynare Working Papers 49, CEPREMAP, May 2016. URL https://ideas.repec.org/p/cpm/dynare/049.html.

Allegret, J.-P. and Benkhodja, M. T. External shocks and monetary policy in a small open oil exporting economy. Post-Print hal-01411705, HAL, 2012. URL https://ideas.repec.org/p/hal/journl/hal-01411705.html.

Andersson, M., Dillén, H., and Sellin, P. Monetary policy signaling and movements in the term structure of interest rates. Journal of Monetary Economics, 53(8):1815–1855, 2006. ISSN 0304-3932. doi: https://doi.org/10.1016/j.jmoneco.2006.06.002.

Andre, P., Haaland, I., Roth, C., and WohlfartJ, o. Narratives about the macroeconomy. Technical report, University of Bonn and University of Mannheim, Germany, May 2022.

Angelov, D. Top2vec: Distributed representations of topics. arXiv preprint arXiv:2008.09470, 2020. URL https://arxiv.org/pdf/2008.09470.

Ash, E., Gauthier, G., and Widmer, P. Relatio: Text semantics capture political and economic narratives. arXiv, 2021. URL https://arxiv.org/pdf/2108.01720.

Ash, E., Gauthier, G., and Widmer, P. Relatio: Text semantics capture political and economic narratives. Political Analysis, 32(1):115–132, 2024.

Athanassiou, P. The application of multilingualism in the european union context, 2006. URL https://www.ecb.europa.eu/pub/pdf/scplps/ecblwp2.pdf?8108c08740d28b7227fda550a5ec39ed.

Baerg, N. Crafting Consensus: Why Central Bankers Change their Speech and how Speech Changes the Economy. Oxford University Press, New York, NY, United States of America, 2020. ISBN 9780190499488.

Baerg, N. and Binder, C. Automated detection of emotion in central bank communication: A warning. National Institute Economic Review, 269:82–91, 2024. doi: 10.1017/nie.2024.31.

Baerg, N. and Lowe, W. A textual taylor rule: estimating central bank preferences combining topic and scaling methods. Political Science Research and Methods, 8(1):106–122, 2020. doi: 10.1017/psrm.2018.31.

Baker, S. R., Bloom, N., and Davis, S. J. Measuring economic policy uncertainty*. The Quarterly Journal of Economics, 131(4):1593–1636, 2016. ISSN 0033-5533.

Bauer, M. D. and Swanson, E. T. A reassessment of monetary policy surprises and high-frequency identification. NBER Macroeconomics Annual, 37:87–155, 2023. doi: 10.1086/723574. URL https://doi.org/10.1086/723574.

Baumgaertner, M. and Zahner, J. Whatever it takes to understand a central banker - Embedding their words using neural networks. MAGKS Papers on Economics 202130, Philipps-Universität Marburg, Faculty of Business Administration and Economics, Department of Economics (Volkswirtschaftliche Abteilung), 2021. URL https://ideas.repec.org/p/mar/magkse/202130.html.

Bańkowski, K., Christoffel, K., and Faria, T. Assessing the fiscal-monetary policy mix in the euro area. Working Paper Series 2623, European Central Bank, December 2021. URL https://ideas.repec.org/p/ecb/ecbwps/20212623.html.

Benes, J., Berg, A., Portillo, R., and Vavra, D. Modeling Sterilized Interventions and Balance Sheet Effects of Monetary Policy in a New-Keynesian Framework. Open Economies Review, 26(1):81–108, February 2015. URL https://ideas.repec.org/a/kap/openec/v26y2015i1p81-108.html.

Bennani, H., Fanta, N., Gertler, P., and Horvath, R. Does central bank communication signal future monetary policy? the case of the ecb. Working paper, Institute of Economic Studies, Faculty of Social Sciences, Charles University in Prague, 2019. URL https://www.econstor.eu/bitstream/10419/203231/1/1666601551.pdf.

Benzécri, J.-P. L'analyse des données: Benzécri, J.-P. et al. L'analyse des correspondances, volume 2. Dunod, 1973.

Berg, A., Portillo, R., Yang, S.-C. S., and Zanna, L.-F. Public Investment in Resource-Abundant Developing Countries. IMF Economic Review, 61(1):92–129, April 2013. URL https://ideas.repec.org/a/pal/imfecr/v61y2013i1p92-129.html.

Bhonde, S., Paikrao, R., and Rahane, K. Text association analysis and ambiguity in text mining. 1324, 11 2010. doi: 10.1063/1.3526195.

Bianchi, F., Terragni, S., and Hovy, D. 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), pages 759–766, Online, aug 2021a. Association for Computational Linguistics. doi: 10.18653/v1/2021.acl-short.96. URL https://aclanthology.org/2021.acl-short.96.

Bianchi, F., Terragni, S., Hovy, D., Nozza, D., and Fersini, E. Cross-lingual contextualized topic models with zero-shot learning. In Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume, pages 1676–1683, Online, apr 2021b. Association for Computational Linguistics. doi: 10.18653/v1/2021.eacl-main.143. URL https://aclanthology.org/2021.eacl-main.143.

Black, M. Vagueness. An Exercise in Logical Analysis. Philosophy of Science, 1937.

Blei, D. M. and Lafferty, J. D. Dynamic topic models. In Proceedings of the 23rd International Conference on Machine Learning, ICML '06, page 113–120, New York, NY, USA, 2006. Association for Computing Machinery. ISBN 1595933832.

Blei, D. M., Ng, A. Y., and Jordan, M. I. Latent dirichlet allocation. J. Mach. Learn. Res., 3(null): 993–1022, March 2003. ISSN 1532-4435.

Blinder, A. S., Ehrmann, M., Fratzscher, M., De Haan, J., and Jansen, D.-J. Central bank communication and monetary policy: A survey of theory and evidence. Working Paper 13932, National Bureau of Economic Research, April 2008. URL http://www.nber.org/papers/w13932.

Blondel, V. D., Guillaume, J.-L., Lambiotte, R., and Lefebvre, E. Fast unfolding of communities in large networks. Journal of Statistical Mechanics: Theory and Experiment, 2008(10):P10008, oct 2008.

Boucekkine, R., Laksaci, M., and Touati-Tliba, M. Long-run stability of money demand and monetary policy: The case of algeria. The Journal of Economic Asymmetries, 24:e00217, 2021. ISSN 1703-4949. doi: https://doi.org/10.1016/j.jeca.2021.e00217. URL https://www.sciencedirect.com/science/article/pii/S1703494921000220.

Boyd-Graber, J., Mimno, D., and Newman, D. Care and feeding of topic models: Problems, diagnostics, and improvements. In Handbook of Mixed Membership Models and Their Applications, CRC Handbooks of Modern Statistical Methods. CRC Press, 2014.

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D., Wu, J., Winter, C., Hesse, C., Chen, M., Sigler, E., Litwin, M., Gray, S., Chess, B., Clark, J., Berner, C., McCandlish, S., Radford, A., Sutskever, I., and Amodei, D. Language models are few-shot learners. In Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M., and Lin, H., editors, Advances in Neural Information Processing Systems, volume 33, pages 1877–1901. Curran Associates, Inc., 2020.

Calvo, G. A. Staggered prices in a utility-maximizing framework. Journal of Monetary Economics, 12(3):383–398, September 1983. URL https://ideas.repec.org/a/eee/moneco/v12y1983i3p383-398.html.

Chen, D. and Manning, C. A fast and accurate dependency parser using neural networks. In Moschitti, A., Pang, B., and Daelemans, W., editors, Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), pages 740–750, Doha, Qatar, October 2014. Association for Computational Linguistics.

Chibi, A., Benbouziane, M., and Chekouri, S. M. Interaction between monetary and fiscal policy : Empirical evidence from algeria. 25th Annual Conference. Kuwait, March 10-12, 2019. Economic Research Forum, 2019. URL https://api.semanticscholar.org/CorpusID:203631167.

Clauset, A., Shalizi, C. R., and Newman, M. E. J. Power-law distributions in empirical data. SIAM Review, 51(4):661–703, 2009. doi: 10.1137/070710111. URL https://doi.org/10.1137/070710111.

de Kok, M., Rebboud, Y., Lisena, P., Raphaël, T., and Tiddi, I. From nodes to narratives: A knowledge graph-based storytelling approach. In CEUR, editor, TEXT2STORY 2024, 7th International Workshop on Narrative Extraction from Texts (Text2Story), colocated with ECIR 2024, 24 March 2024, Glasgow, UK, Glasgow, 2024. CEUR.

De Paoli, B. Monetary policy and welfare in a small open economy. Journal of International Economics, 77(1):11–22, 2009. ISSN 0022-1996. doi: https://doi.org/10.1016/j.jinteco.2008.09.007. URL https://www.sciencedirect.com/science/article/pii/S0022199608001001.

Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., and Harshman, R. Indexing by latent semantic analysis. Journal of the American Society for Information Science, 41(6):391–407, 1990.

Derényi, I., Palla, G., and Vicsek, T. Clique percolation in random networks. Phys. Rev. Lett., 94: 160202, Apr 2005.

Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. Bert: Pre-training of deep bidirectional transformers for language understanding. In Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers), pages 4171–4186, 2019. doi: 10.18653/v1/N19-1423. URL https://aclanthology.org/N19-142.

Diaf, S. Policy preference at central banks: Quantifying monetary policy signals using keyword topic models. WiSo-HH working paper series 69, 2022. doi: https://d-nb.info/1273646363/34.

Diaf, S. CommunityFish: A Poisson-based document scaling with hierarchical clustering. In Abbas, M. and Freihat, A. A., editors, Proceedings of the 6th International Conference on Natural Language and Speech Processing (ICNLSP 2023), pages 59–67, Online, December 2023. Association for Computational Linguistics.

Diaf, S. CliqueCorex: A self-supervised clique-based anchored topic model. In Abbas, M. and Freihat, A. A., editors, Proceedings of the 7th International Conference on Natural Language and Speech Processing (ICNLSP 2024), pages 74–82, Trento, October 2024. Association for Computational Linguistics. URL https://aclanthology.org/2024.icnlsp-1.9/.

Diaf, S. and Fritsche, U. TopicShoal: Scaling partisanship using semantic search. In Proceedings of the 18th Conference on Natural Language Processing (KONVENS 2022), pages 167–174, Potsdam, Germany, 12–15 sep 2022. KONVENS 2022 Organizers. URL https://aclanthology.org/2022.konvens-1.20.

Diaf, S. and Schütze, F. Estimating uncertainty in narrative economics using latent semantic scaling. In JADT 2022 - 16th International Conference on Statistical Analysis of Textual Data, volume Naples, Italy, 6 - 8 July 2022, pages 321–327. 2022.

Diaf, S. and Schütze, F. Uncovering uncertainty in narrative economics: A semantic search approach. In New Frontiers in Textual Data Analysis. Springer Nature Switzerland, 2024. doi: https://doi.org/10.1007/978-3-031-55917-4_1.

Dib, A. Implications of the global financial crisis on the algerian economy. Revue des économies nord Africaines, 7, 2010. URL https://www.univ-chlef.dz/renaf/Articles_Renaf_N_07/Article_17_en.pdf.

Dieng, A. B., Ruiz, F. J. R., and Blei, D. M. Topic modeling in embedding spaces. arXiv preprint arXiv:1907.04907, 2019. URL https://arxiv.org/pdf/1907.04907.

Duncan, O. D. Introduction to structural equation models /. Academic Press,, New York :, 1975. Includes indexes.

ECB, E. C. B. The monetary policy of the ecb, 2011. URL https://www.ecb.europa.eu/pub/pdf/other/monetarypolicy2011en.pdf.

Echterhoff, J. M., Liu, Y., Alessa, A., McAuley, J., and He, Z. Cognitive bias in decision-making with LLMs. In Al-Onaizan, Y., Bansal, M., and Chen, Y.-N., editors, Findings of the Association for Computational Linguistics: EMNLP 2024, pages 12640–12653, Miami, Florida, USA, Nov 2024. Association for Computational Linguistics. doi: 10.18653/v1/2024.findings-emnlp.739. URL https://aclanthology.org/2024.findings-emnlp.739/.

Edison, H. and Carcel, H. Text data analysis using latent dirichlet allocation: an application to fomc transcripts. Applied Economics Letters, 0(0):1–5, 2020. doi: 10.1080/13504851.2020.1730748. URL https://doi.org/10.1080/13504851.2020.1730748.

Eijffinger, S., Mahieu, R., and Raes, L. Estimating the preferences of central bankers: An analysis of four voting records. Workingpaper, Economics, 2013. Pagination: 26.

Eisner, J. Bilexical Grammars and their Cubic-Time Parsing Algorithms, pages 29–61. Springer Netherlands, Dordrecht, 2000. ISBN 978-94-015-9470-7.

Entman, R. M. Framing: Toward clarification of a fractured paradigm. Journal of Communication, 43(4):51–58, 1993. ISSN 0021-9916.

Eshima, S., Imai, K., and Sasaki, T. Keyword assisted topic models. ArXiv, abs/2004.05964, 2020. URL https://arxiv.org/pdf/2004.05964.

Eshima, S., Imai, K., and Sasaki, T. Keyword assisted topic models, 2023. URL https://arxiv.org/abs/2004.05964.

Faia, E. and Iliopulos, E. Financial openness, financial frictions and optimal monetary policy. Post-Print hal-00628737, HAL, November 2011. URL https://ideas.repec.org/p/hal/journl/hal-00628737.html.

Fawareh, H. A., Jusoh, S. S., and Osman, W. S. E. Ambiguity in text mining. 2008 International Conference on Computer and Communication Engineering, pages 1172–1176, 2008. URL https://api.semanticscholar.org/CorpusID:14775254.

Federal Reserve, F. What economic goals does the federal reserve seek to achieve through its monetary policy? Technical report, National Bureau of Economic Research, 2020.

Frankel, J. A. A Comparison of Product Price Targeting and Other Monetary Anchor Options for Commodity Exporters in Latin America. Technical report, 2011.

Galí, J. Monetary Policy, Inflation, and the Business Cycle: An Introduction to the New Keynesian Framework and Its Applications Second edition. Number 10495 in Economics Books. Princeton University Press, 2015. URL https://ideas.repec.org/b/pup/pbooks/10495.html.

Galí, J., López-Salido, J. D., and Vallés, J. Understanding the effects of government spending on consumption. Journal of the European Economic Association, 5(1):227–270, 2007. doi: 10.1162/JEEA.2007.5.1.227. URL https://onlinelibrary.wiley.com/doi/abs/10.1162/JEEA.2007.5.1.227.

Gentzkow, M., Kelly, B., and Taddy, M. Text as data. Journal of Economic Literature, 57(3):535–574, September 2019a. ISSN 0022-0515. doi: 10.1257/jel.20181020. URL https://www.aeaweb.org/articles?id=10.1257/jel.20181020.

Gentzkow, M., Kelly, B., and Taddy, M. Text as data. Journal of Economic Literature, 57(3):535–74, September 2019b.

Goet, N. D. Measuring polarization with text analysis: Evidence from the uk house of commons, 1811–2015. Political Analysis, 27(4):518–539, 2019. ISSN 1047-1987. doi: 10.1017/pan.2019.2.

Goodfriend, M. Monetary mystique: Secrecy and central banking. Journal of Monetary Economics, 17(1):63–92, 1986. ISSN 0304-3932. doi: https://doi.org/10.1016/0304-3932(86)90006-1. URL https://www.sciencedirect.com/science/article/pii/0304393286900061.

Gorodnichenko, Y., Pham, T., and Talavera, O. The voice of monetary policy. American Economic Review, 113(2):548–84, February 2023. doi: 10.1257/aer.20220129. URL https://www.aeaweb.org/articles?id=10.1257/aer.20220129.

Grassi, M. and Tarantino, B. Semtree: tree-based structure learning methods with structural equation models. Bioinformatics, 39(6):btad377, 06 2023. ISSN 1367-4811.

Greenspan, A. Risk and uncertainty in monetary policy. American Economic Review, 94(2):33–40, 2004. ISSN 0002-8282. doi: 10.1257/0002828041301551.

Grimmer, J. and Stewart, B. M. Text as data: The promise and pitfalls of automatic content analysis methods for political texts. Political Analysis, 21(3):267–297, 2013. ISSN 1047-1987. doi: 10.1093/pan/mps028.

Grimmer, J., Roberts, M. E., and Stewart, B. M. Text as Data: A New Framework for Machine Learning and the Social Sciences. Princeton University Press, United States of America, 2022. ISBN 0691207550.

Grootendorst, M. Bertopic: Neural topic modeling with a class-based tf-idf procedure. arXiv preprint arXiv:2203.05794, 2022.

Haldane, A. and McMahon, M. Central bank communications and the general public. AEA Papers and Proceedings, 108:578–83, May 2018. doi: 10.1257/pandp.20181082. URL https://www.aeaweb.org/articles?id=10.1257/pandp.20181082.

Hansen, S., McMahon, M., and Prat, A. Transparency and Deliberation Within the FOMC: A Computational Linguistics Approach*. The Quarterly Journal of Economics, 133(2):801–870, 10 2017. ISSN 0033-5533. doi: 10.1093/qje/qjx045. URL https://doi.org/10.1093/qje/qjx045.

Hansen, S., McMahon, M., and Prat, A. Transparency and deliberation within the fomc: A computational linguistics approach*. The Quarterly Journal of Economics, 133(2):801–870, 2018. ISSN 0033-5533.

Hansson, M. Evolution of topics in central bank speech communication, 2021.

Hartmann, P. and Smets, F. The first twenty years of the European Central Bank: monetary policy. Working Paper Series 2219, European Central Bank, Dec 2018. URL https://ideas.repec.org/p/ecb/ecbwps/20182219.html.

Hasan, A. M., Rassem, T. H., Noor, N. M., and Hasan, A. M. A review of recent trends: Text mining of taxonomy using wordnet 3.1 for the solution and problems of ambiguity in social media. In Peng, S.-L., Son, L. H., Suseendran, G., and Balaganesh, D., editors, Intelligent Computing and Innovation on Data Science, pages 137–152, Singapore, 2020. Springer Singapore. ISBN 978-981-15-3284-9.

Havrilesky, T. and Gildea, J. A. The policy preferences of fomc members as revealed by dissenting votes: Comment. Journal of Money, Credit and Banking, 23(1):130–138, 1991. ISSN 00222879, 15384616. URL http://www.jstor.org/stable/1992770.

Hix, S., Høyland, B., and Vivyan, N. From doves to hawks: A spatial analysis of voting in the monetary policy committee of the bank of england. European Journal of Political Research, 49(6): 731–758, 2010.

Honnibal, M., Montani, I., Van Landeghem, S., and Boyd, A. spaCy: Industrial-strength Natural Language Processing in Python, 2020. URL https://doi.org/10.5281/zenodo.1212303.

Hu, Y. and Khan, L. Uncertainty-aware reliable text classification. In Zhu, F., editor, Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining, ACM Digital Library, pages 628–636, New York,NY,United States, 2021. Association for Computing Machinery. ISBN 9781450383325.

Hubert, P. Qualitative and quantitative central bank communication and inflation expectations. The B.E. Journal of Macroeconomics, 17(1):20160024, 2017. doi: https://doi.org/10.1515/bejm-2016-0024. URL https://www.degruyter.com/view/journals/bejm/17/1/article-20160024.xml.

Husted, L., Rogers, J., and Sun, B. Monetary policy uncertainty. Journal of Monetary Economics, 115:20–36, 2020. ISSN 0304-3932.

Husted, L. F., Rogers, J. H., and Sun, B. Monetary policy uncertainty. International Finance Discussion Papers 1215, Board of Governors of the Federal Reserve System (U.S.), Oct 2017. URL https://ideas.repec.org/p/fip/fedgif/1215.html.

Hüning, H. Swiss national bank communication and investors' uncertainty. The North American Journal of Economics and Finance, 51:101024, 2020. ISSN 1062-9408. doi: https://doi.org/10.1016/j.najef.2019.101024. URL https://www.sciencedirect.com/science/article/pii/S1062940819301366.

IMF. 2018 Article IV Consultation-Press Release; Staff Report; and Statement by the Executive Director for Algeria. Technical report, International Monetary Fund. Middle East and Central Asia Department., Washington D.C., 2018.

IMF. 2023 Article IV Consultation-Press Release; Staff Report; and Statement by the Executive Director for Algeria. Technical report, International Monetary Fund. Middle East and Central Asia Department., Washington D.C., 2024.

Istrefi, K. In fed watchers' eyes: Hawks, doves and monetary policy. Working Paper 725, Banque de France, 2017. URL https://publications.banque-france.fr/en/fed-watchers-eyes-hawks-doves-and-monetary-policy.

Jagarlamudi, J., Daumé, H., and Udupa, R. Incorporating lexical priors into topic models. In EACL, 2012.

Jakobsen, M. E., Shah, R. D., Bühlmann, P., and Peters, J. Structure learning for directed trees. Journal of Machine Learning Research, 23(159):1–97, 2022.

Jean, P.-A., Harispe, S., Ranwez, S., Bellot, P., and Montmain, J. Uncertainty detection in natural language. In Akerkar, R., editor, Proceedings of the 6th International Conference on Web Intelligence, Mining and Semantics, ACM Digital Library, pages 1–10, New York, NY, 2016. ACM. ISBN 9781450340564.

Kádár, Á., Xiao, L., Kemertas, M., Fancellu, F., Jepson, A., and Fazly, A. Dependency parsing with structure preserving embeddings. In Merlo, P., Tiedemann, J., and Tsarfaty, R., editors, Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume, pages 1684–1697, Online, April 2021. Association for Computational Linguistics.

Karim El-Ouaghlidi, Matthias Gomolka, and Jens Orben. Bundesbank speeches: Data report 2019-12. (12), 2019.

Keith, K., Teichmann, C., O'Connor, B., and Meij, E. Uncertainty over uncertainty: Investigating the assumptions, annotations, and text measurements of economic policy uncertainty. Association for Computational Linguistics, (Proceedings of the Fourth Workshop on Natural Language Processing and Computational Social Science):116–131, 2020.

Kerdjoudj, F. and Curé, O. Evaluating Uncertainty in Textual Document. In URSW at ISWC, Bethlehem, United States, 2015. URL https://hal.archives-ouvertes.fr/hal-01740539.

Keynes, J. M. The general theory of employment, interest and money. London : Macmillan, 1936, 1936.

Knight, F. H. Risk, uncertainty and profit. Houghton Mifflin, Boston, 1921.

Kou, L., Markowsky, G., and Berman, L. A fast algorithm for steiner trees. Acta informatica, 15: 141–145, 1981.

Kwon, B., Park, T., Perez-Cruz, F., and Rungcharoenkitkul, P. Large language models: a primer for economists. BIS Quarterly Review, None(None):None, December 2024. doi: None. URL https://ideas.repec.org/a/bis/bisqtr/2412b.html.

Lange, K.-R., Reccius, M., Schmidt, T., Müller, H., Roos, M. W. M., and Jentsch, C. Towards extracting collective economic narratives from texts. Ruhr Economic Papers 963, RWI - Leibniz-Institut für Wirtschaftsforschung, Ruhr-University Bochum, TU Dortmund University, University of Duisburg-Essen, 2022. URL https://ideas.repec.org/p/zbw/rwirep/963.html.

Lauderdale, B. E. and Herzog, A. Measuring political positions from legislative speech. Political Analysis, 24(3):374–394, 2016. doi: 10.1093/pan/mpw017.

Laver, M., Benoit, K., and Garry, J. Extracting policy positions from political texts using words as data. American Political Science Review, 97(02), 2003. ISSN 0003-0554. doi: 10.1017/S0003055403000698.

Le, Q. and Mikolov, T. Distributed representations of sentences and documents. In Proceedings of the 31st International Conference on International Conference on Machine Learning - Volume 32, ICML'14, page II–1188–II–1196. JMLR.org, 2014.

Lefaudeux, B., Massa, F., Liskovich, D., Xiong, W., Caggiano, V., Naren, S., Xu, M., Hu, J., Tintore, M., Zhang, S., Labatut, P., Haziza, D., Wehrstedt, L., Reizenstein, J., and Sizov, G. xformers: A modular and hackable transformer modelling library. https://github.com/facebookresearch/xformers, 2022.

Lehmann, P., Franzmann, S., Al-Gaddooa, D., Burst, T., Ivanusch, C., Regel, S., Riethmüller, F., Volkens, A., Weßels, B., and Zehnter, L. The manifesto data collection. manifesto project (mrg/cmp/marpor). version 2024a, 2024.

Lo, J., Proksch, S.-O., and Slapin, J. B. Ideological clarity in multiparty competition: A new measure and test using election manifestos. British Journal of Political Science, 46(3):591–610, 2016. doi: 10.1017/S0007123414000192.

Lou, X., Hu, Y., and Li, X. Learning linear polytree structural equation models, 2024.

Loughran, T. and McDonald, B. When is a liability not a liability? textual analysis, dictionaries, and 10-ks. The Journal of Finance, 66(1):35–65, 2011. ISSN 00221082.

Lu, B., Ott, M., Cardie, C., and Tsou, B. K. Multi-aspect sentiment analysis with topic models. In 2011 IEEE 11th International Conference on Data Mining Workshops, pages 81–88, 2011.

Lucas, C., Nielsen, R. A., Roberts, M. E., Stewart, B. M., Storer, A., and Tingley, D. Computer-assisted text analysis for comparative politics. Political Analysis, 23(2):254–277, 2015. ISSN 10471987, 14764989. URL http://www.jstor.org/stable/24572972.

Lähner, T. Inconsistent voting behaviour in the FOMC. Applied Economics, 50(14):1617–1643, mar 2018. ISSN 0003-6846. doi: 10.1080/00036846.2017.1368994. Routledge.

Macklem, T. and Vardy, J. 20 years of central bank communications, and lessons for the future. In for International Settlements, B., editor, Central banking in the Americas: Lessons from two decades, volume 127 of BIS Papers chapters, pages 55–67. Bank for International Settlements, december 2023. URL https://ideas.repec.org/h/bis/bisbpc/143-05.html.

Malevergne, Y. and Sornette, D. Collective origin of the coexistence of apparent random matrix theory noise and of factors in large sample correlation matrices. Physica A: Statistical Mechanics and its Applications, 331(3):660–668, 2004. ISSN 0378-4371. doi: https://doi.org/10.1016/j.physa.2003.09.004. URL https://www.sciencedirect.com/science/article/pii/S0378437103007969.

Mandelbrot, B. B. The fractal geometry of nature. W. H. Freeman and Comp., New York, 1982.

Mankiw, N. G. The Savers-Spenders Theory of Fiscal Policy. American Economic Review, 90(2):120–125, May 2000. URL https://ideas.repec.org/a/aea/aecrev/v90y2000i2p120-125.html.

Marchenko, V. A. and Pastur, L. A. Distribution of eigenvalues for some sets of random matrices. Math. USSR-Sb, 1:422–437, 1967. URL https://www.mathnet.ru/links/6db725fdd0b1555b2deea01021d5af48/sm4101.pdf.

Martin, C. H. and Mahoney, M. W. Implicit self-regularization in deep neural networks: evidence from random matrix theory and implications for learning. Journal of Machine Learning Research, 22(1), jan 2021. ISSN 1532-4435. URL https://www.jmlr.org/papers/volume22/20-410/20-410.pdf.

Martin, C. H., Peng, T., and Mahoney, M. W. Predicting trends in the quality of state-of-the-art neural networks without access to training or testing data. Nature Communications, 12(1), jul 2021. ISSN 2041-1723. doi: 10.1038/s41467-021-24025-8. URL http://dx.doi.org/10.1038/s41467-021-24025-8.

Masayuki, K. and Yosuke, T. The Art of Central Bank Communication: A Topic Analysis on Words used by the Bank of Japan's Governors. Discussion papers 19038, Research Institute of Economy, Trade and Industry (RIETI), May 2019. URL https://ideas.repec.org/p/eti/dpaper/19038.html.

Mcauliffe, J. D. and Blei, D. M. Supervised topic models. In Platt, J. C., Koller, D., Singer, Y., and Roweis, S. T., editors, Advances in Neural Information Processing Systems 20, pages 121–128. Curran Associates, Inc., 2008.

McDonald, R., Crammer, K., and Pereira, F. Online large-margin training of dependency parsers. In Knight, K., Ng, H. T., and Oflazer, K., editors, Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL'05), pages 91–98, Ann Arbor, Michigan, June 2005. Association for Computational Linguistics.

McGowan, L. D., Gerke, T., and Barrett, M. Causal inference is not just a statistics problem. Journal of Statistics and Data Science Education, 32(2):150–155, 2024.

McGregor, R. R. Fomc voting behavior and electoral cycles: Partisan ideology and partisan loyalty. Economics & Politics, 8(1):17–32, 1996. doi: https://doi.org/10.1111/j.1468-0343.1996.tb00118.x.

McInnes, L., Healy, J., and Astels, S. hdbscan: Hierarchical density based clustering. The Journal of Open Source Software, 2(11):205, 2017. doi: 10.21105/joss.00205.

McInnes, L., Healy, J., Saul, N., and Großberger, L. Umap: Uniform manifold approximation and projection. The Journal of Open Source Software, 3(29):861, 2018. doi: 10.21105/joss.00861.

Meade, E. E. The FOMC: preferences, voting, and consensus. Review, 87(Mar):93–101, 2005. URL https://ideas.repec.org/a/fip/fedlrv/y2005imarp93-101nv.87no.2,pt.1.html.

Meade, E. E. and Stasavage, D. Publicity of debate and the incentive to dissent: Evidence from the us federal reserve. The Economic Journal, 118(528):695–717, 2008. ISSN 0013-0133.

Mikolov, T., Chen, K., Corrado, G., and Dean, J. Efficient estimation of word representations in vector space, 2013. URL https://arxiv.org/pdf/1301.3781.

Muchlinski, E. Conclusion, pages 224–225. Palgrave Macmillan UK, London, 2011. ISBN 978-0-230-30596-0. doi: 10.1057/9780230305960_7. URL https://doi.org/10.1057/9780230305960_7.

of Algeria, B. Rapport Annuel 2023 : Evolution Economique et Monétaire. Technical report, 2024.

Packard, M. D., Bylund, P. L., and Clark, B. B. Keynes and Knight on uncertainty: peas in a pod or chalk and cheese? Cambridge Journal of Economics, 45(5):1099–1125, 07 2021. ISSN 0309-166X.

Papakyriakopoulos, O., Hegelich, S., Serrano, J. C. M., and Marco, F. Bias in word embeddings. In Hildebrandt, M., editor, Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, ACM Digital Library, pages 446–457, New York,NY,United States, 2020. Association for Computing Machinery. ISBN 9781450369367. doi: 10.1145/3351095.3372843.

Pearl, J. Causality: Models, Reasoning and Inference. Cambridge University Press, USA, 2nd edition, 2009. ISBN 052189560X.

Pfeifer, M. and Marohl, V. P. Centralbankroberta: A fine-tuned large language model for central bank communications. The Journal of Finance and Data Science, 9:100114, 2023. ISSN 2405-9188. doi: https://doi.org/10.1016/j.jfds.2023.100114. URL https://www.sciencedirect.com/science/article/pii/S2405918823000302.

Reimers, N. and Gurevych, I. Sentence-bert: Sentence embeddings using siamese bert-networks, 2019.

Roberts, M., Stewart, B., Dustin, T., and Airoldi, E. The structural topic model and applied social science. Neural Information Processing Society, 2013.

Roos, M. and Reccius, M. Narratives in economics. Journal of Economic Surveys, 38(2):303–341, 2024.

Ruder, S., Peters, M. E., Swayamdipta, S., and Wolf, T. Transfer learning in natural language processing. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Tutorials, pages 15–18, Minneapolis, Minnesota, jun 2019. Association for Computational Linguistics. doi: 10.18653/v1/N19-5004. URL https://aclanthology.org/N19-5004.

Sarno, L. and Taylor, M. P. Official Intervention in the Foreign Exchange Market: Is It Effective, and, If So, How Does It Work? CEPR Discussion Papers 2690, C.E.P.R. Discussion Papers, February 2001. URL https://ideas.repec.org/p/cpr/ceprdp/2690.html.

Schmitt-Grohe, S. and Uribe, M. Closing Small Open Economy Models. NBER Working Papers 9270, National Bureau of Economic Research, Inc, October 2002. URL https://ideas.repec.org/p/nbr/nberwo/9270.html.

Shapiro, A. H. and Wilson, D. J. Taking the Fed at its Word: A New Approach to Estimating Central Bank Objectives Using Text Analysis. Working Paper Series 2019-2, Federal Reserve Bank of San Francisco, Jan 2019. URL https://ideas.repec.org/p/fip/fedfwp/2019-02.html.

Shiller, R. J. Narrative economics. Working Paper 23075, National Bureau of Economic Research, January 2017. URL http://www.nber.org/papers/w23075.

Shiller, R. J. Narrative economics: How stories go viral and drive major economic events. Princeton University Press, 2020.

Slapin, J. B. and Proksch, S.-O. A scaling model for estimating time-series party positions from texts. American Journal of Political Science, 52(3):705–722, 2008. doi: 10.1111/j.1540-5907.2008.00338.x.

Smets, F. and Wouters, R. An estimated dynamic stochastic general equilibrium model of the euro area. Journal of the European Economic Association, 1(5):1123–1175, 09 2003. ISSN 1542-4766. URL https://doi.org/10.1162/154247603770383415.

Straka, M. UDPipe 2.0 prototype at CoNLL 2018 UD shared task. In Zeman, D. and Hajič, J., editors, Proceedings of the CoNLL 2018 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies, pages 197–207, Brussels, Belgium, October 2018. Association for Computational Linguistics.

Straka, M., Hajič, J., and Straková, J. UDPipe: Trainable pipeline for processing CoNLL-U files performing tokenization, morphological analysis, POS tagging and parsing. In Calzolari, N., Choukri, K., Declerck, T., Goggi, S., Grobelnik, M., Maegaard, B., Mariani, J., Mazo, H., Moreno, A., Odijk, J., and Piperidis, S., editors, Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16), pages 4290–4297, Portorož, Slovenia, May 2016. European Language Resources Association (ELRA).

Szarvas, G., Vincze, V., Farkas, R., Móra, G., and Gurevych, I. Cross-genre and cross-domain detection of semantic uncertainty. Computational Linguistics, 38(2):335–367, 2012. ISSN 0891-2017.

Takayanagi, T., Suzuki, M., Kobayashi, R., Sakaji, H., and Izumi, K. Is chatgpt the future of causal text mining? a comprehensive evaluation and analysis, 2024. URL https://arxiv.org/abs/2402.14484.

Titt, R. Investigating the attitude towards ambiguity: Interindividual differences in automatic activations of evaluations of ambiguity. Doctoral thesis, Eberhard Karls Universität Tübingen, 2021. URL http://hdl.handle.net/10900/119290.

Tobback, E., Naudts, H., Daelemans, W., Junqué de Fortuny, E., and Martens, D. Belgian economic policy uncertainty index: Improvement through text mining. International Journal of Forecasting, 34(2):355–365, 2018. ISSN 01692070.

Traag, V. A., Waltman, L., and van Eck, N. J. From louvain to leiden: guaranteeing well-connected communities. Scientific Reports, 9(1), 3 2019. ISSN 2045-2322.

Turenne, N. Clustering and relational ambiguity: from text data to natural data. J. Data Min. Digit. Humanit., 2014, 2013. URL https://api.semanticscholar.org/CorpusID:15143369.

Vafa, K., Naidu, S., and Blei, D. Text-based ideal points. Association for Computational Linguistics, 2020:5345–5357, 2020. doi: 10.18653/v1/2020.acl-main.475.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., and Polosukhin, I. Attention is all you need. In Proceedings of the 31st International Conference on Neural Information Processing Systems, NIPS'17, page 6000–6010, Red Hook, NY, USA, 2017. Curran Associates Inc. ISBN 9781510860964. URL https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf.

Vikner, S. Why German is not an SVO-language but an SOV-language with V2. AU Library Scholarly Publishing Services, 2019. doi: 10.7146/aul.322.218.

Vissing-Jorgensen, A. Central banking with many voices: The communications arms race, 2019. URL https://www.bcentral.cl/es/web/banco-central/contenido/-/detalle/xxiii-annual-conference-of-the-central-bank-of-chile-6/-/detalle/xxiii-annual-conference-of-the-central-bank-of-chile-6.

Warner, B., Chaffin, A., Clavié, B., Weller, O., Hallström, O., Taghadouini, S., Gallagher, A., Biswas, R., Ladhak, F., Aarsen, T., Cooper, N., Adams, G., Howard, J., and Poli, I. Smarter, better, faster, longer: A modern bidirectional encoder for fast, memory efficient, and long context finetuning and inference, 2024. URL https://arxiv.org/abs/2412.13663.

Whelan, K. Central banks communications and monetary policy. Report, European Parliament Committee, 2018. URL www.europarl.europa.eu/cmsdata/153224/WHELAN_Final.pdf.

Wilson, A. T. and Chew, P. A. Term weighting schemes for latent dirichlet allocation. In Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics, HLT '10, page 465–473, USA, 2010. Association for Computational Linguistics. ISBN 1932432655.

Woodford, M. Central bank communication and policy effectiveness. Working Paper 11898, National Bureau of Economic Research, December 2005. URL http://www.nber.org/papers/w11898.

Zarifhonarvar, A. Generating inflation expectations with large language models. Journal of Monetary Economics, 157:103859, 2026. ISSN 0304-3932. doi: https://doi.org/10.1016/j.jmoneco.2025.103859. URL https://www.sciencedirect.com/science/article/pii/S0304393225001308.

Zhao, T., Liu, H., Roeder, K., Lafferty, J., and Wasserman, L. The huge package for high-dimensional undirected graph estimation in r. J. Mach. Learn. Res., 13(null):1059–1062, April 2012. ISSN 1532-4435.

Appendix B

Erklärung

Hiermit erkläre ich, Sami Diaf, dass ich keine kommerzielle Promotionsberatung in Anspruch genommen habe. Die Arbeit wurde nicht schon einmal in einem früheren Promotionsverfahren angenommen oder als ungenügend beurteilt.

Hamburg, den 20.01.2026

Sami Diaf

Eidesstattliche Erklärung

Ich, Sami Diaf, versichere an Eidesstatt, dass ich die Dissertation mit dem Titel: “Essays on Central Bank Communication” selbst und bei drei Zusammenarbeiten mit anderen Wissenschaftlern gemäß den beigefügten Darlegungen nach §6 Abs. 3 der Promotionsordnung der Fakultät für Wirtschafts und Sozialwissenschaften vom 18. Januar 2017 verfasst habe. Andere als die angegebenen Hilfsmittel habe ich nicht benutzt: LaTeX, R, Python, Google Colab, Octave, Dynare, DeepL und die im Quellenverzeichnis angegebene Literatur.

Hamburg, den 20.01.2026

Sami Diaf

Selbstdeklaration

Kapitel 1, 2, 4, 6 sowie 7 und 8 sind in Alleinautorenschaft erstanden.

Abschnitt 3.1 (”Uncovering uncertainty in narrative economics: a semantic search approach”) wurde mit Dr. Florian Schütze geschrieben. Die folgende Schätzung als Prozentsatz von Sami Diafs eigenem Werk wurde zwischen den mit dem Artikel verbundenen Autor vereinbart.

- Konzeption: 50%
- Durchführung: 50%
- Manuskripterstellung: 50%

Abschnitt 3.2 (”Multilingual Topic Modeling on the Bundesbank Speeches”) wurde mit Pr. Ulrich Fritsche geschrieben. Die folgende Schätzung als Prozentsatz von Sami Diafs eigenem Werk wurde zwischen den mit dem Artikel verbundenen Autor vereinbart.

- Konzeption: 50%
- Durchführung: 50%
- Manuskripterstellung: 50%

Kapitel 5 (”Investigating uncertainty in central bank corpora”) wurde mit Dr. Florian Schütze geschrieben. Die folgende Schätzung als Prozentsatz von Sami Diafs eigenem Werk wurde zwischen den mit dem Artikel verbundenen Autor vereinbart.

- Konzeption: 50%
- Durchführung: 50%
- Manuskripterstellung: 50%

Kapitel 8 (”Policy Mix in An Oil Exporting Country: Effectiveness of Countercyclical Measures in Mitigating External Shocks.”) wurde aus einer früheren Zusammenarbeit mit Pr. Ahmed Zakane (ENSSEA und CREAD, Algerien) geschrieben, aber mit einem neuen Textverfahren. Die folgende Schätzung als Prozentsatz von Sami Diafs eigenem Werk wurde zwischen den mit dem Artikel verbundenen Autor vereinbart.

- Konzeption: 70%
- Durchführung: 70%
- Manuskripterstellung: 70%

Hamburg, den 20.01.2026

Sami Diaf

List of publications

Sami Diaf. Policy preference at central banks: Quantifying monetary policy signals using keyword topic models. WiSo-HH working paper series 69, 2022. doi: https://d-nb.info/1273646363/34.

Sami Diaf and Ulrich Fritsche. TopicShoal: Scaling partisanship using semantic search. In Proceedings of the 18th Conference on Natural Language Processing (KONVENS 2022), pages 167–174, Potsdam, Germany, 12–15 sep 2022. KONVENS 2022 Organizers. URL https://aclanthology.org/2022.konvens-1.20.

Sami Diaf and Florian Schütze. Estimating policy uncertainty within monetary policy debates. 5^{th} International Conference on Advanced Research Methods and Analytics (CARMA2023), Sevilla, Spain. URL https://riunet.upv.es/server/api/core/bitstreams/c31d8d25-a7a0-42e1-b387-7e1217e085f2/content

Sami Diaf. CommunityFish: A Poisson-based document scaling with hierarchical clustering. In Mourad Abbas and Abed Alhakim Freihat, editors, Proceedings of the 6th International Conference on Natural Language and Speech Processing (ICNLSP 2023), pages 59–67, Online, December 2023a. Association for Computational Linguistics. URL https://aclanthology.org/2023.icnlsp-1.6/

Sami Diaf. Ambiguity in Central Banking Communication: A Strategic Vagueness Or A Pure Randomness ?. 17^{th} International Conference on Statistical Analysis of Textual Data (JADT 2024), Brussels, Belgium. https://jadt-2024.sciencesconf.org/data/program/Ambiguity_in_Central_Banking_Communication.pdf

Sami Diaf. CliqueCorex: A self-supervised clique-based anchored topic model. In Mourad Abbas and Abed Alhakim Freihat, editors, Proceedings of the 7^{th} International Conference on Natural Language and Speech Processing (ICNLSP 2024), pages 74–82, Trento, October 2024. Association for Computational Linguistics. URL https://aclanthology.org/2025.icnlsp-1.12.pdf

Sami Diaf and Florian Schütze. Multilingual Monetary Policy: Unfolding Language and Policy Preferences of Swiss Central Bankers. 6^{th} International Conference on Advanced Research Methods and Analytics (CARMA2024), Valencia, Spain. URL https://doi.org/10.4995/CARMA2024.2024.17405

Sami Diaf and Florian Schütze. Uncovering uncertainty in narrative economics: A semantic search approach. In New Frontiers in Textual Data Analysis. Springer Nature Switzerland, 2024. URL https://www.springerprofessional.de/uncovering-uncertainty-in-narrative-economics-a-semantic-search-/50037724

Sami Diaf and Ahmed Zakane. Policy Mix in An Oil Exporting Country: Effectiveness of Countercyclical Measures in Mitigating External Shocks. In Dynar Working Papers, CEPREMAP. URL https://www.dynare.org/wp-repo/dynarewp083.pdf

Sami Diaf. LLMs Information Flow Diagnostic: Memory-based Evidence from Random Matrix Theory. In Proceedings of the 8^{th} International Conference on Natural Language and Speech Processing (ICNLSP-2025), pages 103–110, Southern Denmark University, Odense, Denmark. Association for Computational Linguistics. URL https://aclanthology.org/2022.konvens-1.20

Sami Diaf. CausaMap: A Semi-supervised Map For Causal Text Mining. 16^{th} IEEE International Conference on Knowledge Graphs ICKG 2025, November 13-14, 2025 Limassol, Cyprus. URL https://ieeexplore.ieee.org/document/11391993/.