Universidade de Brasília

Faculdade de Economia, Administração e Contabilidade

Departamento de Economia

Natural Language Processing in Economics and Finance: Literature Review and Applications to Monetary Policy Communication

Alexandre Henrique Lucchetti

TESE DE DOUTORADO

PROGRAMA DE PÓS-GRADUAÇÃO EM ECONOMIA

Brasília

2025

Universidade de Brasília
Faculdade de Economia, Administração e Contabilidade
Departamento de Economia

Processamento de Linguagem Natural em Economia e Finanças: Revisão de Literatura e Aplicações para Comunicação de Política Monetária

Alexandre Henrique Lucchetti

Tese de Doutorado submetida ao Programa de Pós-Graduação do Departamento de Economia da Universidade de Brasília como requisito à obtenção do título de Doutor em Economia.

Orientador: Prof. Dr. Daniel Oliveira Cajueiro

Brasília
2025

FICHA CATALOGRÁFICA

Lucchetti, Alexandre Henrique.

Natural Language Processing in Economics and Finance:Literature Review and Applications to Monetary Policy Communication / Alexandre Henrique Lucchetti; orientador Daniel Oliveira Cajueiro. -- Brasília, 2025.

164 p.

Tese de Doutorado (Programa de Pós-Graduação em Economia) -- Universidade de Brasília, 2025.

1. Natural Language Processing. 2. Monetary Policy Communication. 3. Monetary Policy Coordination. 4. Yield Curve. 5. Sentiment Analysis. I. Cajueiro, Daniel Oliveira, orient. II. Título.

Universidade de Brasília
Faculdade de Economia, Administração e Contabilidade
Departamento de Economia

Natural Language Processing in Economics and Finance:Literature Review and Applications to Monetary Policy Communication

Alexandre Henrique Lucchetti

Tese de Doutorado submetida ao Programa de Pós-Graduação do Departamento de Economia da Universidade de Brasília como requisito à obtenção do título de Doutor em Economia.

Trabalho aprovado. Brasília, 07 de março de 2025:

Prof. Dr. Daniel Oliveira Cajueiro,
UnB/FACE/ECO
Orientador

Prof. Dr. Roberto de Góes Ellery Júnior,
UnB/FACE/ECO
Examinador interno

Prof. Dr. José Angelo Divino,
Universidade Católica de Brasília
Examinador externo

Dr. Carlos Alexandre Piccioni,
Banco Central do Brasil
Examinador externo

Abstract

This work consists of three papers on Natural Language Processing (NLP) applied to economics and finance. The first paper surveys the main methods, requisites and applications of NLP in the field. Structured to help shorten the path economic researchers have to trail to get introduced to these methods, our work covers the following topics: (i) useful NLP tasks to economics; (ii) NLP models; (iii) economic and financial textual data for NLP; and (iv) NLP applications in economics and finance. We further contribute by resorting to bibliometric tools to help us visualize the literature map of this field, also providing valuable insights to our survey and the remainder of our work. We finally indicate that there is much room to apply natural language processing to economic issues, but alert that, more than ever, researchers must be careful not to stray away from questions motivated by hypotheses closely tied to economic theories. The second paper provides a forward-looking measure of how central bankers implicitly coordinate their actions, as measured from public manifestations in the form of their speeches' transcripts. In order to do that, we resort to the the central bankers' speeches database made available by the Bank for International Settlements and build a network of similarities that connects central banks, adapting for this context the method proposed by Cajueiro *et al.* (2021). Our results show that our network successfully captures the long-term global importance of central banks overseeing the G10 currencies, with word-level point to evidence of their orthodox approach to monetary policy. We further explore this framework on a dynamic setting, with findings that indicate that coordination tends to increase in times of economic stress, as in the years of the Great Financial Crisis and the period after the Covid-19 pandemic. An evolution analysis on word occurrences then shows that our proposed measure is driven by mentions to policy instruments and economic views proffered by policymakers. The third paper proposes a framework for estimating expectation-embedded multi-dimensional sentiment from monetary policy communication, combining economic fundamentals and state-of-the-art deep learning neural networks. The economic basis of our approach is set by the Litterman and Scheinkman (1991) yield curve decomposition model, with its level, slope and curvature factors accounting for the three dimensions of our sentiment gauge. For text modeling we incorporate in our framework the Bidirectional Encoder Representations from Transformers, BERT (Devlin *et al.*, 2019). In an application to policy communication by the Brazilian Central Bank, our results reinforce the need for more than one sentiment dimension to comprehensively capture the relevant nuances of communication for monetary policy assessment. These results indicate that the added dimensions complement the usual hawk/dove gauge of policy sentiment, and can at times be more relevant in setting the overall tone of policy communication, eventually even preceding shifts in monetary policy stance.

Keywords: Natural Language Processing; Monetary Policy Communication; Monetary Policy Coordination; Yield Curve; Sentiment Analysis.

Resumo

Este trabalho consiste em três artigos sobre Processamento de Linguagem Natural (PLN) aplicada a economia e finanças. O primeiro artigo faz uma revisão sobre os principais métodos, requisitos e aplicações de PLN neste campo de pesquisa. Estruturado de forma a ajudar a encurtar o caminho a ser trilhado pelo pesquisador econômico para ser introduzido a estes métodos, nosso trabalho cobre os seguintes tópicos: (i) tarefas de PLN úteis em economia; (ii) modelos de PLN; (iii) dados textuais em economia e finanças para PLN; e (iv) aplicações de PLN em economia e finanças. Nós ainda contribuímos com o emprego de ferramental de bibliometria para ajudar a visualizar o mapa bibliográfico deste campo de pesquisa, também provendo insights valiosos para a revisão desta literatura e para o restante de nosso trabalho. Por fim, nós indicamos que ainda existe bastante espaço para aplicação de processamento de linguagem natural para problemas econômicos, mas alertamos que, mais que nunca, pesquisadores devem tomar cuidado para não desviar de questões motivadas por hipóteses relacionadas à teoria econômica. O segundo artigo propõe uma medida prospectiva de como banqueiros centrais implicitamente coordenam suas ações, estimada a partir de manifestações públicas na forma de seus discursos transcritos. Para tanto, nós utilizamos a base de dados de discursos de banqueiros centrais disponibilizada pelo BIS (Bank for International Settlements) e construímos uma rede de similaridades que conecta bancos centrais, adaptando para este contexto o método proposto por Cajueiro *et al.* (2021). Nossos resultados mostram esta rede sendo bem-sucedida em capturar a relevância global de longo prazo das instituições responsáveis pelas moedas constituintes do G10, com análise em nível de palavras apontando para evidência de sua abordagem ortodoxa na condução de política monetária. Nós ainda exploramos a capacidade deste arcabouço em uma configuração dinâmica, com resultados indicando que a coordenação tende a aumentar em tempos de estresse econômico, como nos anos da Grande Crise Financeira e no período após a pandemia de Covid-19. Uma análise sobre a evolução do uso de palavras específicas nos discursos mostra que nossa medida de coordenação é influenciada por menções a instrumentos de política monetária e à visão de cenário econômico dos formuladores de política. O terceiro artigo propõe um arcabouço para estimar um índice de sentimento multidimensional considerando expectativas para comunicação de política monetária, combinando fundamentos econômicos e redes neurais profundas de estado da arte. O embasamento econômico da nossa abordagem parte do modelo de decomposição de curva de juros de Litterman and Scheinkman (1991), com seus fatores de nível, inclinação e curvatura representando as três dimensões de nossa medida de sentimento. Para modelagem de texto, nós incorporamos em nosso arcabouço o modelo Bidirectional Encoder Representations from Transformers, BERT (Devlin *et al.*, 2019). Em aplicação à comunicação oficial do Banco Central do Brasil, nossos resultados

reforçam a necessidade de usar mais de uma dimensão de sentimento para capturar as nuances de comunicação relevantes para análise completa sobre a política monetária. Estes resultados indicam que as dimensões adicionais complementam a medida hawk/dove mais comum para medir sentimento nesta área, e por vezes podem ser mais relevantes para definição do tom geral da comunicação, eventualmente até funcionando como indicador antecedente para mudanças na postura da autoridade monetária.

Palavras-chave: Processamento de Linguagem Natural; Comunicação de Política Monetária; Coordenação de Política Monetária; Curva de Juros; Análise de Sentimento.

List of figures

List of tables

Contents

1 Introduction

This work consists of three papers on Natural Language Processing (NLP) applied to economics and finance. The first paper surveys the main methods, requisites and applications of NLP in economics and finance, guided by the following research questions: what is NLP currently capable of and how it can contribute to economic research? What are the methods and techniques to achieve such contributions? What kind of data is needed and suitable for these methods' implementation? And what can we expect to achieve with NLP in economics and finance? These four questions outline the structure of the paper, designed to help shorten the path economic researchers have to trail to get introduced to this field, with topics that cover: (i) useful NLP tasks to economics; (ii) NLP models; (iii) economic and financial textual data for NLP; and (iv) NLP applications in economics and finance. We further contribute by resorting to bibliometric tools to help us visualize the literature map of this field, also providing valuable insights to our survey and the remainder of our work. We finally indicate that there is much room to apply natural language processing to economic issues, but alert that, more than ever, researchers must be careful not to stray away from questions motivated by hypotheses closely tied to economic theories.

With the insights from the first paper, the second and third papers propose novel applications of the discussed techniques to monetary policy communication. The second paper provides a forward-looking measure of how central banks implicitly coordinate their actions, accounting for similarities on what policymakers weigh on the economic outlook, the instruments they rely on and the forward-guidance they communicate, all extracted from public manifestations in the form of speeches' transcripts. In order to do that, we resort to the central bankers' speeches database made available by the Bank for International Settlements and build a network of similarities that connects central banks whose policymakers present speech similarity, adapting for this context the method proposed by Cajueiro *et al.* (2021). This approach thus combines natural language processing with complex network analysis techniques. Our results show that our network successfully captures the long-term global importance of central banks overseeing the G10 currencies. Additionally, word-level analysis over the speeches of policymakers in this group of central banks point to evidence of their orthodox approach to monetary policy. We further explore this framework on a dynamic setting, so we can assess policy coordination evolution. Our findings indicate that coordination tends to increase in times of economic stress, as in the years of the Great Financial Crisis and the period after the Covid-19 pandemic. Additionally, our novel coordination measure is compared to two gauges of coordination extracted from conventional monetary policy instruments, namely policy rates and central bank balance sheet. This analysis shows that coordination as measured by speech similarity is able to capture increases in both metrics,

standing as a more general way to assess whether the global liquidity environment is subject to coordinated policy movements. Lastly, we run an evolution analysis on word occurrences to understand how they can shed light on coordination drivers through time. We find that our proposed measure is driven by mentions to policy instruments and economic views proffered by policymakers.

Finally, the third paper proposes a novel framework for estimating expectation-embedded multi-dimensional sentiment from monetary policy communication, combining economic fundamentals and state-of-the-art deep learning neural networks. In order to do that, we first resort to the Litterman and Scheinkman (1991) yield curve decomposition model — with its level, slope and curvature factors accounting for the three dimensions of our approach — to estimate market participants' expectation shifts as priced along the entire term structure of interest rates in response to monetary policy communication. From there, we propose a two-stage estimation framework to make sure we incorporate all the information regarding monetary policy communication known by economic agents. We do that by first estimating the communication state as known by agents from the Brazilian monetary policy decision statement, released on the same day as its rate decision. The second stage combines the communication estimated from the first stage with new communication from monetary meeting minutes, released six days after the meeting, to infer monetary policy sentiment from how market participants price the new information in the yield curve's factors. For text modeling we incorporate in our framework the Bidirectional Encoder Representations from Transformers, BERT (Devlin *et al.*, 2019). BERT is based on the groundbreaking transformer architecture (Vaswani *et al.*, 2017), which currently provides state-of-the-art results to natural language processing problems. In an application of our framework to the most recent policy communication by the Brazilian Central Bank, our results reinforce the need for more than one sentiment dimension to comprehensively capture the relevant nuances of communication for monetary policy assessment. We show that the first dimension of our sentiment index relates to the hawk/dove gauge usually adopted by conventional monetary policy sentiment analysis. This suggests that these approaches are a special case of our framework. Furthermore, we show that second and third dimensions of our sentiment index can at times be more relevant in setting the overall tone of policy communication, eventually even preceding shifts in monetary policy stance.

2 Language as Data: A Survey of Natural Language Processing for Economics and Finance

2.1 Introduction

Language provides humans with a way to network their brains together (Manning, 2022). It allows not only improved interactions among individuals, but also information sharing, both leveraging rational capabilities to construct and enhance sophisticated concepts and ideas. As a matter of fact, language representations, mostly in the form of unstructured textual data, stand as a very rich yet challenging data source to be explored by almost every branch of social sciences research. In this context, natural language processing (NLP) emerges as a multidisciplinary research field, combining mostly computer science, artificial intelligence and linguistics to extract, quantify and even generate information from human natural languages. This field has been through impressive breakthroughs recently with the advent of transformer neural networks (Vaswani *et al.*, 2017), upscaling its methods' ability to understand underlying concepts from textual data. For social sciences, this means going from text as data (Gentzkow; Kelly; Taddy, 2019), using syntactic and semantic patterns as a means to extract features, to language as data, using quantified ideas from textual language representations.

Our primary contribution lies in establishing a clear connection between NLP (Natural Language Processing) concepts and their application in economics. This clarification highlights the specific aspects within this rapidly evolving research area that should be thoroughly understood by economic and financial researchers to fully capitalize on its potential. Furthermore, we enhance the journey by employing bibliometric tools to visualize the literature landscape in this field. This not only aids in comprehending the research landscape but also enriches our survey with valuable insights.

With that in mind, our paper surveys the main methods, requisites and applications of NLP in economics and finance, guided by the following research questions: what is NLP currently capable of and how it can contribute to the economic research? What are the methods and techniques to achieve such contributions? What kind of data is needed and suitable for these methods' implementation? And what can we expect to achieve with NLP in economics and finance? These four questions outline the structure of the paper, respectively addressed by the topics of: (i) useful NLP tasks to economics; (ii) NLP models; (iii) economic and financial textual data for NLP; and (iv) NLP applications in economics and finance.

Our work aligns with the research of Gentzkow, Kelly, and Taddy (2019) and, more recently, Ash and Hansen (2023), aiming to introduce NLP techniques to economic and financial researchers who may not be familiar with them. Gentzkow, Kelly, and Taddy (2019) work is a well-recognized introduction to the use of textual data in economic research. While it has been widely referenced, it is essential to note that it dates back to a time when significant breakthroughs in natural language processing methodologies had not yet occurred. Nevertheless, it served as a pioneering effort in introducing NLP concepts and methods, offering practical guidance to economic researchers looking to leverage the wealth of in-

formation available in textual data. On the other hand, Ash and Hansen (2023) research provides a comprehensive overview of the most commonly employed textual algorithms in contemporary economics. Additionally, it explores four measurement problems within economics, specifically addressing issues related to measuring document similarity, quantifying economic concepts within raw text, assessing the interrelation of concepts in text, and establishing connections between text and quantitative metadata.

Our work also connects to the significant research conducted by Algaba *et al.* (2020), which focuses on the field of sentiment analysis. In our work, we recognize sentiment analysis as a crucial area at the intersection of economics and natural language processing. Algaba *et al.* (2020)'s work offers a comprehensive overview of both methodological and applied approaches within this field.

We organize our paper as follows: Section 2.2 brings a broad overview of the literature by applying bibliometric tools to map what have been done so far, providing input for the remainder of our work; Section 2.3 introduces NLP tasks, discussing how sentiment analysis and topic modeling, among others, can be applied to economic research; Section 2.4 proceeds to describe the main NLP models, tracking their evolution over time; Section 2.5 focuses on textual economic data, going from the most usual data sources to some practical advice on how to work with such data; Section 2.6 brings the applications themselves, discussing how NLP has been used by the economic literature; and finally, Section 2.7 concludes our work and outlines frontier research topics.

2.2 Literature Mapping

We begin by resorting to bibliometric tools to understand how NLP techniques have been applied in economics and finance and how the field has evolved through the last few years, while trying to draw a picture of what to expect from emerging research and their possibilities. We write the next subsections based on Zupic and Čater (2015) and Donthu *et al.* (2021)'s recommendations to perform bibliometric analyses. However, it is important to state that we do not mean to perform a complete bibliometric analysis here, since we are not interested in assessing research performance, for instance. We direct our efforts towards presenting the mapping of topics and visualizing their relations. Thus, according to the classification of tools and techniques in Donthu *et al.* (2021), we engage in a *science mapping* assessment, supported by network analysis methods and visualization.

2.2.1 Scope and data

For this analysis, we use a dataset consisting of 2,501 scientific publications sourced from the Web of Science (WOS) Core Collection. The dataset spans a period from 1992, with the earliest paper in our collection dated that year, extending through to 2023. This database,

besides being one of the most commonly used for bibliometric studies (Zupic; Čater, 2015; Donthu *et al.*, 2021), was also widely used in recent finance literature (Khan *et al.*, 2022; Chen *et al.*, 2023).

In order to retrieve the bibliometric data, we build the search query inspired in Khan *et al.* (2022) and Chen *et al.* (2023), with relevant keywords trying to capture the intersection of NLP and economics and finance literature. For that end, we used a set of keywords for each of these two fields, as shown in Table 2.1, and queried the intersection (by applying an AND clause between the two sets). In complement to that search and still following the surveyed literature, and in order to guarantee that the resulting universe is relevant to our analysis, we turned to the Web of Science Category filter to keep only topics akin to our focused fields. However, since we are working in an interdisciplinary area, namely computer science and economics and finance, we noted that simply querying for the intersection of the two categories resulted in neglecting an important part of the literature we needed for a comprehensive assessment. In a similar fashion, querying the union would result in including research from fields strange to our research goal, just as would including only one of our target fields. Our solution was then to perform a negative screening to rule out every category that appeared in the initial search and should not be a part of our analysis universe[1]. It is worth mentioning that, as recommended in the bibliometric literature guidelines (Zupic; Čater, 2015; Donthu *et al.*, 2021), our query building process comprises many iterations of criteria adjustment and result assessment in order to get to the most reliable universe of research to our study.

Criteria	Keywords
Field: Economics / Finance	financ* OR econom* OR “monetary polic*” OR “fiscal polic*” OR investment* OR “asset pricing” OR stock* OR “fixed income” OR commodit* OR forex OR bitcoin OR cryptocurrenc* OR “liquidity risk” OR “credit risk” OR “market risk” OR insurance*
Field: NLP	nlp OR “natural language processing” OR “sentiment analysis” OR “text classification” OR “text summarization” OR “topic modeling” OR “bag of words” OR “n-gram” OR ngram OR “tf-idf” OR tfidf OR “term frequency - inverse document frequency” OR “latent semantic analysis” OR “latent dirichlet allocation” OR “word embedding” OR “sentence embedding” OR word2vec OR doc2vec OR “large language model”

Table 2.1 – WOS Query Keywords

[1] As an example, we included ‘MEDICAL INFORMATICS’ and ‘ENGINEERING CHEMICAL’ in the negative screening to rule out every work related to these categories. The assumption that was confirmed in the results is that categories like these would have little to no intersection with our target field.

2.2.2 Tools and methodology

As for the bibliometric tool used in our analysis, VOSviewer (van Eck; Waltman, 2010) is our software of choice, following what has previously been applied to the finance literature in Khan *et al.* (2022). The rationale behind this choice is the software's specialization in network visualization, a feature that suits perfectly our science mapping objective for this section. It does so by applying the *visualization of similarities* (VOS) mapping technique[2].

In order to visualize the topics covered in the targeted literature, as well as how they relate to each other, we apply a co-word analysis (Donthu *et al.*, 2021) to the titles and the abstracts of the 2,501 scoped papers. In this kind of approach, the terms — or words — mentioned are the unit of analysis, defining the nodes of the network. Thus, the size of each node shows how frequent the term is in the literature. The frequency of their co-occurrence in documents defines the links between them. We used a binary counting approach, so we do not get misleading signals from multiple occurrences of some terms in a single paper[3]. Furthermore, we only consider terms with more than 20 occurrences in the dataset, so we focus our analysis in the most relevant nodes in the network, and keep the 60% most relevant terms according to the VOSviewer *relevance score*[4]. This last step ensures that recurrent terms with no actual meaning to the analysis — such as "paper" or "interesting result" — are left out so they do not obfuscate relevant terms in the network.

2.2.3 Literature Network and Findings

We now get to the network representation for the literature of NLP applied to economics and finance, depicted in Figure 2.1. In this network, we can see three clusters clearly outlined[5]. We will refer to them as the *NLP cluster*, represented by the green nodes, the *Economy Cluster*, represented by the red nodes, and the *Finance Cluster*, in the blue nodes. It is interesting diving deeper into each one of them, as the three show some relevant aspect of the literature.

2.2.3.1 Network Clusters

Let us first turn to the *NLP Custer*, in green. It is rather straightforward to understand our labelling choice, since this cluster contains most of the NLP models and techniques commonly applied to our fields of interest (there is one exception, though, as we will see when we get to where topic modeling stands, further ahead). Here, we can find a wide range of models, from the basic *lexicon* (Guthrie *et al.*, 1996) approaches to the groundbreaking *transformer* network (Vaswani *et al.*, 2017). In between, we note terms referring to NLP

[2] For more on the VOS mapping technique, refer to Eck and Waltman (2007).
[3] Binary counting only considers the presence or absence of a term in each research paper.
[4] For more on VOSviewer's relevance score, refer to van Eck and Waltman (2011)
[5] Cluster representation is also based on the VOS mapping technique (Eck; Waltman, 2007).

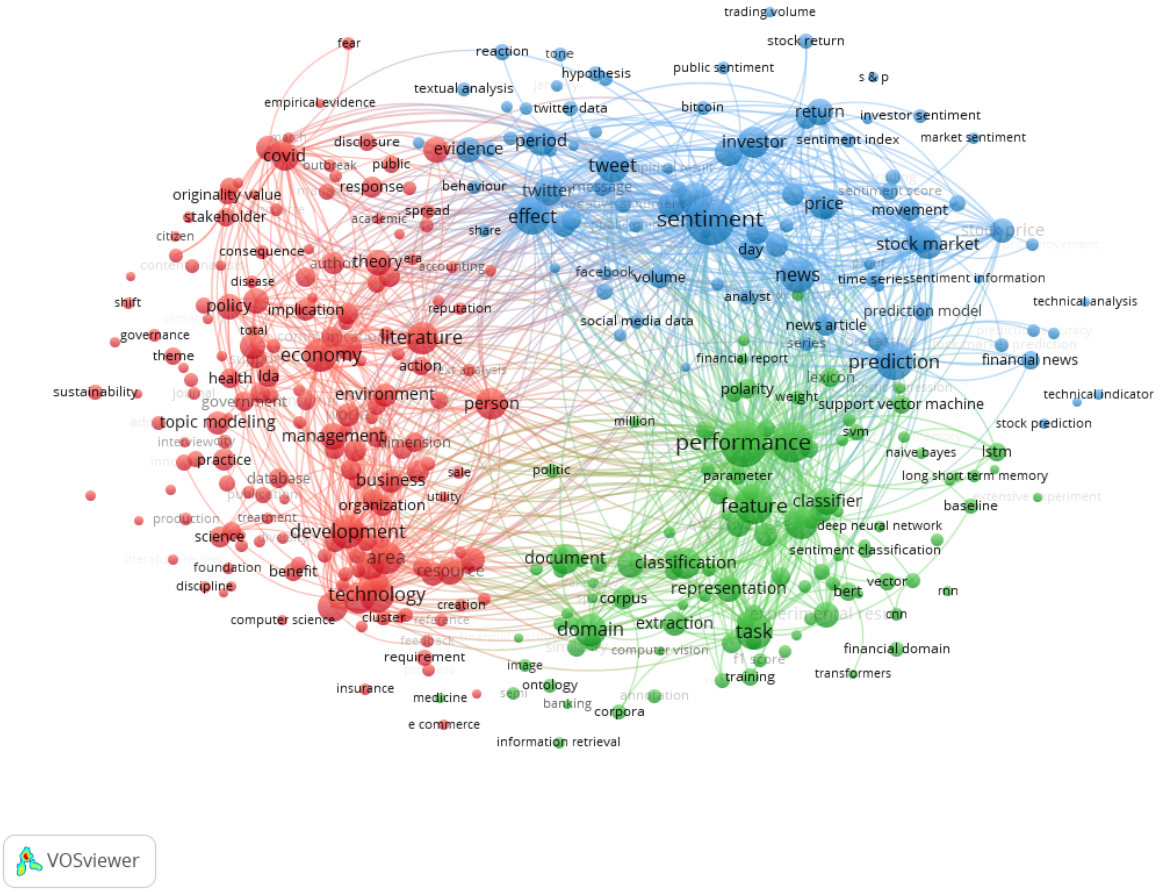


Figure 2.1 – NLP in Finance - Literature Network - Clusters

models and related machine learning techniques such as *term frequency inverse document frequency* (Salton; Buckley, 1988), *support vector machine* (Cortes; Vapnik, 1995) (and their respective abbreviations *tf idf* and *svm*), *logistic regression* (Berkson, 1944; Berkson, 1951), *naive bayes* (Duda; Hart *et al.*, 1973; Domingos; Pazzani, 1997) and *word embedding* (Mikolov *et al.*, 2013a). Furthermore, approaches related to *deep neural networks* (Goodfellow; Bengio; Courville, 2016) are also present. Some of these are *convolutional neural network* (Zhang *et al.*, 1988; LeCun *et al.*, 1989; Li *et al.*, 2021), *recurrent neural networks* (Jordan, 1986; Elman, 1990) and *long short term memory* networks (Hochreiter; Schmidhuber, 1997; Sutskever; Vinyals; Le, 2014), again all with the respective abbreviations: *cnn*, *rnn* and *lstm*. Finally, *transformers* are represented not only by its own node, but also by *attention mechanism* (Bahdanau; Cho; Bengio, 2014) and *bert* (Devlin *et al.*, 2019). Despite being possible to identify even more NLP model-related nodes, we should move on the other interesting insights from this cluster. We must keep in mind, however, that we come back to the NLP models' section of the network to assess the fast-paced developments of NLP in the recent years and how these methods

have been incorporated to the economic and financial literature.

Still on the green cluster, we can also identify language processing task-related nodes, backing one more time the rationale behind calling this the *NLP Cluster* (there are, once again, exceptions that will be discussed ahead, but witout prejudice to the argument presented here). Besides the trivial *task* node, one of the most connected nodes in this cluster, by the way, *classification*, for "document classification" (Sebastiani, 2002), and *extraction*, which in turn is strongly connected to *feature* — resulting in "feature extraction" (Shaker *et al.*, 2022) —, are good examples of task nodes in this cluster[6]. However, maybe the most insightful conclusion from this segment of the network comes from looking at the most central node in it: *performance* (King, 1996; Liang *et al.*, 2022). Besides being connected to almost every model node, it has the strongest connection of the *task* node. This result suggests that the studied branch of the literature focuses a fair amount of effort in assessing NLP model performance in economics and finance applications (Mishev *et al.*, 2020).

Moving on to the *Financial Cluster*, in blue, we begin our analysis by maybe the most meaningful result in our bibliometric exercise: the *sentiment* node. This node refers to "sentiment analysis", a popular NLP task that extracts document tone according to same predefined scale (more on this in Section 2.3) (Algaba *et al.*, 2020). This node is not only the most connected node in its own cluster, but the most connected node in the entire network. Moreover, this node is the largest one, indicating that this term is the most mentioned among the selected papers. All this suggests that this is the preferred task for NLP applications in the economics and finance literature (Zhao *et al.*, 2021; Petropoulos; Siakoulis, 2021; Shapiro; Sudhof; Wilson, 2022; Niţoi; Pochea; Radu, 2023).

However globally important in this bibliometric network, we must understand what makes this node the center of what we call the *Financial Cluster*. First, to convince the reader that this can in fact be labelled *Financial Cluster*, we point out to the most mentioned terms in it. Among them, we find *stock market*, *stock price*, *investor*, *price* and *return*, to name a few. Less mentioned terms, thus smaller nodes, also add up to this interpretation, with *financial news*, *s&p*, *bitcoin* and *profit* as examples. Therefore, turning back to the *sentiment* node and recalling that it is placed in the center of the cluster, we can infer that sentiment analysis is the most widespread NLP task to the specific field of finance as well (Daniel; Neves; Horta, 2017; Sohangir *et al.*, 2018; Mishev *et al.*, 2020).

In addition to it, continuing with our analysis in the *Financial Cluster*, as we still have plenty of insight to extract from it, we turn to the still-not-mentioned largest nodes here: textual data nodes. These are nodes such as *twitter*, *tweet* and *news*, backed by the somewhat smaller but yet relevant nodes *social media* and *facebook*. Here we can see which are the most usual textual data applied in NLP research in finance. Since social media and news

[6] We will cover NLP tasks in detail in Section 2.3.

data allow for higher frequency market monitoring when compared to traditional economic data, it is easy to see the value of this kind of data to the research field (Wang *et al.*, 2015; Souma; Vodenska; Aoyama, 2019; Agarwal, 2019).

And, lastly, to exhaust the *Financial Cluster*'s insights, we must also take a look at the frontier between this cluster and the *NLP Cluster*, mentioned before. There, we can see a group of nodes related to *prediction* and *forecasting*. Besides these two, *prediction model*, *prediction accuracy* and *stock market prediction* illustrate what we refer to as NLP applications, as these outline one possible purpose of using NLP in economics and finance. As a matter of fact, financial markets predictions are one of the most popular NLP applications in our field of interest (Jiang, 2021; Farimani; Jahan; Fard, 2022). This and other applications will be further discussed in Section 2.6.

Finally, we get to the *Economy Cluster*. Being the most heterogeneous in the network, the criteria for naming this cluster was to look to its most relevant node: *economy*. However not as straightforward as the previous two, this cluster's label is somewhat backed by the other central nodes as well, specially *management*, *business* and *policy*. Nevertheless, the most interesting fact about this one resides in another exception to the NLP tasks and models described in the first cluster: the *latent dirichlet allocation* node (Blei; Ng; Jordan, 2003). Alongside with its abbreviation, *lda*, this node represents the most popular NLP model outside the *NLP Cluster* in our literature network. Used to perform *topic modeling* analysis (Boukus; Rosenberg, 2006), which is represented by a neighbouring node, this model's unsupervised learning traits in fact make it adequate to be used in a heterogeneous range of subjects (Correia; Mueller, 2021; Roos; Reccius, 2023; Borup *et al.*, 2023). Thus, it could also be considered as a defining node for this cluster, albeit less central than the one chosen in our analysis.

2.2.3.2 Literature Evolution

Having the literature network fully characterized and its main insights discussed, we can move to some assessment on how this network seems to be shaping itself across time. Figure 2.2 shows the same co-word occurrence network for the NLP applications in economics and finance, although now with nodes coloured according to the average publication year of the papers in which each term appears.

As seen, the representation in Figure 2.2 is helpful to understand the recent trends in the literature, specially in order for us to assess the evolution of the topics and methods applied over time. In a broad analysis, we can see that most of the newest terms are located in the *Economy Cluster*, with a great proportion of them with relation to some extend to the *covid* and the *pandemic* nodes. It is as expected, since this is in fact a subject that emerged in the last few years (Borup *et al.*, 2023). Moreover, terms related to ESG aspects, such as *sustainability* and *governance* also mark the average most recent terms in this literature.

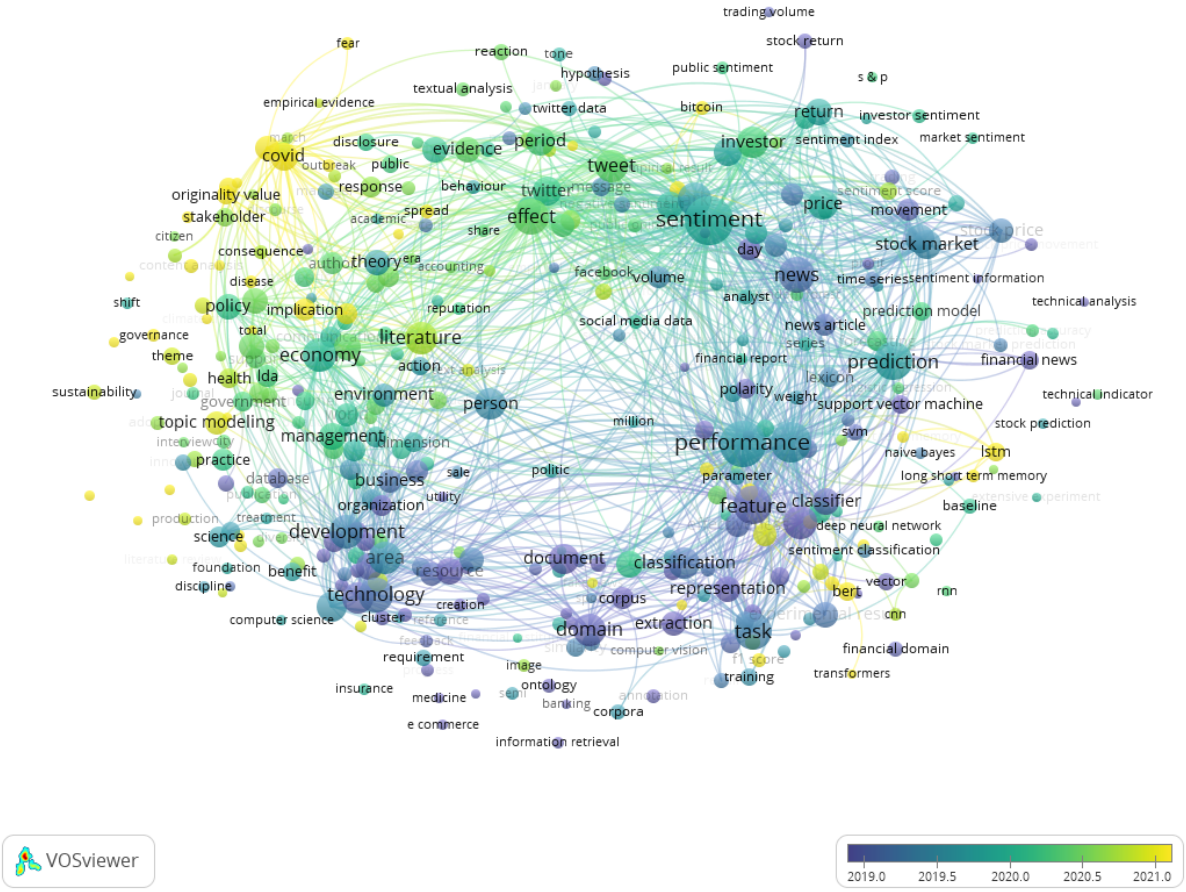
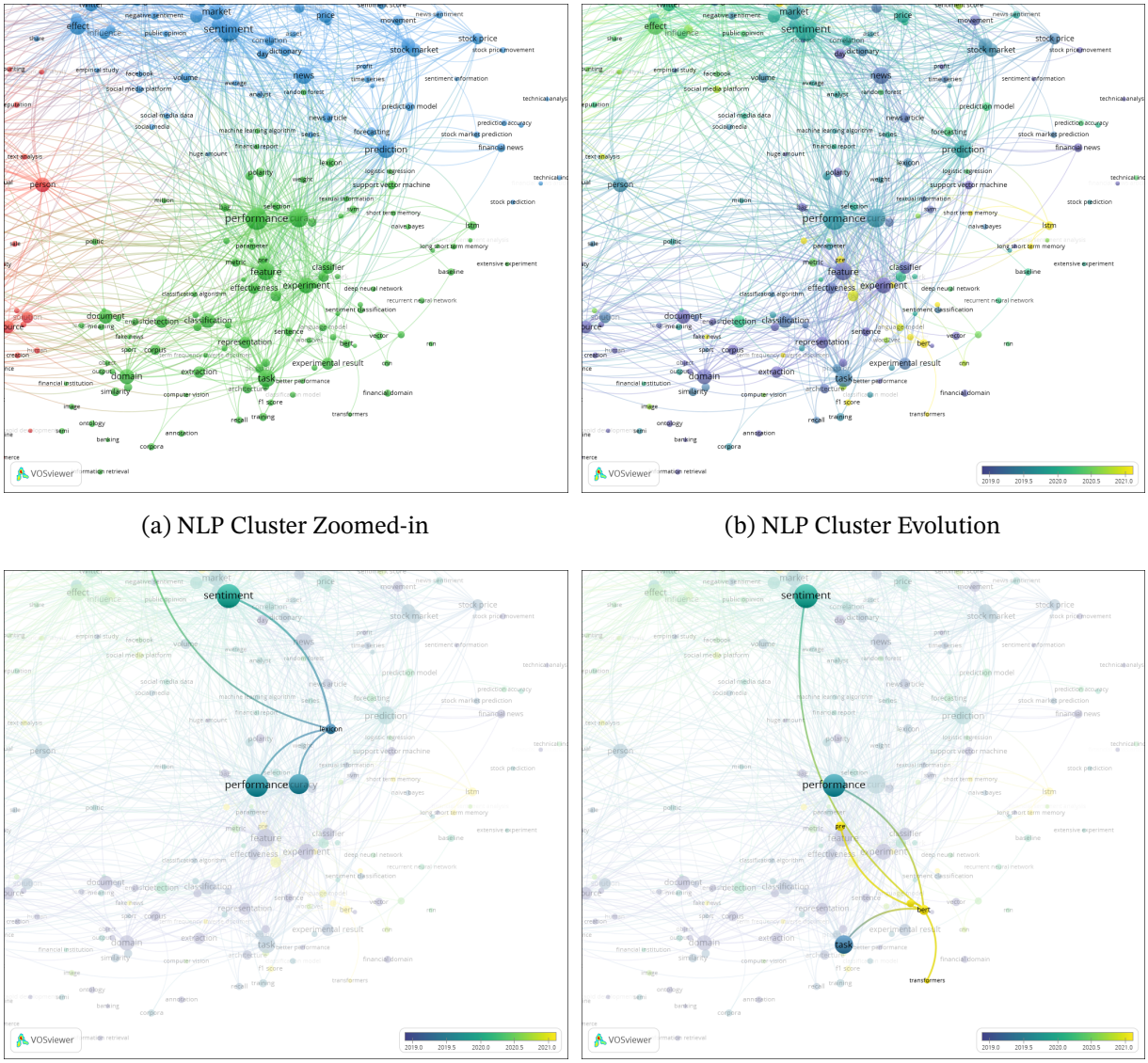


Figure 2.2 – NLP in Finance - Literature Network - Evolution

Furthermore, *bitcoin* and *cryptocurrency* represent the most recent mentions in average for the *Financial Cluster* (Passalis *et al.*, 2022). As for the *NLP Cluster*, it is interesting to make a zoomed-in analysis to see what we can infer for the models in light of the recent breakthroughs. Figure 2.3 shows that analysis.

In figure 2.3, we first see the zoomed-in *NLP Cluster* and its evolution, respectively, in panels 2.3a and 2.3b. More on panel 2.3b, we can see clearly that some bright yellow nodes, showing more recent terms to the literature, stand out. The two largest, thus referring to terms with more occurrences in the universe of papers in scope, are *lstm* (Hochreiter; Schmidhuber, 1997; Sutskever; Vinyals; Le, 2014) and *bert* (Devlin *et al.*, 2019), two of the most recent models used to perform NLP tasks (these models will be addressed in Section 2.4).

As a matter of fact, still on Figure 2.3, the evolution assessment over the *NLP Cluster* draws a perfect picture of the recent history of NLP models as they have been developed and applied to economics and finance. The model nodes discussed early now range from



(a) NLP Cluster Zoomed-in

(b) NLP Cluster Evolution

(c) Lexicon Node Connections

(d) BERT Node Connections

Figure 2.3 – NLP Techniques Evolution Over Time

blue nodes — with terms such as *lexicon*, *bag*, for bag-of-words, and *svm*, referring to older approaches — to green nodes — representing intermediary models in the NLP timeline, being represented by *neural network*, *rnn* and *cnn* — and finally getting to yellow nodes — with *bert* and *lstm* as the best examples as mentioned before.

Panels 2.3c and 2.3d illustrate this evolution highlighting two examples: Panel 2.3c brings the dictionary-based representative node *lexicon* and its connections, whereas panel 2.3d shows the transformer-based *bert* model node, also with its connections (which, not surprisingly, has a strong link to the *transformers* node, alongside the *pre* node, presumably capturing the mentions in the literature of the pre-training trait of the BERT model). It is also interesting to see the common connections on these two examples: despite being clearly separated by some time in the NLP model timeline, in addition to being relatively

distant in the network representation considering the NLP models region, both of the nodes have strong connections to the *sentiment* node. This supports our previous conclusion that sentiment analysis is the most widespread NLP task found in the economic and financial literature, even when we look through a time perspective.

Finally, we can safely conclude from the visualization of the literature network through the lens of term occurrence evolution, that we observe a long dated pattern of assessing the performance of state-of-the-art NLP models to economic and financial research questions, a pattern that culminates in the most recent trend of resorting to transformer based models. Hence, we can expect to see this as the most prolific branch of this bibliometric network for the near-term future.

2.3 NLP Tasks

Having the bibliometric map as a guide to our survey through the next sections, we now move on to the discussion of NLP tasks most used by our target research field. With economic and financial applications in mind, the choice of NLP tasks are the next step to research question formulation. In other words, researchers should always ask themselves if the NLP task at hand is the most suitable for tackling the research problem of interest. This step should be done prior to NLP model selection, as it will mostly guide this choice to adequate alternatives. Thus, it must be carefully understood, as it will make the link between the actual NLP application and the broad range of methods currently available.

2.3.1 Sentiment Analysis

Sentiment analysis consists of a textual classification problem, in which documents are usually categorized as positive, negative or neutral based on the semantic tone of its text. In a broader sense, applications to economics and finance also adopt different scales according to the studied topic: whether bullish or bearish sentiment for stock market analysis or dovish or hawkish sentiment for monetary policy related texts. As a matter of fact, the classification techniques usually applied in sentiment analysis can also be extrapolated to other classification scales, not remaining restricted to positive-negative ranges. As an example of such approach, Figueredo, Mueller, and Cajueiro (2022) use a machine learning technique to classify politicians' speeches as left, center or right based on previously labeled data. Moreover, the approach in this task is usually of supervised learning, making it necessary to feed the model with labeled data in the training phase in order for it to perform latter classifications.

Early examples of sentiment analysis endeavors in finance resorted to dictionary methods to predict financial or operating performance of companies. Tetlock (2007) relate journalistic sentiment — as a proxy for investor sentiment — to stock market levels. He finds

that negative sentiment in this textual data can effectively predict lower returns and higher volatility on the following days, however with limited effects. Tetlock, Saar-Tsechansky, and Macskassy (2008), turning to individual companies' news articles, find that high occurrences of negative terms predict lower financial performance. Moreover, Heston and Sinha (2017) indicate that positive net news sentiment, accounting for positive minus negative word frequencies, are associated with higher stock returns, while negative net sentiment result in lower stock returns on firm level. Furthermore, works assessing earnings conference calls and press releases find positive relations between positive tone and stock returns and operating performance (Price *et al.*, 2012; Doran; Peterson; Price, 2012; Davis *et al.*, 2015; Davis; Piger; Sedor, 2012).

2.3.2 Topic Modeling

Topic modeling is the NLP task of determining the topics of a set of documents. Usually performed by applying dimensionality reduction and clustering algorithms to term-document matrices (which we will discuss in Section 2.4.3), this task leverages unsupervised learning techniques to reach outputs without the need of prior knowledge or assumption[7]. Either by functioning as a first layer of the analysis, aimed at selecting relevant documents for a specific latter study, or by providing latent topics as features, or even by helping explore the underlying dimensions of a given text corpus, topic modeling consists of one of the most employed NLP task in the economic literature (Boukus; Rosenberg, 2006; Correia; Mueller, 2021; Angelico *et al.*, 2022; Roos; Reccius, 2023).

Despite being a strength of this task in terms of flexibility of applications, the unsupervised nature of its most usual approach comes hand in hand with some challenges. First, latent topics extracted from text corpus are usually in the form of a set of words, which can make them often difficult to interpret. The researcher usually has to analyse the output to evaluate the meaning of each latent topic extracted. Furthermore, as the tools to perform this task are usually wholly data driven, one cannot target the analysis towards extracting specific topics of interest[8] (Mcauliffe; Blei, 2007).

2.3.3 Event Extraction

The event extraction task consists of gathering information about event instances from textual data, such as news articles or social media posts. It can be seen as composed by two steps: event detection, focused on identifying and classifying event triggers; and argument extraction, identifying arguments of the event and labeling their roles (Liu; Luo; Huang,

[7] One should note, though, that topic modeling is not a task performed exclusively by unsupervised learning techniques. For a supervised approach, refer to Mcauliffe and Blei (2007).

[8] Of course there are extensions that tackle these issues. For an example, refer to (Chang *et al.*, 2009).

2018; Grishman; Westbrook; Meyers, 2005). This task has knowledge graphs as one of the main tools adopted by researchers to extract event information from textual data (Deng *et al.*, 2019; Liu *et al.*, 2019b; Cheng *et al.*, 2022). It also provides some insight to better understand the underlying rationale behind the NLP model, a matter so important that has a branch of the artificial intelligence literature dedicated to it (Arrieta *et al.*, 2020; Chen *et al.*, 2023).

Diving deeper into knowledge graphs, they are a way to structure data with the focus on relations, central to event detection. On the graphs' nodes, they describe entities (such as agents, places, dates or concepts), and on their edges, the relations between these entities. Figure 2.4 shows an example of event extraction from Brexit's Wikipedia page [9].

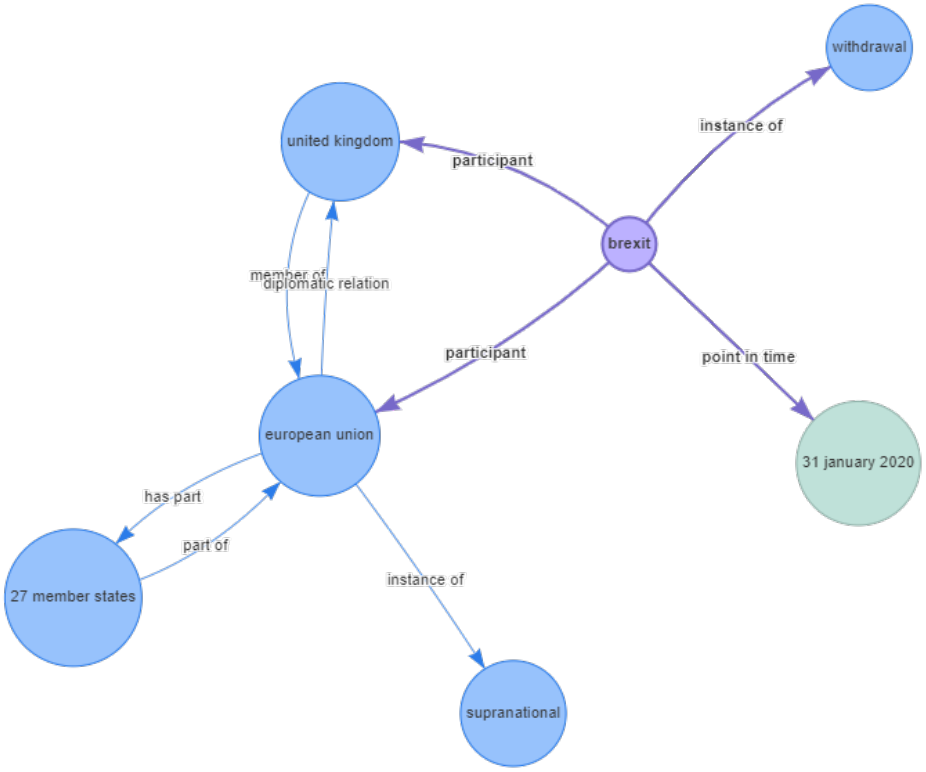


Figure 2.4 – Knowledge Graph Example: Brexit

Similar to event extraction itself, the process of gathering relations from textual data to form knowledge graphs can be seen as a two tasks problem. First, the entities are extracted from text as in named entity recognition (NER). Second, relation classification (RC) checks whether there exists any pairwise relation between the extracted entities (Zeng *et al.*, 2014;

[9] This graph was created by REBEL, a BART based relation extraction model, available at https://huggingface.co/spaces/ml6team/Knowledge-graphs (Cabot; Navigli, 2021).

Cabot; Navigli, 2021). Despite being of secondary interest to the economics and financial literature when compared to previously discussed tasks, Liu *et al.* (2019b) stand as an example of its application to stock market predictions.

2.3.4 Summarization

The task of automatic text summarization is based on generating a short piece of text from an original, longer document input, with the most important information preserved. The most common type of NLP summarization task is known an extractive summarization, in which the output is composed of the important sentences from the original text, extracted and joined together. It is opposed to abstractive summarization, which includes terms that were not originally present in the input document and involves writing the summary from scratch (Liu, 2019; Cajueiro *et al.*, 2023).

Automatic text summarization is commonly an accessory task in economic and financial research, as it is usually not enough to answer interesting research questions by itself. However, although secondary to academic applications, it is an important task for other purposes, mostly when researchers or analysts have to read large amounts of text from which the most important information could be readily extracted via NLP summarization. Chen *et al.* (2019) is an example of turning to extractive text summarization over a large corpus of financial news to select the most relevant ones based on time, topic and category. The authors latter apply a classification model to predict foreign exchange markets' movements.

2.3.5 Document Similarity

One last natural language processing task worth mentioning for its economic and finance applications is the one focused on measuring document similarity. The purpose of this task in the literature of our interest is usually to extract from textual data some underlying relation between agents that is not present in other kinds of data. It might be applied to news or social media data when aiming at the extraction of public perception of similarity, or it might be applied to corporate disclosure or product description, when the objective is to identify business similarities. The application of this task can contribute to industrial organization research, investment diversification, financial contagion, among others.

Albeit a straightforward task for humans, extracting semantic similarity can be rather challenging for algorithms. One of the most popular measures of document similarity is to apply cosine similarity to textual vector representations (which we will see in Section 2.4). Taking documents d_1 and d_2 represented by vectors a and b, cosine similarity can be calculated as:

$$\text{cosine similarity}(d_1, d_2) = \frac{\sum_i a_i b_i}{\sum_i a_i^2 \sum_i b_i^2},$$

where i indexes each vector's element, be it word count, sentence embedding or other dimension of text representation. From the above equation, we can see that it consists of dot product similarity, in the numerator, scaled by the vectors' lengths. Figure 2.5 brings examples of this measure applied to vector pairs.

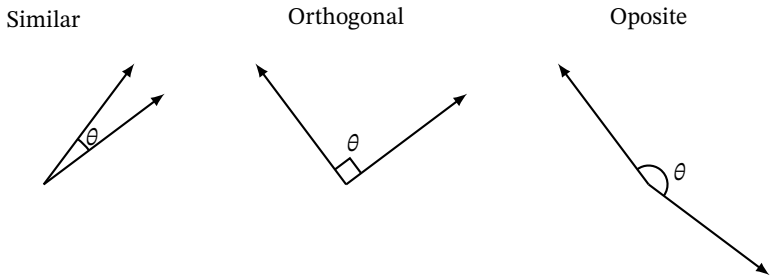


Figure 2.5 – Cosine Similarity

Still with the document similarity task in mind, a branch of the literature resorts to network theory, with networks built from textual data (Viegas *et al.*, 2019; Cajueiro *et al.*, 2021). Usually having term-document matrices as starting point for building networks — whether it be with or without dimensionality reduction techniques — this approach allows to frame the research questions in a different setup, not only providing further insights when compared to plain NLP methods, but also leveraging the network theory toolbox to assess such targeted relations. Document similarity has been widely used to extract companies' businesses similarity from corporate disclosure or news articles, whether to measure competition or risk exposures, including contagion risk (Hoberg; Phillips, 2016; Cajueiro *et al.*, 2021).

2.4 NLP Models

We dedicate this section to text modeling itself, with information retrieval, text representation and feature extraction techniques. Generally speaking, these models are core to NLP tasks, used to extract some semantic meaning to be treated in subsequent phases of research. One should keep in mind, though, that alongside with text representation models, the reviewed literature also usually adopts statistics, basic machine learning or deep learning models in order to give meaning to the information extracted from textual data. Among the most common are Logistic Regression (logit) (Berkson, 1944; Berkson, 1951), Support Vector Machines (SVM) (Cortes; Vapnik, 1995), K-Nearest Neighbor (KNN) (Cover; Hart, 1967), Random Forests (Breiman, 2001), gradient boosting (Friedman, 2001) and Naive Bayes

models (Duda; Hart *et al.*, 1973; Domingos; Pazzani, 1997)[10].

The text representation models in this section can be classified by whether or not they consider term relations when extracting features and quantifying concepts from textual data. For models that do not consider word relation, thus not representing word order or relative position, we will use the broad definition of bag-of-words (BoW). We will begin by discussing such models and move on to more sophisticated approaches. By the end of this section, we will dedicate some extra energy to understanding transformer networks, given their current relevance to NLP applications.

2.4.1 Lexicons

Harris (1954)'s distributional semantic hypothesis postulates that the meanings of words are related to their relative occurrence contexts and distributional properties. Building from that hypothesis and as previously mentioned, bag-of-words approaches model texts as an unordered collection of words, leaving aside their sequence of occurrence. Narrowing down to the most popular approaches in this category, dictionary techniques — also known as word lists — make use of tabulated collection of terms with associated attributes to measure the tone of documents. This is usually achieved by means of counting terms related to each sentiment. Furthermore, when created for very specific purposes, dictionaries are often referred to as lexicons (Loughran; Mcdonald, 2016).

The most commonly adopted dictionaries by economic and financial research are Henry (2008), Harvard IV-4, Diction and Loughran and Mcdonald (2011). Henry (2008) was one of the first word lists created specifically for the financial domain, build from companies' earnings press releases and adopted by other works mostly to study earnings conference calls (Price *et al.*, 2012; Doran; Peterson; Price, 2012; Davis *et al.*, 2015). Harvard IV-4 and Diction are general purpose dictionaries among the first made available to researchers. Harvard IV-4 is adopted by some of the most influential early papers in sentiment analysis in finance (Tetlock, 2007; Tetlock; Saar-Tsechansky; Macskassy, 2008; Heston; Sinha, 2017). This excerpt of the literature usually focused on assessing stock or financial performance based on news sentiments extracted by means of Harvard IV-4 positive and negative categories. Similarly, Diction's terms are usually grouped into positive and negative categories to infer sentiment from companies' earnings press releases and annual reports (Davis; Piger; Sedor, 2012; Davis; Tama-Sweet, 2012). Finally, for one of the most adopted dictionaries in finance (Kearney; Liu, 2014), Loughran and Mcdonald (2011) categorizes words in six different lists to build a comprehensive domain specific financial lexicon. Backing the rationale for a domain specific financial dictionary, Loughran and Mcdonald (2011) argue that the general purpose

[10] Since these models are not in the scope for our discussions, we recommend Bishop and Nasrabadi (2006) and Izenman (2008) for references.

ones fail to correctly infer sentiment by incorrectly classifying regular words to the business communications, such as cost or liability. Their dictionary has been widely applied to firm's business communications, news articles or social media content (Feldman *et al.*, 2010; Dougal *et al.*, 2012; Liu; McConnell, 2013; García, 2013; Chen *et al.*, 2014; Huang; Teoh; Zhang, 2014).

Dictionary-based methods are still widespread in the economic literature (Gentzkow; Kelly; Taddy, 2019), with recent works still dedicating to craft lexicons focused on specific questions (Apel; Grimaldi; Hull, 2022; Shapiro; Sudhof; Wilson, 2022; Ravassi, 2020). García, Hu, and Rohrer (2023) stand as a recent example of dictionary construction, resorting to machine learning techniques for it. They suggest that this approach outperforms the largely adopted dictionary by Loughran and Mcdonald (2011).

2.4.2 Term-document matrices

Still in the broader domain of Bag of Words (BoW), it becomes evident that mere counting of terms in documents does not suffice for effective document differentiation, especially when considering the influence of document length. This limitation led to the development of the Term Frequency - Inverse Document Frequency (TF-IDF) matrix, a very popular solution for this challenge[11]. The TF-IDF matrix is a $N_V \times N_D$ matrix that establishes a relation between a term and a document, where N_V is the vocabulary size and N_D is the number of documents in the dataset:

$$\mathbf{M}_{\text{tfisf}} = \begin{array}{c} \\ w_1 \\ w_2 \\ \vdots \\ w_{N_V} \end{array} \begin{array}{c} \begin{array}{cccc} d_1 & d_2 & & d_{N_D} \end{array} \\ \begin{bmatrix} \omega_{1,1} & \omega_{1,2} & \cdots & \omega_{1,N_D} \\ \omega_{2,1} & \omega_{2,2} & \cdots & \omega_{2,N_D} \\ \vdots & \vdots & \cdots & \vdots \\ \omega_{N_V,1} & \omega_{N_V,2} & \cdots & \omega_{N_V,N_D} \end{bmatrix} \end{array} \tag{2.1}$$

where each row is a term w and each column is a document d.

Thus, we may write the weight ω_{ij} that quantifies the importance of term i in document j as

$$\omega_{i,j} = \begin{cases} \frac{f_{\text{tf}}(\text{tf}_{i,j}) \times f_{\text{idf}}(\text{df}_i)}{n_j} & \text{if } \text{tf}_{i,j} > 0 \\ 0 & \text{if } \text{tf}_{i,j} = 0 \end{cases}, \tag{2.2}$$

where $f_{\text{tf}}(\text{tf}_{i,j})$ is the weight associated with the frequency of term i in document j, $f_{\text{idf}}(\text{df}_i)$ is the weight with the document frequency of the term i and n_j is the normalization term that takes into account the length of document j. Thus, the factor ω_{ij} depends on three different factors. The first factor f_{tf} (*local factor*) relates to the term frequency and captures

[11] For an early reference, see Salton and Buckley (1988).

the significance of a term within a specific document. The second factor f_{idf} (*global factor*) relates to the document frequency and gauges the importance of a term throughout the entire document. It calculates the inverse frequency of each term, taking into account the number of documents containing that term. This results in higher weights for terms that occur in a smaller number of documents, thus emphasizing terms that are unique or specific to particular documents. The third factor n (normalization) adjusts the weight to account for varying document lengths, ensuring comparability across documents.

Therefore, the TF-IDF framework addresses two main issues inherent to simple word counts: the overemphasis of term frequency in longer documents and the uniform treatment of terms regardless of their distribution across documents. By combining these three factors, TF-IDF accomplishes a dual objective: it normalizes term frequency within documents (TF) to address the issue of document length, and it scales the importance of terms based on their distinctiveness across the document corpus (IDF). There are many different ways to define these factors and we may find a collection of them in (Baeza-Yates; Ribeiro-Neto, 2008), (Manning; Raghavan; Schütze, 2008) and (Dumais, 1991).

A particular example of the evaluation of the weight ω is available in Loughran and Mcdonald (2011):

$$\omega_{i,j} = \begin{cases} \frac{1+\log(\text{tf}_{i,j})}{1+\log(a_j)} \cdot \log\left(\frac{N_D}{\text{df}_i}\right), & \text{if } \text{tf}_{i,j} \geq 1 \\ 0, & \text{otherwise,} \end{cases}$$

where $\text{tf}_{i,j}$ represents the raw count of term i in document j, df_i is the number of documents containing term i and a_j is the average word count in document j. Thus, $f_{\text{tf}}(\text{tf}_{i,j}) = \text{tf}_{i,j}$, $f_{\text{idf}}(\text{df}_i) = \frac{N_D}{\text{df}_i}$ and $n_j = a_j$.

When in the domain of sentiment analysis, text representations from these term-document approaches are usually latter combined with some supervised classification step, ranging from regression models (Tetlock, 2007; Tetlock; Saar-Tsechansky; Macskassy, 2008; Tetlock, 2010; Tausch; Zumbuehl, 2018) to more sophisticated machine learning approaches[12].

2.4.3 Latent Dirichlet Allocation (LDA)

Although Latent Dirichlet Allocation (LDA) (Blei; Ng; Jordan, 2003) is not strictly a model for representing text, it deserves particular attention in this discussion due to its prominence as a topic modeling algorithm within the economics and finance literature. LDA assumes that documents are composed of a mixture of topics, where each topic associates with a specific distribution of words. For mathematical clarity, we define the vocabulary $V =$

[12] See Kumbure *et al.* (2022) for a recent review.

$\{w_1, \ldots, w_i, \ldots, w_{N_V}\}$ as the set of all unique words across documents, and $I_V = \{1, \ldots, N_V\}$ as the indices of these words, with N_V denoting the total number of unique terms or the vocabulary size. A document $d = [w_{i_1}, \ldots, w_{i_k}, \ldots, w_{i_{l_d}}]$ is represented as a sequence of l_d words (not necessarily unique), indexed from the vocabulary, where $1 \leq k \leq l_d$ and $i_k \in I_V$, and V^d signifies the subset of the vocabulary present in document d.

LDA assumes a fixed, known number of topics K, and models each document d in a corpus $D = \{d_1, \cdots, d_{N_d}\}$ through the following generative process:

1. Determine the document's length l_d by drawing from a Poisson distribution with parameter ξ.
2. Draw the document's topic distribution θ from a Dirichlet distribution with a K-dimensional vector of positive parameters α[13].
3. For each word w_n in the document:
   a) Choose a topic z_n based on a Multinomial distribution parameterized by θ.
   b) Select a word w_n according to a Multinomial probability conditioned on topic z_n, where this probability is specified by a $K \times N_V$ matrix $\beta = p(w^j = 1|z^i = 1)$, for all $j \in \{1,...,N_V\}$ and $k \in \{1,...,K\}$.

Estimating the Latent Dirichlet Allocation (LDA) model involves determining the set of parameters that best explain the observed collection of documents under the assumption that each document is a mixture of a certain number of topics, namely the document length distribution parameter ξ, per-document topic distributions parameters α and per-topic word distributions β. There are different techniques to estimate it (Chauhan; Shah, 2021). A common method familiar to economists is the application of the Expectation-Maximization (EM) algorithm. The EM algorithm comprises two main steps: the Expectation (E) step and the Maximization (M) step. In the E-step, the algorithm calculates the expected values of the latent variables, which include the topics and their corresponding word assignments, based on the current estimates of the model parameters. Then, in the M-step, the algorithm maximizes the likelihood of the observed data, given the expected assignments of these latent variables.

Determining the optimal number of topics in a LDA model is a critical step for ensuring the model's effectiveness in capturing meaningful and distinct themes from a corpus of documents. One of the most effective approaches for this determination is the use of coherence measures. Coherence measures provide a quantitative way to evaluate the interpretability and quality of the topics generated by the model. It counts how many times certain word pairs appear in the same document. Higher scores suggest that the words in a topic are

[13] Note that the basic assumption of the LDA is that each document is a mixture of K topics. Thus, the variable θ represents this mixture for a particular document, where θ_i (the i-th element of θ) indicates the proportion of the document that is allocated to the i-th topic. Since θ is a distribution, its elements sum up to 1, with each value being between 0 and 1, representing the proportion of the corresponding topic in the document.

more related, indicating that the topic is more coherent. By calculating coherence scores for topics, we can compare different sets of topics to see which are more understandable and meaningful.

2.4.4 Word Embeddings

One of the most significant limitations of Bag of Words (BoW) approaches is their inability to capture the sequence of words in text representation, which is a critical aspect of textual meaning. Models for generating word embeddings marked a new era in NLP — in which algorithms could better relate word meanings when analyzing text — and also opened the doors to transfer learning in natural language representation — making in possible to leverage applications of large models pre-trained on huge amounts of textual content. First word embedding models used absolute embedding, with each word being represented by a single vector, independent of context. Latter evolution enabled having a context-based word embedding, with context also playing a role in determining each word's representation vector. Although the incorporation of deep learning adds complexity to the models and challenges their explainability, these techniques have been shown to significantly enhance the performance of textual representation (Wang; Yuan; Wang, 2020; Mishev *et al.*, 2020; Zhao *et al.*, 2021).

The main advance brought by word embedding models is their ability to classify words that appear in the same context as semantically close, assigning them with similar vectors. Furthermore, the high dimension space representation of semantics in these models enables one to perform meaningful calculations in the form of mathematical operations over word representation. Figure 2.6 presents an example of such calculations, relating "bonds", "fixed income" securities issued by companies, to "stocks", securities representing companies' "equity". We can see how the relation between those terms in the context of financial investments are represented in the embeddings vector space.

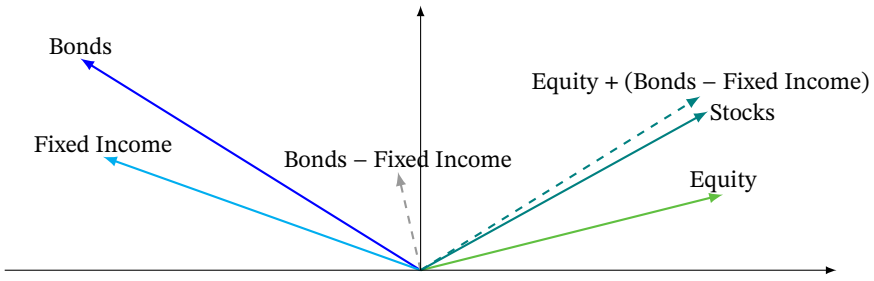


Figure 2.6 – Word Embeddings

The concept of word embeddings is inspired by classical n-gram models. An n-gram model, starting with a segment of text, predicts the next word based on the previous $n-1$ words, thereby estimating the probability of a given word's occurrence in this sequence. This probability is calculated by counting how often the n-th word appears in a sequence following

the previous $n-1$ words. Similarly, skip-gram models extend this concept by including the possibility of skipping k words in the sequence for analysis. Unlike classical n-gram and skip-gram models that rely on explicit word occurrence counting, word embedding models use neural networks to learn the characteristics of words that are likely to appear in a sequence.

Neural networks are advanced computational paradigms inspired by the biological neural networks that constitute animal brains. These models are intrinsically nonlinear, making them exceptionally effective for tasks such as nonlinear regression analysis and the classification of binary or multinomial responses (Haykin, 2009; Goodfellow; Bengio; Courville, 2016). A key feature of neural networks is their architecture, which is organized into a sequence of layers. The data flows from one layer to the next, starting from the input layer, where the independent variables (referred to as such by economists) are introduced, to the output layer, which produces the predicted outcome (akin to the dependent variable in economic models). Each layer consists of multiple neurons, or transformation units, that perform a specific operation: they calculate a weighted linear combination of outputs from the neurons in the preceding layer, and then apply a nonlinear transformation to this sum. A fundamental exemplar of neural network architecture is the multilayer perceptron (MLP), a fully connected network that, despite not being the focus of this discussion, is central due to its foundational role in understanding and developing various neural network models, including those considered in our work. An MLP is characterized by its organization into n_l layers, each consisting of M_l parallel neurons. In layer l, every neuron transforms the linear combination of its M_{l-1} inputs (the outputs of the $l-1$ layer) through a nonlinear function. The output of neuron j in layer l is given by:

$$y_j^{(l)} = h^{(l)}\left(\sum_{i=1}^{M_{l-1}} w_{ij}^{(l)} y_i^{(l-1)} + w_{0j}^{(l)}\right), \; j = 1, \cdots, n_l,$$

where the expression within the parentheses, $a_j^{(l)} = \sum_{i=1}^{M_{l-1}} w_{ij}^{(l)} y_i^{(l-1)} + w_{0j}^{(l)}$, denotes the activation of neuron j. The function $h^{(l)}$ is the activation function for neurons in layer l. The network's input, represented by $y_i^0 = x_i$ where $i = 1, \cdots M_0$, indicates that the neural network possesses M_0 inputs. Similarly, the network's output, denoted by $y_i^{n_l} = y_i^0$ for $i = 1, \cdots M_{n_l}$, highlights that the neural network has $M_{n_l} = M_o$ outputs. The weight matrices W^l, for $l = 1, \cdots, n_l$, also referred to as coefficients in econometrics, are essential parameters that we must estimate. The estimation process typically involves optimization techniques such as minimizing the squared error or maximizing a likelihood function through gradient descent methods, thereby refining the network's performance in predictive tasks. Although computing gradients for the output layer is straightforward, the challenge intensifies when dealing with the intermediate layers. This complexity arises due to the need for a systematic method to propagate the error information backward from the output towards the input layer, a process named "backpropagation". The backpropagation algorithm employs a recursive

approach to compute the gradients of intermediate layers effectively. The essence of this algorithm lies in its ability to use the computed gradients of layers closer to the output (the downstream layers) to calculate the gradients of the preceding layers (the upstream layers). This method depends on the principle of the chain rule from calculus, which allows for the decomposition of the derivative of a function with respect to its inputs into products of derivatives through each layer of the network. The process begins by evaluating the error at the output layer (or the likelihood of it). This error (or the likelihood of it) is then used to compute the gradient of the loss (performance) function with respect to each weight in the output layer directly. Subsequently, for each layer moving upstream, the algorithm recursively applies the chain rule to determine the gradients of the loss (performance) function with respect to the weights of that layer. This involves two key steps. The derivative of the loss (performance) function with respect to the activations of the neurons in the current layer, which is obtained by propagating the gradients back from the subsequent layer. The derivative of the activations of the current layer with respect to its weights, which, when multiplied by the result from the previous step, yields the gradients necessary for updating the weights. By iterating this process across all layers from the output back to the first hidden layer, backpropagation efficiently calculates the gradients required for weight adjustment across the entire network. One important characteristic of neural network models is their universal approximation capability. Given a sufficient number of neurons per layer, an adequate number of layers, and the right choice of nonlinear transformations, neural networks can approximate any continuous function to a desired degree of accuracy.

Several neural network architectures are available for generating word embeddings, with the Continuous Bag of Words (CBOW) architecture of the word2vec model being a significant example. This architecture uses contextual words from a text segment to predict the central word (Mikolov *et al.*, 2013a; Mikolov *et al.*, 2013b). For illustration, consider Adam Smith's quote, "Mercy to the guilty is cruelty to the innocent." within a CBOW model utilizing a window size of 2. This approach enables the use of multiple inputs to train (estimate) the neural network. For instance, the contextual words before the central word are “mercy", “to" and after the central word are “guilty", and “is" that we may use to predict “the", or words before the central word are “to" "the" and after the central word are "is" and "cruelty" that we may use to predict "guilty", among other combinations. Imagine having access to a vast text corpus, such as the entirety of Wikipedia or a complete collection of economic literature. This would allow the generation of millions of inputs for neural network training.

The CBOW model's architecture includes a neural network with an input layer sized according to the number M_w of contextual words, an intermediate layer of size M_h, and an output layer whose size, M_V, corresponds to the vocabulary. Contextual words fed into the neural network are represented as one-hot vectors, meaning each word is represented by a vector sized according to the vocabulary, with all elements set to zero except for the element

representing the word, which is set to one. We keep the parameters of this neural network model in two matrices: the contextual matrix V, which maintains the parameters linking the input layer with the hidden layer with size $M_V \times M_h$, and the output matrix W, which contains parameters for connecting the hidden layer to the output layer with size $M_h \times M_V$.

Mathematically, the output of the l-th intermediate neurons is

$$y_l^h = \sum_{w=1}^{M_w} e'_w V_l, \text{ for } l = 1, \cdots, M_h,$$

where the activation function of the intermediate neuron is linear, e_w is the one-hot vector associated to the contextual word l and V_l is the column l of the contextual matrix V. In addition, the output l of the neural network gives the probability of the central word be the l-th word of the dictionary. This output is evaluated as

$$y_l^o = h^o \left(\sum_{i=1}^{M_h} y_i^h w_{il} \right), \text{ for } l = 1, \cdots, M_V,$$

where y_i^h is the output of the neuron i of the hidden layer, w_{il} is the element of i-th row and the l-th column of the matrix W and h^o is the softmax function given by

$$h^o(x_i) = softmax(x_i) = \frac{\exp(x_i)}{\sum_{j=1}^{M_V} \exp(x_j)},$$

also used in multinomial logit models.

Despite the numerous details involved in its estimation, the process resembles that of estimating a multinomial logit model, aiming to predict the central word by maximizing the likelihood function associated with the multinomial logit model. The word embedding for a specific word is derived by averaging the row vector of the contextual matrix and the column vector of the output matrix, both associated with that word. Thus, the word embeddings are vectors of dimension $d_{\text{embedding}} = M_h$. Another architecture available in the word2vec model is the so-called skip-gram model that predicts the contextual words from the central word. In the example above that uses the Adam Smith's quote, we could use the word "guilty" to predict "to" "the" "is" and "cruelty".

In machine learning literature, models are typically categorized into supervised and unsupervised models, each with distinct approaches to learning from data. The supervised learning paradigm uses labeled training data to learn a function that can predict outcomes for unseen data. For example, consider the task of classifying emails into "spam" and "non-spam". To train a supervised model for this task, we require a dataset with emails that are already labeled as spam or not. This labeled dataset enables the model to learn the characteristics of spam and non-spam emails. After the model has learned these characteristics by adjusting its parameters, it can then be applied to a new, unlabeled dataset to classify emails as spam

or non-spam. Unsupervised learning algorithms do not rely on labeled data. These models discover inherent structures within the data without the need for explicit labels. Examples include TF-IDF models and Latent Dirichlet Allocation (LDA) models, as introduced respectively in Sections 2.4.2 and 2.4.3. These models identify patterns and relationships in the data, such as grouping similar documents, without predefined categories or labels. A special case within the spectrum of learning paradigms is self-supervised learning. This approach, which can seem like a subset of supervised learning, uses unlabeled data by generating appropriate labels from the data itself. An example is training a CBOW model for generating word embeddings. Although it appears to require labels (the central word in a context), these "labels" are directly extracted from the input data, without the need for external labeling. Hence, CBOW and similar models are termed self-supervised, as they fabricate their own supervision signal from the unlabelled data.

Other classical models for generating word embeddings are GloVe and ELMo. Glove (Global Vectors for Word Representation) uses two architectures for word representation (Pennington; Socher; Manning, 2014). Besides the one used in the skip-gram method, GloVe employs global matrix factorization, responsible for capturing global relations between words. However, both word2vec and GloVe have an important drawback: they both result in a one-to-one mapping for word representation, with each word mapped to only one vector, independent of context. This means that words with more than one meaning would have a single vector representation for all of them[14]. ELMo (Embeddings from Language Models) specifically addresses this issue sequence models of language explored in Section 2.4.5, standing as one of the first embedding technique to successfully deliver contextual word embeddings (Peters *et al.*, 2018). In order to do so, it resorts to a deep learning model pre-trained on large corpora of textual data. At this point, transfer learning — a feature so important to other fields of deep learning, such as computer vision — was beginning to be a reality for NLP as well (Malte; Ratadiya, 2019). [15]

In terms of practical applications, embedding models usually provide input in the form of semantic meaningful features, for latter analysis performed by other models. In a finance-directed work, Minh *et al.* (2018) propose a domain specific sentiment word embedding model, Stock2Vec, trained on stock news and Harvard IV dictionary.

Besides word embedding models, further evolution was made to sentence embedding, capturing in a semantic vector of high dimension a representation of the entire sentence. Doc2Vec (Le; Mikolov, 2014) stands as an example of this kind of textual representation approach. Li, Zhao, and Moens (2023) bring a recent overview of embedding models.

[14] Maybe the most straightforward example of these words is "bank", which can have different meanings when associated with different contexts. Take, for instance, "bank account" or "river bank".

[15] For a survey on vector space models of text representation, seeTurney and Pantel (2010).

2.4.5 Sequence models

Sequence-to-sequence deep learning models used for tasks of NLP are NN models in which the inputs and outputs are sequences of tokens of varying sizes. In order to explore the most recent approaches of sequence-to-sequence models we need to trace back to the first architectures of Recurrent Neural Networks (RNN) (Jordan, 1986; Elman, 1990) and limitations of the MLP to deal with the problem of predicting words in a sequence. Imagine that we want to use a MLP perceptron to deal with the n-grams problem of predicting the next word in a sequence described in Section 2.4.4. In order to do that we need to fix the size of the sequence like we did fixing the size of the context window in the CBOW model. Second, since the structure of the MLP is fixed, the same word in different positions would be treated differently. RNNs are sequence models that deal directly with these two modeling constraints of the standard multilayer perceptrons that are essential to model sequences. They resemble the structure of time series models when they have a latent model such as the structure of state space models (Hamilton, 2020). We may summarize the vanilla RNN by the set of equations:

$$s^i = g(W_{ss}s^{i-1} + W_{sx}x^i + b_s) \tag{2.3}$$

and

$$y^i = g(W_{ys}s^i + b_y), \tag{2.4}$$

where x^i is a word in a piece of text, y^i is the word we want to predict, s^i is the state of the RNN, W_{ss}, W_{ax}, W_{ys}, b_s and b_y are the parameters of the network that we need to learn and s^0 is a vector of zeros. Figure 2.7 represents this model. Like the other models of language, we try to predict the probability of the next word. Therefore, we may estimate the parameters of this model in a self-supervised fashion using the backpropagation algorithm adapted to this case.

Suppose, for instance, that your text includes the quote due to Milton Friedman “There is no such thing as free lunch.". Thus, x^1 = “there”, x^2 = “is”, x^3 = “no”, x^4 = “such”, x^5 = “thing”, x^6 = “as”, x^7 = “free”, y^1 = “is” y^2 = “no”, y^3 = “such”, y^4 = “thing”, y^5 = “as”, and y^6 = “free”, y^7 = “lunch”. Thus, this algorithm tries to maximize the probability that the token “is" arises when the token “there" is an input, to maximize the probability the token “no" happens when "is" is the input and s_1 is the state of the system generated by “there" and so on. Unfortunately, vanilla RNNs are not good at dealing with long sequences. Problems that arise in this context are the so-called *exploding* and *vanishing* gradients (Bengio; Frasconi; Simard, 1993; Pascanu; Mikolov; Bengio, 2013; Ribeiro *et al.*, 2020). This happens naturally due to the algorithm of backpropagation, which is based on the chain rule of calculus, and the consequent multiplication of the same shared matrix of the parameters several times.

In order to overcome the difficulties of exploding and vanishing gradients, two important models of RNNs were introduced, namely Long Short-Term Memory (LSTM) (Hochreiter; Schmidhuber, 1997; Gers; Schmidhuber; Cummins, 2000) and Gated Recurrent Units (GRU) (Cho *et al.*, 2014) empirically explored in (Chung *et al.*, 2014). The basic idea behind these models is to replace the simple units of the vanilla RNN model with complex units which include gates that control the flow of information that passes from one unit to the other.

A natural extension of the vanilla RNNs is to consider the case where the input sequence and output sequence have different sizes. These models are called *Encoder-Decoder* models and they aim at mapping one sequence to another sequence (Sutskever; Vinyals; Le, 2014; Vinyals; Le, 2015), where the encoder (decoder) is the part of the model that deals with the input (output) sequence. In tasks like translation, where the model must grasp the full meaning of a sentence before producing its equivalent in another language, this separation enables a deep contextual understanding of the input and a coherent, contextually accurate output.

Figure 2.8 presents an example of this topology, where each unit of this model is the RNN unit (vanilla, LSTM or GRU). In this type of model, the intention is, given a sequence, to predict another sequence not necessarily of the same size. For example, consider that we want to use a sequence-to-sequence model to translate the quote due to Friedrich Hayek “The price system is a mechanism for coordinating knowledge" to Spanish. We may have x^1 =“the", x^2 =“price", x^3 =“system", x^4 =“is", x^5 =“a", x^6 =“mechanism", x^7 =“for", x^8 =“coordinating", x^9 =“knowledge", y^1 =“El", y^2 =“sistema’", y^3 =“de", y^4 =“precios", y^5 =“es", y^6 =“un’", y^7 =“mecanismo", y^8 =“para", y^9 =“coordinar", y^{10} =“el’" and y^{11} =“conocimiento" in Figure 2.8, where $T_x = 9$ and $T_y = 11$.

Applications of sequence models in finance are mostly twofold. One approach is to use the last hidden state s^{T_x} of recurrent units (or, similarly, the encoder output in the

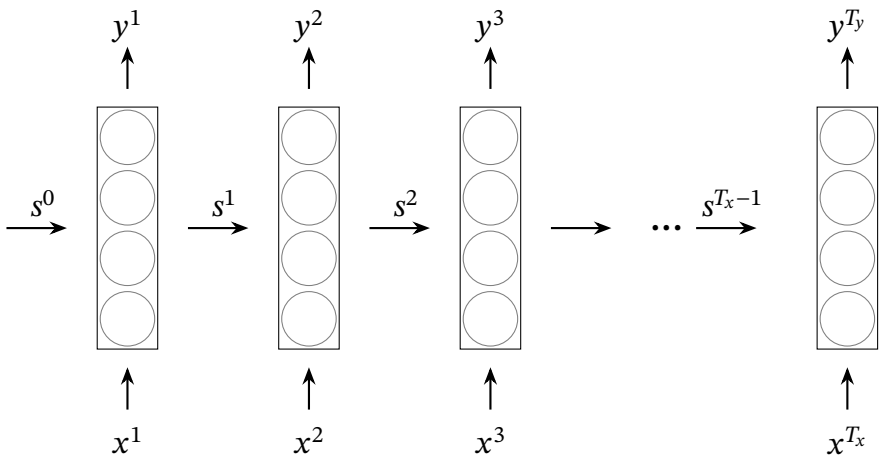


Figure 2.7 – Vanilla RNN.

case of Encoder-Decoder models) to capture the semantic representation of textual data and perform some NLP task, such as sentiment analysis. This approach is thus focused on textual sequences, compiling meaningful information gathered from the many iterations of the hidden state passed through recurrent units (Wang; Nulty; Lillis, 2020; Maia *et al.*, 2021). However, sequence models are also widely adopted to perform financial time series forecasting. In this alternative approach, the modeled sequence is the time series itself, with past data as input and future data as output. One should note, though, that this approach is not centered on textual data, but can adopt financial texts as one of many additional inputs (Jiang, 2021; Kumbure *et al.*, 2022).

2.4.6 Transformers

We finally get to the nowadays state-of-the-art text models. In a different direction than that of previous approaches, which used recurrent neural networks to model text in a sequence to sequence manner (Jordan, 1986; Elman, 1990), an archetypal transformer is a neural network with the presence of three fundamental components (Vaswani *et al.*, 2017): the utilization of word embeddings, the application of position encodings and the incorporation of attention mechanisms. Word embeddings, as introduced in Section 2.4.4, transform words into vectors of real numbers. Position encoding plays a vital role in preserving the order of words. To accomplish this, a unique vector is added to the embedding of each word, encoding its position within the sequence. Attention Mechanisms allows the model to selectively concentrate on various segments of the input data (Bahdanau; Cho; Bengio, 2014). Within a transformer, this mechanism ensures that each word can relate to every other word in a sequence. The computation of attention vectors involves three matrices: the query matrix, holding information about the query or target; the key matrix, containing details about the data available for query; and the value matrix, storing information on the data to be retrieved. These matrices are respectively used to generate the respective query, key and value vectors that facilitate the model's understanding of context and relationships within the text. Transformers actually work with multi-head attention that is the idea of evaluating several attention models simultaneously.

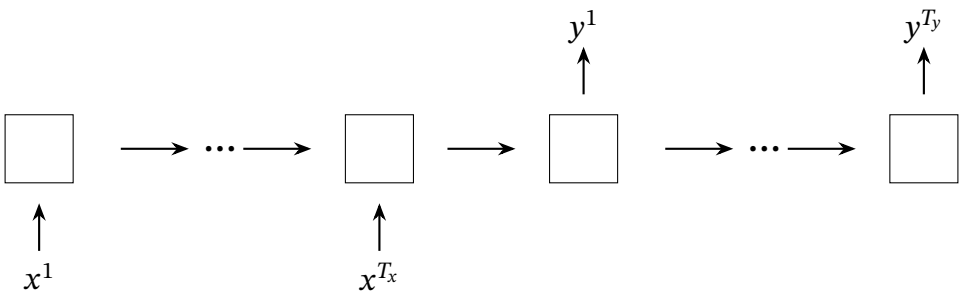


Figure 2.8 – Sequence-to-sequence models.

Although there are many different transformers' architectures, a typical transformer (Vaswani *et al.*, 2017) runs through a sophisticated architecture that comprises encoder and decoder components working in a parallel construction. The process begins with the encoder, which starts converting one-hot vectors into word embeddings. In the beginning, these word embeddings are filled by random numbers, but they are updated as the transformer starts understanding the meanings among the words. After generating the word embeddings associated with the words that are being inputted, the transformer has to indicate the position of each word in a sentence since it does not have a sequencial structure as the models presented in Section 2.4.5. Positional encoding describes the location or position of an entity in a sequence so that each position is assigned a unique representation. Transformers use a clever positional encoding scheme, where each position/index is mapped to a vector. More precisely, they use

$$PE_{(k,2i)} = \sin \frac{k}{n^{2i/d}}, \text{ for even positions}$$

and

$$PE_{(k,2i+1)} = \cos \frac{k}{n^{2i/d}}, \text{ for odd positions,}$$

where k is the position of an object in the input sequence, d is the dimension of the word embedding vectors, n is a pre-defined scalar (Vaswani *et al.* (2017) used $n = 10000$) and $0 \leq i < d/2$ is used to map the indexes of the word embedding vector. Consider, for instance, the quote due to Ludwig von Mises "Capital does not reproduce itself". Suppose for simplicity that $n = 100$ and the dimension of the word embeddings $d = 4$. The first word "capital" receives the positional encoding given by $(PE(0,0), PE(0,1), PE(0,2), PE(0,3)) = (0,1,0,1)$. The second word "does" receives the positional encoding given by $(PE(1,0), PE(1,1), PE(1,2), PE(1,3)) = (0.84, 0.54, 0.09, 0.99)$. The third word "not" receives the positional encoding given by $(PE(2,0), PE(2,1), PE(2,2), PE(2,3)) = (0.91, -0.42, 0.20, 0.98)$. Finally, the forth word receives the positional encoding given by $(PE(3,0), PE(3,1), PE(3,2), PE(3,3)) = (0.14, -0.99, 0.30, 0.95)$. It is worth mentioning that Vaswani *et al.* (2017) tested to generate the positional encoding by learning, but they did not find best results. Thus, each word is then represented by the sum of the original word embedding associated with each word added by the vector encoding. We call E is the matrix that concatenates the vectors that are the sum of the word embeddings associated with the words in a sentence and the correspondent position encoding. Thus, in this example this matrix has size 5×4 and, in general, this matrix has size given by the number of elements in the sentence times the dimension of the word embeddings.

Each encoder layer consists mainly of multi-head self-attention mechanisms. Each single self-attention mechanism generates queries, keys, and values to capture the contextual relationships within the input sequence. The goal of self-attention is to estimate the relative

relevance of the keywords compared to the query word for the same entity. Mathematically, the self-attention mechanism is evaluated using

$$A(q,k,v) = softmax\left(\frac{qk^T}{\sqrt{d_K}}\right)v,$$

where $q = EQ$, $k = EK$ and $v = EV$, Q, K and V are matrices of the same dimensions $d \times d_K$ (d_K is a pre-defined dimension of the Q, K and V matrices) that are estimated. In particular, Q and K matrices help to find correlations in the input sequence. On the other hand, the scaling factor $\sqrt{d_K}$ is used to address the gradient vanishing problem of the softmax function, since for large values of d_K, the dot products grow large in magnitude, pushing the softmax function into regions where it has extremely small gradients. Furthermore, note that the product qk^T is a square matrix of size given by the length of the sentence. The highest values arise between terms that have “aligned" vectors. For instance, using the example above with the quote of Ludwig von Mises “Capital does not reproduce itself". The product qk^T could be something like

	Capital	does	not	reproduce	itself
Capital	3.1	0.8	1.2	1.8	1.5
does	0.8	2.8	2.2	1.5	0.8
not	1.2	2.2	2.5	2.1	0.9
reproduce	1.8	1.5	2.1	2.5	1.7
itself	1.5	0.8	0.9	1.7	2.2

The softmax function is evaluated in a row basis looking at the most important information in each row of the matrix. This result is used to weight the value v matrix. The idea of multi-head mechanism is that H different combinations of matrices Q, K and V are used simultaneously. The output of the H attention heads are then concatenated and weighted by matrix W^o, resulting in embedding of the same dimension d of input embeddings. Finally, the result passes through a residual connection layer, which adds the embeddings with position encoding from before the attention mechanism, and can also be further passed through normalization and through feed forward layers. Encoders are usually stacked so that the output of one encoder layer is passed as input to the next one. Figure 2.9 depicts the inner workings of a transformer encoder as just described.

The decoder, in parallel, starts with its unique multi-head attention mechanism, designed to focus on different positions of the encoder’s output. The decoder’s initial multi-head attention block processes the output embeddings, enriched with position encoding, to maintain the sequence order in the generation process. It utilizes the queries generated from these embeddings while employing the keys and values from the encoder to integrate the input context. Following this, the decoder layers, mirroring the structure of the encoder layers but with an additional cross-attention mechanism, work to generate the output sequence

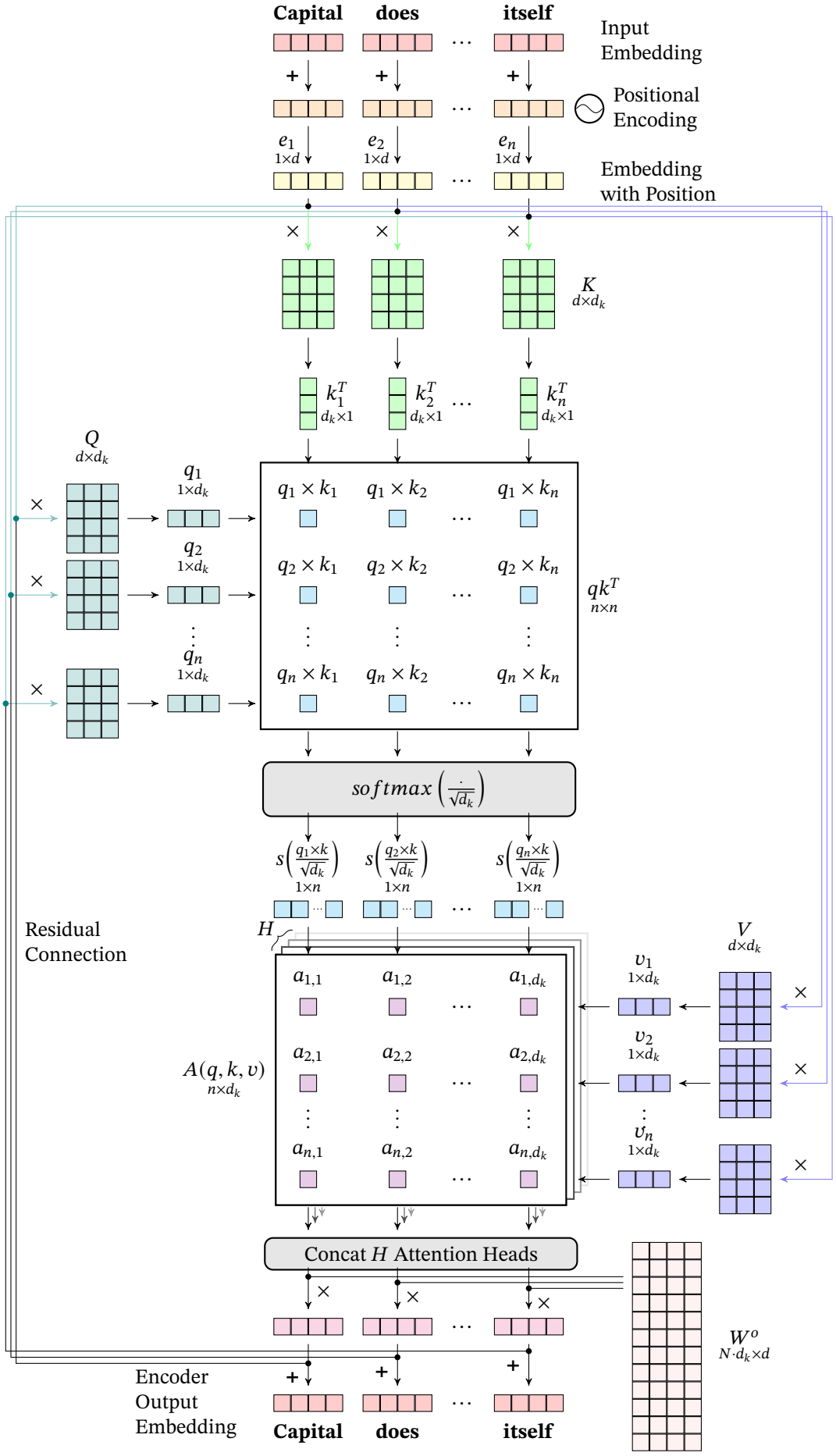


Figure 2.9 – Transformer encoder.

recursively. Except for inputs, cross-attention calculation is the same as self-attention. Cross-attention combines asymmetrically two separate embedding sequences of same dimension (the input sequence and the output sequence), in contrast self-attention input is a single embedding sequence. This is helpful for activities such as translation where we need to compare two different sequences.

Although the transformer architecture we have described is composed of two separate parts (encoder-decoder structure), this architecture is not unanimous practice and many different architectures are possible. There are, for instance, encoder-based transformer frameworks that focus exclusively on the encoder part of the original transformer architecture such as BERT (Bidirectional Encoder Representations from Transformers) (Devlin *et al.*, 2019). These models are particularly suited for tasks that require understanding or encoding information without the need for generating new text sequences. This representation is usually latter used for tasks that require a certain level of language understanding, such as sentence classification.

On the other hand, there are also decoder-based transformer frameworks like GPT and its iterations (Radford *et al.*, 2019; OpenAI, 2023). They are designed to generate text in an autoregressive manner. In these models, the generation process involves predicting the next token in a sequence given all the previous tokens, making them particularly well-suited for tasks like text completion, content generation, and language modeling. This approach has recently gained attention due mostly to its models' ability to understand tasks through user input prompts.

Deepening on transformer-based models, we will begin by the most popular in scientific literature: BERT. As previously mentioned, BERT is an encoder-based transformer model proposed by Devlin *et al.* (2019). This model brought new standards to NLP, achieving state-of-the-art performance in most tasks. As an encoder-based model, it focuses on providing context-rich embeddings from text — for both single words and entire sentences — which can be further used to perform various NLP tasks. More than that, BERT enables fine-tuning by just plugging an output layer to the network, such as a classification layer for sentiment analysis, and providing a relatively small labeled dataset. In this step, all parameters trained in the pre-training stage are fine-tuned to the specific task/data at hand.

Besides its bidirectional trait[16], that allows more context to be incorporated in word representations, BERT's pre-training strategy is one of the main determinants for its performance success. It consists of two unsupervised tasks: Masked Language Model (MLM) and Next Sentence Prediction (NSP). The former first masks a percentage of terms in the text sequence, which must be predicted by the model. Devlin *et al.* (2019) argue that this bidirectional training objective leads to a more powerful model than either left-to-right

[16] The bidirectionality means that, conceptually, the model has the capacity to understand and integrate the context from both the left and the right sides of a token simultaneously.

or shallow concatenation of left-to-right and right-to-left approaches. The latter feeds the model with sentence pairs, with 50% actual sequential sentences and 50% random sentences from the corpus. The embeddings of both sentences are then used to predict if the pair is in fact a sequence. Devlin *et al.* (2019) demonstrate that this step is very beneficial to tasks that involve two sentences, such as question and answering. Pre-training is performed on BookCorpus (800M words) and English Wikipedia (2,500M words).

Since Devlin *et al.* (2019), the NLP literature has developed a handful of derived BERT models, either by (i) improving some of its features or by (ii) focusing on specific tasks or domains. Representing the first group, ALBERT (Lan *et al.*, 2020) addresses scalability issues by standing as a lightweight successor, outperforming BERT in several tasks while lowering memory consumption and increasing training speed; RoBERTa (Liu *et al.*, 2019a) suggests the original BERT model was undertrained, thus improving pre-training procedure and increasing training corpus size and; DistilBERT (Sanh *et al.*, 2020) significantly reduces the size of the original model, retaining most of its language understanding capabilities. On the second group, we highlight FinBERT (Araci, 2019), a BERT architecture further pre-trained on financial text corpus.

Aside from BERT, other transformer-based models have been developed. The popular GPT, currently in its fourth iteration (Radford *et al.*, 2019; OpenAI, 2023) leverages a transformer decoder-based architecture to perform generative language modeling. XLM (Lample; Conneau, 2019) is a transformer-based cross-lingual model. And BART (Lewis *et al.*, 2019) uses a transformer encoder-decoder architecture, combining the ability to perform both natural language understanding and generation.

One should also note that transformers have important limitations as well. One of the most relevant is related to model efficiency, specially when faced with the challenge of processing long sequences. Formally, this happens because self-attention has a complexity of $\mathcal{O}(T^2 \cdot D)$ where T is sequence length and D is the model's hidden dimension. From that representation, we can see that transformer complexity increases by the order of the squared length of textual sequences modeled. To address this issue, there have been proposed models based on transformer variants by the NLP literature (Lin *et al.*, 2021). The Longformer (Beltagy; Peters; Cohan, 2020) resorts to a sparse attention modification, making use of the fact that the attention matrix is usually very sparse; and Linear Transformer (Katharopoulos *et al.*, 2020) use linearized attention to replace the original dot product attention. Furthermore, Transformer-XL (Dai *et al.*, 2019) relaxes the fixed length restriction — enabling longer-term dependencies and avoiding context fragmentation — by resorting to a recurrence mechanism and a different position encoding scheme.

Finally, it is interesting to highlight some practical considerations about using transformer-based models in economics and finance. It is worth mentioning that, as a way to leverage the transfer learning capabilities of transformers, avoiding the costly training phase, one

can resort to pre-trained versions of such models in open-source communities. Qiu *et al.* (2020) argue that pre-training performed on large corpus enables transformer models to learn universal language representations, which can be then transferred to domain-specific applications through fine tuning. We point to *huggingface* as one of the most complete open-source repositories, with implemented models and pre-trained parameters made readily available through the *transformer* Python package[17] (Wolf *et al.*, 2020). However, there are other packages made available by authors themselves[18].

Beginning with model choice, there are two main paths to follow. First, domain specific version of transformer models provide some ready-to-use alternatives for economic problems. Models such as FinBERT (Araci, 2019) or CentralBankRoBERTa (Pfeifer; Marohl, 2023) are examples of this approach[19]. If the choice is for one of such models, it is possible to apply it in an out-of-the-box fashion, without the need to organize some problem specific training dataset or to resort to the improved computational power needed in the training phase of large language models. If this is the chosen path, one only has to prepare the data for it to fit the format expected by the model, usually using the tools provided with the chosen model, and perform the NLP task as discussed in Section 2.3.

However, if the application is very specific or if the researcher needs improved adherence to the problem at hand, some further training might be necessary for the model to perform accordingly. This further step is called fine tuning, which consists of feeding application-specific training data to a model with general language knowledge so that pre-trained parameters are tilted to new values. Note that this can be done to both: the previously mentioned domain-specific models or to general versions of models (e.g. BERT or RoBERTa). Fine tuning requires only a fraction of the computational resources and data needed for pre-traning, as it does not require the training step to capture all the language nuances and structures, but only the ones specific for the application at hand, such as jargon or technical aspects of domain-specific texts.

In spite of being much simpler than training a large language model from the beginning, fine tuning is, as a matter of fact, a deep learning model training step. Thus the researcher has to pay attention to hyperparameters, special model parameters that control the learning process, i.e., the estimation of the other parameters of the model. As these hyperparameters are usually very sensitive to model choice, papers that propose new models usually bring some recommendations for preferred hyperparameter values, which can be used as a guideline for the researcher in this training step[20]. Furthermore, some practical

[17] https://github.com/huggingface/transformers.

[18] One example is BERT by Devlin *et al.* (2019), Turc *et al.* (2019), available at https://github.com/google-research/bert.

[19] FinBERT is available at https://huggingface.co/ProsusAI/finbert and CentralBankRoBERTa is available at https://huggingface.co/Moritz-Pfeifer/CentralBankRoBERTa-agent-classifier

[20] For an example of such hyperparameters recommendation for the BERT model, see Devlin *et al.* (2019).

applications might not be very rich in terms of data availability, which is particularly true for some economic applications. Moreover, some NLP tasks, such as sentiment analysis, require annotated data for model training, which in turn can make generating new data even more demanding or expensive. If limited data availability is an issue, one possible solution might be to adjust hyperparameters accordingly. There is dedicated literature that one can resort to in these cases, the *few sample learning* research (Lu *et al.*, 2023). For BERT, we refer to Zhang *et al.* (2020).

2.5 Textual Data

Our main goal in this section is to provide practical directions for the most used data sources in previous works. Moreover, we will also highlight previous contributions in terms of annotated data, crucial for any supervised learning task. Then, by the end of the section, we provide some commentary on data preparation methods, listing the most common preprocessing steps and providing further reference on the subject.

2.5.1 Data Sources

Maybe the two most straightforward types of textual data for economics and financial research are news articles and social media data. As mentioned before, both of them allow for higher frequency updates when compared to traditional economic data. Regarding news, there are many providers, such as *The New York Times* (https://www.nytimes.com/), *The Wall Street Journal* (https://www.wsj.com/), *Bloomberg* (https://www.bloomberg.com/) and *Thomson Reuters* (https://www.reuters.com/), in a non-exhaustive list. The choice of which provider to be used should be mostly guided by domain specialization and geographic relevance, matching those of the research topic at hand. For social media, the most popular sources in the literature are X (former Twitter — https://twitter.com/), Stocktwits (https://stocktwits.com/) and Seeking Alpha (https://seekingalpha.com/). Some of them provide API for automatic data capturing, allowing for data selection criteria parameterization.

When working with this kind of textual data, be it in the form of news articles or social media content, one of the first challenges researchers face involves the selection of relevant pieces of news. It is an important step in the extraction of information from text given that such data is always marked by high levels of noise, alongside with redundancy. The approaches adopted by the reviewed literature are twofold. On the one hand, most papers focus on applying a set predefined rules in order to achieve such selection. On the other hand, some works resort to NLP techniques, modelling this problem as tasks that range from text summarization to information extraction (Farimani; Jahan; Fard, 2022).

Furthermore, besides the sources mentioned thus far, another type of widely adopted data — mostly by financial researchers seeking company-specific information — is corporate

disclosures data. Loughran and Mcdonald (2016) point to 10-K annual filings with the Security Exchange Commission (SEC)[21], recommending that the use of this document be focused on specific sections, such as Management Discussion and Analysis (MD&A).

Going for company level to macroeconomic level, maybe the single most adopted type of data by economic researchers is central bank communication. The popularity of such textual data comes hand in hand with the amount and quality of data available, made possible by the transparency standards for monetary policy and its challenge of steering economic agents expectations both on prospective inflation and future interest rates.

Regarding the sources for central bank communication documents, we point to two main data sources. First, the Bank for International Settlements (BIS) provides an aggregated source of central bankers' speeches ranging from 1997 to current date[22]. Second, we can find communication documents for most central banks made publicly available on their own websites. These documents usually range from monetary policy decision statements, press conference transcripts, monetary authority meeting minutes to even full meeting transcripts. Table 2.2 brings the sources of minutes or statements for the most relevant central banks globally[23].

2.5.2 Annotated Data

Supervised learning techniques are the ones that require labeled data for the model to learn patterns that lead to the correct output in the training phase, and then replicate such patterns to new, unseen information. Considering the tasks discussed in Section 2.3 and practical applications in economics and finance, sentiment analysis is the most adherent to this approach. However, there are also examples for topic modeling and automatic summarization in the literature, as mentioned when discussing those tasks.

With that need in mind, some previous works provide valuable annotated data for sentiment analysis model training or fine-tuning. This is the case of Niţoi, Pochea, and Radu (2023), who provide 1,998 central bank communication sentences labeled as hawkish, dovish or neutral[24]. Similarly, but in a more general sentiment analysis approach, one of the most popular datasets in financial applications is the Financial Phrase-Bank (Malo *et al.*, 2014). It consists of 4,845 financial news sentences labeled by 16 experts as positive, neutral or negative[25]. One of the advantages of this dataset, it is also annotated according to experts'

[21] https://www.sec.gov/edgar/searchedgar/companysearch.

[22] https://www.bis.org/cbspeeches/index.htm. Moreover, Hansson (2021) provides a dataset from this source, which he used in his topic modeling study.

[23] These sources are meant as guidelines and were gathered on November/2023. The provided URL may not be active in the future.

[24] Niţoi, Pochea, and Radu (2023) make both the data and their sentiment index available at https://sites.google.com/view/bert-cbsi/.

[25] The Financial Phrase-Bank is available at https://huggingface.co/datasets/financial_phrasebank.

Country	URL
USA	https://www.federalreserve.gov/monetarypolicy/fomccalendars.htm
Euro Zone	https://www.ecb.europa.eu/press/accounts/html/index.en.html
England	https://www.bankofengland.co.uk/monetary-policy-summary-and-minutes/monetary-policy-summary-and-minutes
Japan	https://www.boj.or.jp/en/mopo/mpmsche_minu/minu_all/index.htm
China	http://www.pbc.gov.cn/en/3688229/3688311/3688329/index.html
Sweden	https://www.riksbank.se/en-gb/press-and-published/minutes-of-the-executive-boards-monetary-policy-meetings/
Norway	https://www.norges-bank.no/en/topics/Monetary-policy/Monetary-policy-meetings/?tab=newslist
Canada	https://www.bankofcanada.ca/publications/summary-governing-council-deliberations/
Australia	https://www.rba.gov.au/monetary-policy/rba-board-minutes/
Brazil	https://www.bcb.gov.br/en/publications/copomminutes
Mexico	https://www.banxico.org.mx/publications-and-press/minutes-of-the-board-of-governors-meetings-regardi/minutes-regarding-monetary-po.html
Chile	https://www.bcentral.cl/en/web/banco-central/areas/monetary-politics/monetary-policy-meeting-rpm

Table 2.2 – Central bank communication sources

agreement. This allows researchers to choose between using a more restricted text corpus, with higher levels of label consensus, or a broader corpus, giving up some consensus in such classification.

Turning to supervised topic modeling, we refer to the Twitter Financial News dataset[26]. It provides 21,107 documents annotated with 20 labels, presenting a comprehensive coverage of financial topics. Finally, despite not specific to the economics and finance domains, Cajueiro *et al.* (2023) provide a broad list of annotated data sources for automatic text summarization, which usually consist of text and one or more summaries. Their list includes many news data, such as DUC (Over; Dang; Harman, 2007) and CNN/Daily Mail (Hermann *et al.*, 2015).

2.5.3 Preprocessing

Before closing our discussions about textual data, we must not forget an important part of working with unstructured data such as text: preprocessing. This step is performed prior to feeding NLP models with data and assures some of its noise is removed, besides meeting some of the models' requirements, such as fitting sequence length restrictions. The main preprocessing steps are tokenization, stop-word removal, special character and punctuation removal, lowercase conversion, stemming and substituting specific terms, such as full url

[26] https://huggingface.co/datasets/zeroshot/twitter-financial-news-topic.

paths with the term "url". Not all of these steps are performed to every model and there can be additional steps for specific tasks or models[27].

As examples of preprocessing steps, tokenization is the process of converting text into smaller regular parts, called tokens, so that NLP models can more easily recognize, assign meaning and relate them to other terms. Stop-word removal consists of removing word that are common in documents, but which carry no intrinsic meaning, such as "a" and "the". And stemming consists of reducing inflected words to their base form, such as "finance", "financing" and "financial", that become "financ".

2.6 Applications in Economics and Finance

We will now assess how surveyed works put all the previously discussed pieces together to tackle a handful of research questions. Applications such as stock market predictions, monetary policy assessment and many more will be addressed in this section.

The first application we will discuss is one of the most prolific in this field: financial markets prediction (Xing; Cambria; Welsch, 2018; Farimani; Jahan; Fard, 2022). This branch of the literature usually adopts text classification techniques, with the majority of works resorting to sentiment analysis approaches. Bollen, Mao, and Zeng (2011) use a sentiment analysis tool based on lexicons to investigate whether a large-scale collection of tweets can track public mood state. They find that this textual data can indeed track public mood, however with not all of the mood dimensions identified having significant predictive power over the stock market. Daniel, Neves, and Horta (2017) propose a dictionary-based sentiment analysis application and, along with some other available tools to perform this task, evaluate the influence of Twitter on financial markets.

Furthermore, features extracted from textual data are often combined with other sources of information to perform forecasts. As a direct example, Li, Wu, and Wang (2020) combine stock price data trough technical indicators and textual news data trough sentiment analysis, derived from dictionary approaches, to predict stock market trends. In a similar fashion, Jin, Yang, and Liu (2020) perform stock market prediction by combining sentiment analysis, extracted from social media text data through CNN (Kim, 2014), with market data. In order to perform price predictions, they resort to an attention mechanism enhanced LSTM.

Kumbure *et al.* (2022) perform a broad review on machine learning for stock market prediction. They find a growing use of deep learning techniques and textual data in recent articles. Similarly, Passalis *et al.* (2022) examine a handful of deep learning models to extract sentiment from various textual sources, including news stories, in order to evaluate trading

[27] For a more in depth discussion of data preprocessing in NLP, see Chai (2023).

strategy performance for Bitcoin. In the same direction, Wang, Yuan, and Wang (2020) analyse articles from Seeking Alpha, the online crowd-sourced content service provider for financial markets, to assess the performance of a deep learning model, namely LSTM, against traditional machine learning approaches (Naive Bayes, SVM, Logistic Regression and XGBoost). Their results show the LSTM approach outperforming the other models.

Once in the realm of NLP applied to financial markets prediction, news data and social media content are among the most widely used types of textual data, as we have seen before. Agarwal (2019) perform an extensive exploratory review on the impact of the diffusion of information available on the Internet — from online news articles to social media platforms — over stock markets and investor behaviour, surveying works from 1992 to 2017. On an applied work, Souma, Vodenska, and Aoyama (2019) use sentiment analysis over news stories from Thomson Reuters to predict financial markets' sentiment. Through deep learning techniques, they extract sentiment polarity, whether positive, neutral or negative, based on stock returns right after news publications. Moreover, Calomiris and Mamaysky (2019) assess the impact of news over stock market risk and return for large set of countries and Chen *et al.* (2022) resort to attention-based encoders to perform stock market predictions from financial news. Wang *et al.* (2015) test classification machine learning models over social media content to extract sentiment. They find SVM classifiers to outperform Naive Bayes and Decision Trees.

Given the growing interest of NLP applications to economics and financial research questions, a branch of the literature specialized on assessing the performance of NLP methods to such applications. These works, usually in the intersection between economic and computer science literatures, provide researchers with valuable reference for model selection, analysing their performance through a broad range of evaluation metrics. As an example, in a review of machine learning for stock market prediction, Kumbure *et al.* (2022) consider metrics among four categories, which they name (i) accuracy-based, (ii) error-based, (iii) return-based and (iv) statistical tests. They find that root mean square error (RMSE), mean absolute percentage error (MAPE), mean absolute error (MAE) and mean square error (MSE) are the most common among the error-based metrics and the Sharpe Ratio is the most frequent among return-based metrics. Accuracy as an evaluation metric stands in the first place among the accuracy-based metrics in terms of literature adoption, followed by hit ratio, applied to classification-based prediction works. Finally, they find that a paired t-test was the most widely adopted metric among the so called statistic tests category.

It is important to note, though, that how NLP models are evaluated, and the after all selection of which model is more adequate, depends on the NLP task as hand — this is one of the reason for our choice of how structuring this paper, by the way. For the most popular setup for sentiment analysis, as binary classification problem, metrics such as F1 score are widely adopted. F1 score is the harmonic mean between precision, measuring how accurate the model was in its positive outputs, and recall, measuring how sensitive the model was to

detecting positive items in the dataset. As Saha, Gao, and Gerlach (2022) write:

$$F1\text{-}score = \frac{2 \times Precison \times Recall}{Precision + Recall}$$

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN},$$

with TP for true positives, FP for false positives and FN for false negatives.

Another common metric used to evaluate binary classification performance is Matthews Correlation Coeficient (MCC). Is has the advantage of considering true and false positives and negatives and can be used to evaluate performance with classes of different sample sizes. This is the main evaluation metric Mishev *et al.* (2020) adopt in their comparison of a broad range of financial sentiment analysis models. It is calculated as:

$$MCC = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP + FP)(FN + TN)(FP + TN)(TP + FN)}},$$

with TP for true positives, TN for true negatives, FP for false positives and FN for false negatives.

In the just mentioned work, Mishev *et al.* (2020) find transformer models showing superior performance against other machine learning and deep learning approaches, with results comparable to expert's opinions. Facing that result, they suggest that the main recent progress in sentiment analysis is driven by improvements in representation methods, which feed the semantic meaning of the words and sentences into the models. Furthermore, Mishev *et al.* (2020) show that distilled versions of NLP transformers, such as DistilBERT (Sanh *et al.*, 2020), can be efficient alternatives to their large counterparts. They argue that these models stand as relative light-weight and cost-saving models, suitable for production environments.

Additionally, Farimani, Jahan, and Fard (2022) study the literature in the intersection of behavioral economics and the NLP field. They conduct a broad survey on text representation models, from bag of words approaches to word embedding techniques, and how previous papers combine such text representation with financial market data to perform stock market predictions. Among their finding, a growing adoption of word embedding in this branch of the literature, especially after Vaswani *et al.* (2017) and the spread of transformer-based models, with BERT (Devlin *et al.*, 2019) and its variants as the most popular ones (Araci, 2019; Chen *et al.*, 2019; Mishev *et al.*, 2020; Zhao *et al.*, 2021; Shapiro; Sudhof; Wilson, 2022).

Attention-based techniques applied to financial prediction have been present in research papers even when not through transformer networks. Liu *et al.* (2018) use a two-level attention mechanism to measure the importance of each word and each sentence in news'

titles and contents, in order to predict stock price movements; Liu *et al.* (2022) resort to an attention mechanism and explicitly model events and noise over a fundamental stock value state for news-driven stock movement prediction; and Yang *et al.* (2018) use dual-stage attention mechanism, first to identify influential financial news for stock price prediction and second to allocate different weights to different days in terms of their contribution to stock price movement. They then turn to these attentions' results, adding to knowledge graphs, to address the problem of explainability, a key issue to deep learning models. Furthermore, with the challenge of understanding the underlying financial fundamentals that drive model's outputs in mind, Deng *et al.* (2019) argue that, despite having achieved promising results in stock trend prediction, most of deep neural network models are not interpretable for humans. Thus, they propose an approach that combines knowledge graphs from financial news data with convolutional neural networks to predict stock price trends. Similarly, Cheng *et al.* (2022), Kim *et al.* (2019), with distinct models, also highlight the interpretability of combined knowledge graph and attention mechanism approaches to predict financial time series.

In an approach resembling those of knowledge graphs, a branch of the literature also combines other graph-based techniques with textual data. Gao *et al.* (2022) use an attention-based graph approach and combine topics extracted from companies' descriptive documents and historical returns. In a similar fashion, building from a complex network background, Cajueiro *et al.* (2021) develop a framework to measure similarity between news stories of different companies in order to model indirect contagion between them. Besides the two main application of their method, namely measuring the risk of bad news pass through from one company to another and providing additional input for central banks to deal with indirect contagion, their approach might also be used by portfolio managers to add another dimension to diversification, since the news similarity network they propose capture links between companies not usually captured by traditional correlation networks, but that can significantly affect investment performance.

Other research area that leverages NLP techniques to analyse highly relevant economic textual data is the one targeted at monetary policy. This application of language processing in economics is actually manifold, ranging from future policy decision forecast to financial markets' impacts of monetary policy and even to inferring implied policy objectives, hidden on monetary authorities' manifestations. As an early example, Boukus and Rosenberg (2006) perform topic modeling on the Federal Open Market Committee's communication through the lens of Latent Semantic Analysis (LSA). In a different direction, Chague *et al.* (2015) use a dictionary approach combined with dimensionality reduction to extract central bank sentiment and assess its impact over the Brazilian term structure of interest rates and Shapiro and Wilson (2022) use NLP to extract central banks' objective function from their communication, including implicit inflation target.

Applications of NLP in the field of monetary policy has evolved with the developments of new language processing methods. Petropoulos and Siakoulis (2021) use a dictionary-based approach and combine features extracted from central bank communication with machine learning to predict financial markets' behaviour. They find that central bankers expectations have predictive power over financial market turbulence. Moreover, entering the era of transformer-based models, Niţoi, Pochea, and Radu (2023) propose a fine-tuned BERT model to analyse sentiment of central bank communication, which outperforms commonly used dictionary-based approaches in anticipating policy rate changes.

Other research topic that has been gaining attention recently involves understanding the role narratives play in economic fluctuations (Shiller, 2017). It is a problem usually tackled by topic modeling techniques when addressed through the NLP toolbox (Roos; Reccius, 2023). In an assessment of news-driven busyness cycle, Larsen and Thorsrud (2019) use LDA to extract latent yet interpretable topics from news data. They find that many of these topics have predictive power over economic variables, including asset prices. Also applying LDA, but this time to answers of an open-ended questionnaire, Borup *et al.* (2023) find narratives (in the form of topics) containing novel information not present in other sources of text data.

In another very interesting NLP application, Baker, Bloom, and Davis (2016) use a word list based news search to assess economic policy uncertainty. They construct three groups of terms, relating to economy, policy and uncertainty, and measure the frequency of news articles containing terms from all of the three groups over the number of articles in the same newspaper and month.

Despite not a direct economic application, but certainly one of economic relevance, some works aim to analyse political tone in textual data. Gentzkow and Shapiro (2010) assess political slant in news articles by building a dictionary of terms used by either Democratic or Republican parties' Congress members in speeches. In a similar fashion Figueredo, Mueller, and Cajueiro (2022) analyse speeches by Brazilian Senators from 1995 to 2018. However, they focus on political polarization issues, leveraging classification accuracy as a means to measure it, with a supervised classification approach on top of tf-idf text representation for feature extraction. They find that polarization has indeed increased since the 2003-2007 legislature.

Finally, other noteworthy applications range from mutual funds industry (Solomon; Soltes; Sosyura, 2014; Hillert; Niessen-Ruenzi; Ruenzi, 2014) to M&A and IPO analysis (Ahern; Sosyura, 2014; Ferris; Hao; Liao, 2013; Loughran; McDonald, 2013) and even to to fraud detection (Purda; Skillicorn, 2015).

2.7 Concluding Remarks and Further Research

In this paper, we conducted a survey of natural language processing applied to economics and finance. Starting by drawing a map of the literature by the means of bibliometric tools, we outlined how this interdisciplinary field shapes itself, providing insights on the recurrent patterns through which NLP tasks, models and data are combined to contribute with actual applications to our field of interest. The literature mapping exercise also sheds light on the fast-paced evolution of natural language processing techniques in recent years, and on how its adoptions in economic and financial research has adapted to such advances.

At the end, the insights from the bibliometric exploratory analysis also shaped the structure of our work. Building from the interactions of NLP tasks, models and data, we were guided by questions formulated to address each of these features individually, then combining them to discuss applications themselves. Also adding to the findings from literature visualization, we found this as the most straightforward path for an introduction of the subject to the uninitiated economic and financial researchers, hopefully disseminating NLP techniques and shortening the path to aggregating more practitioners to it in order to leverage the richness of information in the form of textual data to the field.

Finally, the applications we discussed are manifold. Nonetheless, we can still see plenty of directions for this promising field to expand. As an example, we point to improved dimensions of monetary policy measures extracted from textual data. Despite the applications we mentioned in Section 2.6, we still see a gap in the literature relating the effects of central bank communication to overall yield curve movements, notwithstanding the discussion on the topic in the economic literature (Wright, 2012; Amaral, 2013). Furthermore, issues related to monetary policy international coordination (Taylor, 1985; Taylor, 2013) can also benefit from central bank communication assessments through the lens of NLP.

For another direction, machine learning explainability concerns are already a topic tackled among computer science researchers. This line of works aims to provide interpretable results, driving machine learning — and specially deep learning — away from a black box approach, in which one can't understand the rationale behind model's outputs, and closer to a theory-driven approach. Moreover, we must note that, since the interpretation of the outcomes are highly application-dependent, this is a topic in which we see room for economics and financial researchers to contribute. Some works have been done to assess the explainability of more general machine learning (Carmona; Dwekat; Mardawi, 2022; Chen *et al.*, 2023), with space for specific developments to NLP.

All in all, in light of the recent progress in natural lagnguage processing models, we find appropriate to conclude our work with the recommendation of Li (2010), Loughran and Mcdonald (2016), one that is so relevant nowadays as it was back then: "the literature needs to be less centered on finding ways to apply off-the-shelf textual methods borrowed from

highly evolved technologies in computational linguistics and instead be more motivated by hypothesis closely tied to economic theories" (Loughran; Mcdonald, 2016).

3 Measuring Monetary Policy Coordination from Central Bankers' Speeches

3.1 Introduction

In our globally connected economy, where trade and capital flows interlink nations, it is crucial for monetary authorities to consider international counterparts in policy-making. A key, yet unresolved question emerges: How can we effectively measure international monetary policy coordination, especially considering the wide array of instruments at policymakers' disposal?

In this paper, we provide a novel, forward-looking measure of how central banks implicitly coordinate their actions, accounting for similarities on what policymakers weight on the economic outlook, the instruments they rely on and the forward-guidance they communicate, all extracted from public manifestations in the form of speeches' transcripts. We use the central bankers' speeches database made available by the Bank for International Settlements[1] and we build a network of similarities that connects central banks whose policymakers' present speech similarity. Our method advances the static framework of Cajueiro *et al.* (2021), originally designed to measure indirect contagion from news-based similarities between companies, into a dynamic setup to analyze coordination among central bankers. First, using a long-term time series, we investigate if there are fundamental relations among the countries' monetary policy and, by resorting to a community detection technique (Blondel *et al.*, 2008), we find three clear communities. From this analysis, one of them clearly stands out: the one that contains 9 of 10 central banks overseeing the G10 currencies. Moreover, the missing G10 central bank, the European Central Bank (ECB), is included in a community containing only Euro Zone institutions. When these institutions are left out of the analysis, since they are not responsible for monetary policy, the European group is dissolved and the ECB migrates to the G10 community. This result suggests that this network successfully captures the long-term global importance of that group of institutions. Furthermore, word analysis over the speeches of policymakers in this group of central banks point to evidence of their orthodox approach to monetary policy. Second, we explore how the connections in the speech similarity network has evolved through time. Our findings indicate that coordination tends to increase in times of economic stress, as in the years of the Great Financial Crisis and the period after the Covid-19 pandemic. Additionally, our novel coordination measure is compared to two gauges of coordination extracted from conventional monetary policy instruments, namely policy rates and central bank balance sheet. This analysis shows that the coordination as measured by speech similarity is able to capture increases in both metrics, standing as a more general way to assess whether the global liquidity environment is subject to coordinated policy movements. Finally, we run an evolution analysis on word occurrences in speeches and on how they shed light on coordination drivers through time. We find that our proposed measure is driven by mentions

[1] https://www.bis.org/cbspeeches/.

to policy instruments and economic views proffered by policymakers.

Our paper is mostly related to Armelius *et al.* (2020) since they resemble our current efforts leveraging textual data to analyze international relations between monetary policy across countries. Using natural language processing to analyze monetary policy sentiment spillovers between economies, they suggest that there are regular patterns in central bank communication that are not explained by economic and financial relations. Furthermore, our present paper uses an extended version of the Armelius *et al.* (2020) dataset. However, differently from us their sentiment analysis approach focuses on spillovers among economies, whereas our speech similarity network aims at explicitly proposing an objective measure of coordination extracted from communication data.

Also related to our work is the recent contribution of Can *et al.* (2024). Like ours, their goal was to assess international monetary policy coordination, however diverging in methodology and data. Whereas we use textual data and complex networks analysis methodologies, they use machine learning models and macroeconomic data, such as countries' policy rates, output growth and consumer inflation, among others, to perform an empirical evaluation of how groups of developed and emerging economies integrate their policy decisions. They find that global financial conditions are a key driver of policy decisions, with greater effects over the more financially vulnerable emerging economies, who tend to suffer capital outflows in events of global monetary tightening, which in turn leads them to also increase their policy rates.

An interesting strand in this literature explores how communication between central bank governors may promote monetary policy coordination. In particular a recent study due to Imisiker (2016) examine the role of communication between central bank governors in monetary policy coordination. In order to do that, he examines the impact of governors' attendance at the Global Economy Meetings at the Bank for International Settlements (BIS) over monetary coordination as measured by correlation of cyclical components of money market rates.

Our paper also connects to the research that explores how central bank communication affects economic agents' expectations. In general, we find that clearer communication from central banks enhances monetary policy effectiveness by influencing such expectations. For instance, clear communication can lead to more accurate private sector forecasts for short and long-term interest rates and can decrease both the variation and bias in inflation rate expectations (Swanson, 2006; Neuenkirch, 2012; Jitmaneeroj; Lamla; Wood, 2019). Additionally, recent research by Hansen and McMahon (2016), Labondance and Hubert (2017), Shapiro and Wilson (2022), and Fasolo, Graminho, and Bastos (2022) applies natural language processing to examine the effects of central bank communication on market and economic variables.

Moreover, our work naturally relates to the literature that explores monetary policy

coordination and cooperation among central banks and their effects. While cooperation involves sharing central banking techniques and information, and identifying common issues, coordination usually refers to policy actions formally agreed upon by groups of policymakers, aimed at achieving positive results for the international monetary system as a whole. It is worth mentioning that there is a long and historical discussion about the role of rule-based monetary policy and new tools to establish the implications and benefits of cooperation and coordination between countries (Jensen, 1999; Sutherland, 2004; Liu; Pappa, 2008; Taylor, 2013; Kapur; Mohan, 2014; Figueroa; Padilla, 2022; Bordo, 2021).

Finally, our model leverages natural language processing to identify similarities in policymakers' speeches. From there, our methodology maps the problem of coordination between central bankers into a question of building a complex network of speech similarities. An advantage of this approach is to use techniques specialized in network modeling (Caldarelli; Chessa, 2016; Latora; Nicosia; Russo, 2017) to tackle the issue of international monetary policy coordination. In this context, several studies have explored ways to measure similarity among institutions from different perspectives. Correlation networks, for instance, have been widely used to identify similar behaviors among various players in financial markets (Mantegna, 1999; Bonanno; Vandewalle; Mantegna, 2000; Tabak; Cajueiro; Serra, 2009; Tabak; Serra; Cajueiro, 2010; Getmansky; Lo; Pelizzon, 2012; Cont; Schaanning, 2019). Additionally, networks based on shared board memberships, such as board of directors networks, reveal that companies with overlapping boards often make strategic decisions simultaneously (Davis; Greve, 1997; Battiston; Weisbuch; Bonabeau, 2003). Furthermore, the structure of networks can also influence how participants communicate and replicate behaviors in financial markets, as demonstrated in studies of communication and imitation dynamics (Tedeschi; Iori; Gallegati, 2009; Tedeschi; Iori; Gallegati, 2012).

We structure our work as follows. Section 3.2 focus on the review of the pertinent literature. Section 3.3 describes our data. Section 3.4 presents the methodology employed in our study. Section 3.5 discusses our main findings and Section 3.6 concludes our work.

3.2 Literature review

Our paper stands in the intersection of three economic research streams: the literature on (i) international monetary policy coordination; (ii) on cross border monetary policy spillovers; and (iii) on monetary policy communication.

The current branch of the literature on monetary policy coordination had its dawn with Hamada (1976), followed by Oudiz *et al.* (1984), Canzoneri and Gray (1985) and Canzoneri and Henderson (1991). These works constitute what can be understood as the first generation of international policy coordination models (Canzoneri; Cumby; Diba, 2005). Building from a Mundell-Fleming framework, these Keynesian-tradition approaches model policymakers'

strategic interactions through a game theory perspective. Each country acts according to the maximization of a utility function that penalizes deviations from desired levels of output, inflation and international reserves. These works showed that — even in a framework with floating exchange rates — in the absence of coordination the Nash equilibrium is not Pareto efficient[2]. However, the consensus in such literature is that gains from coordination seemed to be fairly modest at best. This is particularly true for empirical exercises, as the ones provided by Oudiz *et al.* (1984).

This early literature explores various aspects through which policy coordination might result in Pareto improvements over the non-cooperative Nash equilibria. Canzoneri and Gray (1985) highlight that non-cooperative behavior might result in both too expansionary or too contractionary solutions, depending on a set of structural features of the world economy, such as the extent to which wage rates are indexed and how exposed world economies are to the markets affected by exogenous shocks. On the other hand, works such as Canzoneri and Henderson (1991) explore the structure of international monetary policy games, the incentives policymakers have to deviate from cooperative strategies and the mechanisms that can replicate the gains from coordination on non-coordinated behavior. They analyze one-shot and multiperiod finite horizon-games, reflecting policymakers' finite terms of office; two and three country setups for exploring the effects of coalitions in monetary policy coordination; the effects of time inconsistency emerging from the possibility of announcements of central banks' prospective strategies and their incentives to deviate from these strategies when the time come for their implementation; and incomplete information games, where policymakers might not be certain about other countries preferences or strategies.

Despite the arguments for monetary policy coordination, Rogoff (1985) warns that it might even be counterproductive. Whereas international coordination might actually generate better output-inflation tradeoff compared to non-coordinated policies once there is no incentive to manipulate the currency and the terms of trade, there might emerge a credibility problem with the private sector. Rational wage setters realize that the policymakers' incentives to inflate prices are greater in the presence of international policy coordination, leading to higher nominal wage rates. This results in systematically higher inflation when compared to the noncooperative solution.

The second generation of policy coordination models (Canzoneri; Cumby; Diba, 2005), includes the works of Obstfeld and Rogoff (1995), Corsetti and Pesenti (2001), Obstfeld and Rogoff (2002), Devereux and Engel (2003), Benigno and Benigno (2003), Benigno and Benigno (2006), Benigno and Benigno (2008), Clarida, Galı, and Gertler (2002), Canzoneri,

[2] As Oudiz *et al.* (1984) define, the Nash solution in absence of coordination corresponds to the case where each country maximizes its utility function, taking other countries' policies as given. As for the cooperative solution, it corresponds to the case where every policymaker acts as to maximize a collective utility function, defined as a weighted average of each country's own utility function.

Cumby, and Diba (2005). These are New Open Economy Macroeconomics[3] models with microfoundations that incorporate optimizing agents, monopolistic competition and nominal rigidities into dynamic general equilibrium frameworks. In these models, the stabilization goal of monetary policy is usually embedded in households' utility functions through consumption stabilization, providing means to explicit assess coordination effects over welfare.

Tackling the intertemporal limitations of the aggregative Keynesian models, which, among other aspects, fail to describe the impacts of monetary policy over production decisions, the seminal work of Obstfeld and Rogoff (1995) propose a model of international policy transmission provided with microfoundations. Their two-country framework allows the assessment of international welfare spillovers of monetary and fiscal policies. Once again, analyzing the results of unilateral monetary changes, they find the terms of trade and current account effects to be of second-order importance. These results are in line with the previously reported limited gains from policy coordination.

Corsetti and Pesenti (2001) provide a model that captures the macroeconomic spillover effects between two interdependent economies' monetary and fiscal policies while accounting for monopolistic supply in production and monopoly power of a country in trade. They find that expansionary monetary policy can actually reduce national welfare, as it deteriorates domestic consumers' purchase power in global markets. In turn, most of the benefits of such monetary expansion can be perceived by foreigners. This result, that contrasts with the competitive devaluation argument[4], is shown to be more relevant in smaller and more open economies. Corsetti and Pesenti (2005) then build on this framework to analyze optimal monetary policy rules, also investigating the benefits of policy coordination. They find that gains from coordination are possible and depend in a non-monotonic way on the level of exchange rate pass-through to domestic prices. In the extreme cases of perfect or absent pass-through there are no gains to coordination, since there is no policy spillover through the exchange rate channel. However, middle cases do provide some welfare improvement. Nonetheless, they do not further explore the magnitude of such gains.

Following the New Open Economy Macroeconomics tradition on a stochastic microfounded model, Obstfeld and Rogoff (2002) argue that, even in the presence of significant international economic integration from goods and financial markets, the need for policy coordination is again of second-order importance to the macroeconomic stabilization benefit of monetary policy. They analyze rule-based monetary frameworks and find that, as domestic monetary rules improve and international asset markets become more complete, the uncoordinated Nash solution to the monetary policy game converges to the cooperative solution. These policy rules should not target real market imperfections, such as monopoly

[3] See Lane (2001) for a survey on New Open Economy Macroeconomics.

[4] For a deep dive on how Corsetti and Pesenti (2001)'s result contrasts with the competitive devaluation argument, we refer to Corsetti *et al.* (2000).

distortions or capital market imperfections, and instead should be designed to offset nominal rigidities.

Most of this literature merges with research on macroeconomic spillovers across countries, usually focusing on spillovers through the terms of trade channel. However, there are also works that try to understand and measure the different mechanisms through which one country's monetary policy might impact other economies[5]. On a recent contribution, Bianchi and Coulibaly (2024) explore a financial channel that acts through the world real interest rate, making financial integration in global capital markets relevant to their framework. They find that, in the absence of coordination, countries can fail to internalize the impact of their monetary policy decisions over the world real interest rate, affecting other central banks' ability to stabilize output and inflation. The effect of monetary policies over the world real interest rate is also explored by Fornaro and Romei (2022), who conclude that monetary policy coordination can lead to gains in terms of output levels.

Other contributions to the monetary policy spillover research are in Rey (2015), Kalemli-Özcan (2019), Acharya and Pesenti (2024). Rey (2015) analyzes how global financial cycles and capital flows interact with central economies' monetary policy spillovers. Kalemli-Özcan (2019) explores the risk premia channel of US monetary policy spillovers. She shows that emerging market economies are more exposed to this channel, since capital flows are more sensitive to risk perception in such economies. In turn, Acharya and Pesenti (2024) show that spillovers tend to be magnified by the introduction of a precautionary savings channel and a real income channel in a setup with heterogeneous agents.

Motivated by the empirical observation that coordinated monetary policy movements across countries can have results greater than the sum of their expected individual effects, Caldara *et al.* (2024) developed a model to assess globally synchronous monetary tightening, such as the one observed in 2022, when major economies started rising policy rates to fight the inflation pressures that began on the previous year. They argue that monetary policy spillovers are mutually reinforcing, with episodes of synchronous tightening being associated with tighter financial conditions and larger effects on economic activity. Their model's results are based on spillovers through the financial channel, which provides nonlinear amplification when faced with global monetary tightening and affects output and financial conditions with greater magnitude than inflation, thus worsening the monetary policy tradeoff[6]. Their result shed light in the need for accurately measuring the state of coordination of global monetary policy, since the nonlinearities in spillovers depend on such measure to reach adequate policy recommendation.

[5] For discussion and empirical assessment on monetary policy spillovers, we refer to Kearns, Schrimpf, and Xia (2023).

[6] For other recent works that challenge the consensus view that the gains from monetary policy coordination are limited, see Dedola, Karadi, and Lombardo (2013), Bodenstein, Corsetti, and Guerrieri (2024), Fornaro and Romei (2022).

Finally, our work also relates to the literature on communication as an instrument of monetary policy. As a matter of fact, Blinder *et al.* (2008) argue that central bank communication has gained importance to become one of its main tools, used to guide market expectations, and thus move asset prices[7]. They provide a comprehensive survey, going through the evolution of the understanding of the role of communication by both researchers and policymakers, as well as exploring works on its theoretical foundation and empirical evidence, highlighting that the latter approach responds for the majority of papers surveyed. Nonetheless, their discussion on the theoretical background for communication as a tool of monetary policy provides great insight into its importance, showing that it does not depend exclusively on deviations from the rational expectation hypothesis. It also finds support in features such as nonstationarity of the ever-changing economic environment; or in asymmetric information between the public and the policymaker, especially about the policy rules that drive central banks' decisions, which are in turn nonstationary themselves and usually much more complex than simple Taylor Rules (Svensson, 2003).

It is also worthy noting that the empirical literature on central bank communication has recently shifted with the advances in the fields of Machine Learning and Natural Language Processing. As one of the latest examples, in arguably one of the most sophisticated approaches, Gorodnichenko, Pham, and Talavera (2023) develop a deep learning model to extract voice emotions in press conferences after the Federal Open Market Committee policy decision. Controlling for the meeting decision itself and for text sentiment, they find statistical significance in the impact of the measured voice emotions over asset prices. This example illustrates how this branch of the literature is currently developing, building from state of the art methods to provide novel perspectives to questions both recent and long lasting. Our paper positions itself among these works.

3.3 Data

We use the textual data that contains central bankers' speeches available by the Bank for International Settlements (BIS)[8]. It consists of English transcripts or translations of select and policy relevant speeches proffered by central bankers of 119 institutions around the world over the years ranging from 1997 to 2024. In total, there are 18,970 speeches in the entire database. However, in order to guarantee minimum representation and make feasible the evolution analysis, we rule out institutions with less than 100 speeches in the entire database (this represents at least around 3.7 speeches per year on average). After applying this filter, we remain with 16,842 speeches for 41 institutions. This dataset is an extended

[7] Recent works also highlight central bank communication as a means to promote monetary policy accountability and thus enhance its credibility (Blinder *et al.*, 2024; Ehrmann *et al.*, 2024).

[8] https://www.bis.org/cbspeeches/.

version of the one used by Hansson (2021) to assess the evolution of topics in central bankers' speeches and Armelius *et al.* (2020) to study the spillover of central bankers' sentiment across countries. The rationale behind our data choice lays on the fact that this set comprises a broad range of economies, which makes it suitable for coordination assessment, and also an extended time period, making it suitable for the evolution analysis.

To apply the methodology to build the coordination network from our textual speech database, we first have to pre-process the documents. This is done by passing speeches through steps that are usual to the natural language processing field. These steps are: (i) tokenization, which breaks down the text into a list of standardized elements by removing undesired information, such as punctuation, converting words to lowercase and removing accents; (ii) stopword removal, which consists in removing common words that do not add meaning to the speeches, such as "the" and "and"; (iii) stemming, which reduces inflected words to their root form; and (iv) removing rare words, in our case the ones that appear less than 5 times in a speech or in less than 10 speeches in the database.

Moreover, as a means to provide benchmark to our speech similarity policy coordination measure, we also build similarity networks from conventional monetary policy instruments. In order to do so, we use two additional measures, also provided by the BIS: central bank policy rates[9] and central bank total assets[10]. Both of the additional series are filtered to match our speech database in terms of time range and institutions covered.

Central bank policy rates consist of nominal interest rates defined by policymakers as the widespread main instrument to establish the monetary policy stance[11]. This database, with monthly series curated by the BIS for more than 40 advanced and emerging market economies, allows us standardize our assessment for the broad set of countries in our study, which is necessary to measure policy coordination. Notwithstanding, Forbes, Ha, and Kose (2024), who use the same database to study monetary policy cycles throughout a set of advanced economies in a long-term time frame, highlight that, despite being the primary tool for monetary policy, policy rates may have limitations at times. This is specially relevant for times of shifts in policy framework or targets, or for times when other monetary policy instruments may have assumed priority over interest rates. The former case, when monetary frameworks change, despite being incorporated in our proposed speech similarity measure, is not as straightforward to be captured by conventional instruments' assessment[12]. On the other hand, to address the former case, considering alternative instruments, we augment the benchmark analysis by using central bank total assets data. Thus, central bank total assets are used to capture the effects of quantitative easing (QE) and quantitative tightening (QT)

[9] https://data.bis.org/topics/CBPOL.

[10] https://data.bis.org/topics/CBTA.

[11] For the methodology of central bank policy rates data series, see Bank for International Settlements (2024a).

[12] See Ball (2010) for a discussion on monetary policy regimes.

instruments on monetary policy stance[13]. We do so by measuring balance sheet increases (QE) or reductions (QT) from the first difference of the quarterly time series of central bank total assets. This database covers more than 50 advanced and emerging market economies and are referenced in U.S. dollars for every country.

3.4 Methods

We divide this Section in three subsections. In Section 3.4.1 we present how we adapt the method from Cajueiro *et al.* (2021) to build the coordination network in a dynamical framework. In Section 3.4.2, we review the community detection method due to Blondel *et al.* (2008). Finally, in Section 3.4.3 we show the methodology for the coordination networks used as benchmark to our study, built from policy rates and central bank total assets data.

3.4.1 The construction of the speech coordination network

To build our monetary policy coordination network, we must first identify the similarities between the discourses of central bank pairs. They will represent the links in the network. In order to do that, we begin by exploring some concepts from the field of natural language processing. Let N_V be the total number of distinct words. First, we identify each of these words, w_i, by the index i, with $I_V = \{1, \ldots, N_V\}$ representing the set of all indexes. Then, we define the vocabulary $\mathcal{V} = \{w_1, \ldots, w_i, \ldots, w_{N_V}\}$ as the set of all distinct words, from every speech in our database. The speech $s_j = [w_{i_1}, \ldots, w_{i_k}, \ldots, w_{i_{L_j}}]$ is then defined as a sequence of L_j non-unique words, with $1 \leq k \leq L_j$ and $i_k \in I_V$. The set of all speeches is written as $\mathcal{S} = \{s_1, \ldots, s_j, \ldots, s_{N_S}\}$. Finally, the vocabulary of a specific speech s_j is represented by $\mathcal{V}^{s_j}$.

To capture the overall properties of the discourse of a given central bank in a specific time interval, instead focusing on individual speeches, we use the concatenation of all of its speeches during the assessment period. Therefore, making $\mathcal{C}$ as the set of central banks in our database, we define s^k as the concatenation of speeches of central bank k, for $k \in \mathcal{C}$. From hereon we will refer to the concatenation of speeches of central bank k, s^k, as simply the speech of central bank k. Furthermore, we define $\mathcal{S}^C$ as the set of all s^k, with N_C elements, where N_C in the number of central banks.

With that, we build the term-speech matrix $\mathbf{M}$ as a $N_V \times N_C$ matrix with the frequency of words that occur in the collection of speeches associated with each central bank in a given period of time. In this matrix, rows correspond to words w_i for $i \in I_V$ and columns

[13] For the methodology of central bank total asset data series, see Bank for International Settlements (2024b).

correspond to the speeches s^k for $k \in C$:

$$\begin{array}{c} \\ w_1 \\ w_2 \\ \vdots \\ w_{N_V} \end{array} \begin{array}{c} \begin{array}{cccc} s^1 & s^2 & & s^{N_C} \end{array} \\ \begin{bmatrix} n_{11} & n_{12} & \cdots & n_{1N_C} \\ n_{21} & n_{22} & \cdots & n_{2N_C} \\ \vdots & \vdots & \cdots & \vdots \\ n_{N_V 1} & n_{N_V 2} & \cdots & n_{N_V N_C} \end{bmatrix} \end{array}, \tag{3.1}$$

where n_{ik} represents the number of times word w_i appears in the speech s^k.

Moving to the coordination network itself, which assumes the form of a speech similarity network, it has central banks as nodes and the links between them defined as a function of the probability of association of their speeches. This way, the coordination between the speeches of a central bank k and another central bank l depends explicitly on the words that appear in the speeches of both central banks. One simple way to summarize the information about the words that appear simultaneously in speeches of two different central banks k,$l \in C$ is given by the function

$$q_{k,l} = \cos(\theta_{k,l}) = \frac{\sum_{i=1}^{N_V} f^k(w_i) f^l(w_i)}{||f^k||\ ||f^l||}, \tag{3.2}$$

where $\theta_{k,l}$ is the angle between the vectors f^k and f^l, for k,$l \in C$, and $f^c(\cdot)$ represents a measure of the importance of each word $w_i \in \mathcal{V}_W$ in s^c. Note that when $f^k, f^l \geq 0$, which is the case for our methodology as we will see ahead, we have $\theta \in [0, \pi/2]$.

The field of natural language processing proposes several different forms for the function f^c. Our methodology applies the so-called Entropy Model. This model provides a similarity evaluation closer to the one that results from actual people's perception (Pincombe, 2004). The Entropy model measures the importance of the word i in the speech of central bank k, by the normalized

$$f^k(w_i) = \omega_{\text{local}}(i,k) \times \omega_{\text{global}}(i), \tag{3.3}$$

where

$$\omega_{\text{local}}(i,k) = \log_2(\phi_{ik} + 1) \tag{3.4}$$

and

$$\omega_{\text{global}}(i) = 1 + \frac{\sum_{k=1}^{N_C} p_{ik} \log_2 p_{ik}}{1 + \log_2 N_C}, \tag{3.5}$$

with $\phi_{ik} = \frac{n_{ik}}{\sum_{k=1}^{N_V} n_{ik}}$, $p_{ik} = \frac{n_{ik}}{\sum_{k=1}^{N_C} n_{ik}}$ and the definition of $0 \log_2(0) = 0$ that is consistent with the $\lim_{t \to 0} t \log_2 t = 0$.

On the one hand, the local weight ω_{local} measures the importance of a word inside a speech as a function of its frequency. On the other hand, the global weight ω_{global} measures

the importance of a word in a speech when compared to its presence in all other speeches of the sample. Basically, it is a measure of how exclusive is the presence of a word in a speech. Thus, a word that appears simultaneously in several speeches of the sample is less important than a word that appears specifically in few speeches.

Finally, for the identification of relevant connections in the speech similarity network, we use to the randomization algorithm described by the following steps:

1. Randomly choose a given token frequency $\overline{n}$;
2. Randomly choose two central banks l and k;
3. If both central banks have tokens with the frequency $\overline{n}$, randomly choose one token with frequency $\overline{n}$ from each central bank, namely w_l and w_k;
4. If token w_l does not belong to the speeches of central banks k (s^k) and the token w_k does not belong to the speeches of central bank l (s^l), exchange these tokens. Note that this process does not modify the token frequency distribution of the speeches of each central bank presented in Eq. (3.1).

This randomization algorithm is proposed by Cajueiro *et al.* (2021) to rule out links that may appear due to the presence of random words that are common in the speeches of different central banks. In order to determine the random connections, with certain degree of confidence, the algorithm permutes words between speeches of different central banks to generate random speeches with the same $n_{i,k}$ (thus the same ϕ_{ik} and p_{ik}) for $i \in I_V$ and $k \in C$. The number of occurrences of a term i in the speech of central bank k, n_{ik}, is the main variable to determine the importance of a word given by Eq. (3.3) using Eqs. (3.4) and (3.5). In our case, if we do not constrain the randomization algorithm to keep the number of different terms constant, we might treat central banks in very different monetary frameworks or different moments in the economic cycle the same way. It is important to note that different monetary policy frameworks are usually related to different sets of words. This is also true for different policy goals, that may vary according to the moment of a country in the macroeconomic cycle. Therefore, some central banks may have to rely on a greater number of terms to communicate their complex monetary frameworks or goals, whereas others may only need a limited number of words to do so[14]. All in all, the adopted randomization algorithm results in speeches with the same characteristics as the original, but using random words from the complete vocabulary.

In our work, we apply this method both for the long-term dataset of speeches and for specific periods to measure the coordination between the monetary policy of different

[14] For instance, a monetary authority pursuing a dual mandate of price stability and full employment may have to use a greater number of terms to communicate these different dimensions than a central bank in an implicit inflation targeting framework. Or a central bank of an economy going through a period of inflationary pressures will probably need more terms to communicate its strategy than a monetary authority with a neutral policy stance.

central banks. Since monetary policy is highly dependable on economic cycle, besides being susceptible to both domestic and international shocks which can shape the objectives and instruments of monetary authorities, we must account for time varying similarities. In order to incorporate that aspect, we run the described methodology in a dynamic setup, by segmenting our analysis in disjoint shorter time frames. This allows us not only to infer monetary coordination with improved precision, but also to assess how it has evolved through time.

3.4.2 The method to identify the communities

In our work, we use the Louvain method for community detection (Blondel *et al.*, 2008). It is an algorithm that identifies non-overlapping groups of nodes in a network by optimizing modularity. Modularity, introduced by Newman and Girvan (2004), is a measure used in network theory to quantify the strength of division of a network into communities, defined as:

$$Q = \frac{1}{2m} \sum_{ij} \left[A_{ij} - \frac{k_i k_j}{2m} \right] \delta(c_i, c_j),$$

where A_{ij} is the weight of the edge between nodes i and j; k_i and k_j are the sums of the weights of the edges attached to nodes i and j; m is the sum of all of the edge weights in the network; and $\delta(c_i, c_j)$ is a delta function that equals 1 if nodes i and j are in the same community and 0 otherwise.

The Louvain algorithm operates iteratively in two main phases: local modularity optimization and graph aggregation. Phase 1, local modularity optimization, is initialized with each node in the network assigned to its own community. The algorithm then evaluates the modularity change that would occur if one node was moved to a neighboring community. If moving the node results in a positive change in modularity, the node is reassigned to the community that produces the largest modularity increase. This process is then repeated for all nodes until no further improvement is possible. At this point, the network reaches a local maximum in modularity, with nodes grouped into optimal communities. The algorithm then moves to phase 2, graph aggregation. Here, the communities identified in phase 1 are treated as single nodes, and a new graph is constructed. The edges between nodes in different communities are aggregated into the new graph's edges, with their weights representing the total weight of the edges between the original nodes. The resulting graph is thus a simpler representation of the original network, with nodes corresponding to entire communities. It is also analogous to the initialized network in phase 1, with each of the new nodes assigned to its own community. The algorithm then returns to phase 1, repeating the two phases on the aggregated graph and further optimizing modularity. This iterative process continues

until no significant modularity improvement is possible. The Louvain method is efficient and scalable, making it well-suited for analyzing large networks.

3.4.3 The conventional policy instruments benchmark networks

As mentioned before, we use two additional measures to build alternative policy coordination networks from conventional monetary policy instruments: policy interest rates and central bank balance sheet.

Despite standing as the main instrument in policymakers' toolbox, measuring monetary policy stance from interest rates is not as straightforward when we are working with panel data with many different economies for a long time frame. Simply using the first difference of the time series does not account for the magnitude of monetary easing or tightening, as it represents a better reference for policy cycles than for its stance. Alternatively, using the rate's level does not provide enough information as well, as the same level of interest rates can mean much different stances for different countries. Indeed, the degree of easing or contraction in policy rates depend upon the country-specific monetary policy framework and upon many different economic variables, with the most relevant being the unobservable neutral interest rate. As a matter of fact, the neutral interest rate is itself subject of discussion in the economic literature, with works dedicating special attention to the challenges of its estimation (Jordà; Taylor, 2019; Kiley, 2020; Ferreira; Shousha, 2023). Moreover, also as discussed in the referred works, the neutral interest rate for a given country is not constant in time, meanings that its estimation should also be constantly updated throughout the studied time frame for it to be an adequate reference for policy stance. This adds to the challenge, as the updates usually require incorporating real-time estimations of the output gap, another unobservable variable with its own problems for estimation (Orphanides; Norden, 2002).

Therefore, in order to account for the country-specific time-varying nature of the reference for monetary policy stance, while not incurring in unnecessary added complexity to our benchmark measures of coordination, we standardize the approach across countries by comparing the current level of policy rates to their long-term average. Precisely, we calculate the percentage deviation from the long-term average:

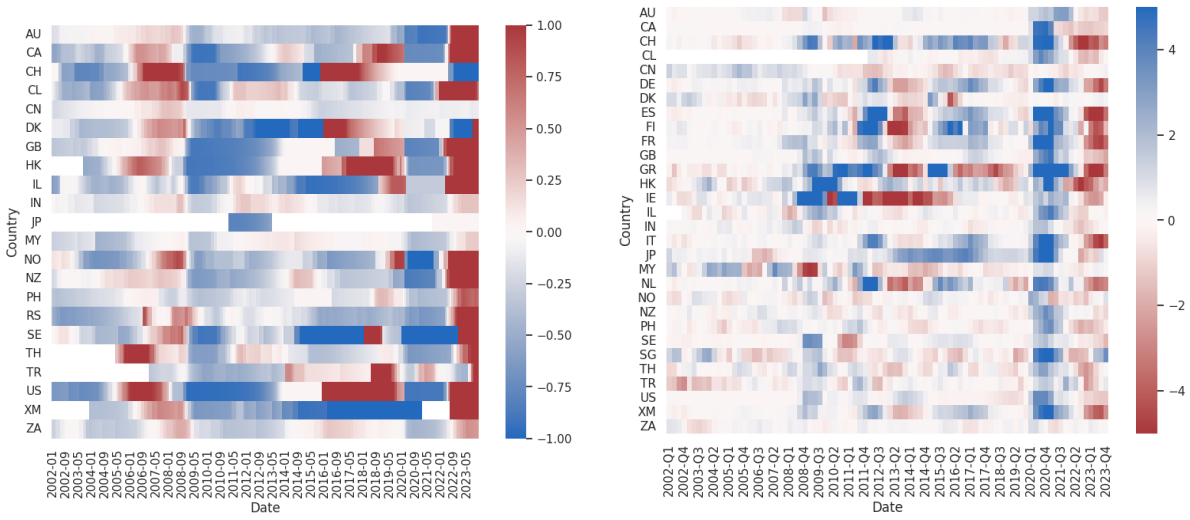
$$S_i(i_{k,t}) = \frac{i_{k,t} - \overline{i_{k,t}}}{\overline{i_{k,t}}}, \tag{3.6}$$

where $S_i(i_{k,t})$ represents the policy rate stance for central bank k in period t, $i_{k,t}$ represents the policy interest rate for central bank k in time t and $\overline{i_{k,t}}$ stands for its long-term average, given by:

$$\overline{i_{k,t}} = \frac{\sum_{\tau=1}^{N} i_{k,t-\tau}}{N}. \tag{3.7}$$

In the above equation, N calibrates the time frame in which we calculate the mean for the interest rate. Fundamentally, it should match the average monetary policy cycle, assuming that cycles usually contain periods of deviation from the neutral rate, with above periods representing monetary contractions whereas below periods represent easing monetary stance. It is a clear simplification, which works for our parsimonious benchmark assessment. In this paper we use $N = 60$ monthly periods.

To measure QE/QT effects over monetary stance, we begin by taking the first difference of the central bank total asset time series. This gives us the amount by which the balance sheet increased, in times of quantitative easing, or decreased, in times of quantitative tightening, always measured in U.S. dollars for enabling comparison between countries. We adopted this measure instead of percentage changes, as we want the magnitude of the policy instrument to be independent of the magnitude of the balance sheet in the first place. Moreover, as a means to smooth the policy stance measure, accounting for times when central bank asset purchasing or selling programs have skipped periods, we take the four-quarters moving average. Figure 3.1 resumes the evolution by country for both of our benchmark monetary policy stance measures.



(a) Policy rate stance.

(b) QE/QT stance.

Figure 3.1 – Benchmark monetary policy measures.

We then proceed to the construction of the two benchmark coordination networks, each from one of the just discussed measures of monetary policy stance. In order to do so, we begin by separating the entire period in smaller time frames, as to match the frames in the speech coordination networks. This step is relevant, since we want the new networks to be benchmarks in the evolution analysis presented in Section 3.5. After that, we build the networks from the correlation of the policy stance between country pairs in each time frame. The rationale for our choice for correlation as the measure of the connection between

countries relates to the filter of statistically significant edges in the networks. Correlation provides a straightforward t-test for significance assessment, therefore eliminating the need for building a randomization algorithm suited for the new measures[15]. We thus conclude the network construction by filtering relevant connections based on a 5% significance test. The formula for the correlation t-test is

$$t_{\rho_{x,y}} = \frac{\rho_{x,y}\sqrt{n-2}}{\sqrt{1-\rho_{x,y}^2}}, \tag{3.8}$$

where $\rho_{x,y}$ is the Pearson correlation coefficient between policy stances x and y, which in turn represent the nodes of the benchmark networks, and n is the number of observations within the time frame used for assessing policy coordination.

3.5 Results

As we argued before, many factors may affect the speech similarity used as input in our assessment of policy coordination. These factors are present in different dimensions of central bankers' communications. On the one hand, some of them are inherent to economic fundamentals, defined mainly by policy framework and institutions, and relate, for instance, to monetary policy mandates and the required accountability and transparency imposed on policy decisions. On the other hand, some result from economic cycles, depending on whether economies are in synchronous stages or face the same shocks, leading to coordinated shifts in policy stance and, therefore, possibly similar communication of central bankers' strategies. Despite the latter being responsible for most of the coordination we are interested in, it is also important to understand the fundamental characteristics that may influence policy coordination. The next subsections explore each of these dimensions.

3.5.1 Long-term speech similarity

In order to understand structural characteristics that drive speech similarities, we begin our analysis by running the proposed methodology on a long-term time frame so we can analyze persistent patterns in central bankers' communications. The time window in this first study ranges from 2002 to 2023 and include every institution that meets the minimum number of speeches in the database as described in Section 3.3.

Figure 3.2 shows the long-term speech similarity network representation. It is a fully connected network with link strength between central banks given by speech similarities according to Eq. (3.2). We also applied the Louvain community detection algorithm (Blondel

[15] Note that the randomizing word occurrence algorithm used in speech similarity network does not apply to the additional monetary policy stance measures.

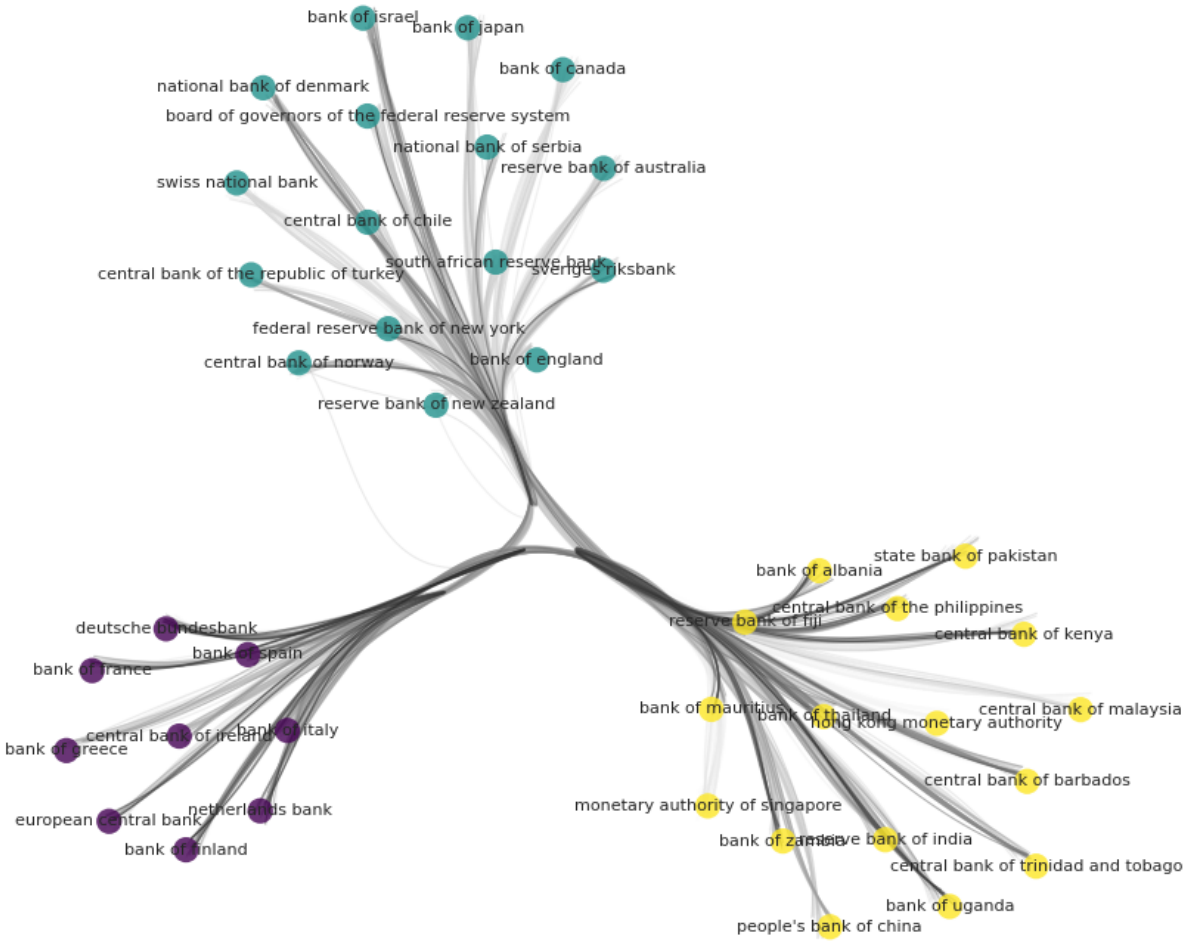


Figure 3.2 – Network representation for long-term speech similarity, from 2002 to 2023.

et al., 2008) described in Section 3.4.2 to understand how central banks grouped in the period of our study. In this analysis one group stands out: a community that comprises most of the central banks overseeing G10 currencies[16]. As a matter of fact, 9 of 10 G10 currencies are represented in this community, with the exception being the European Central Bank (ECB). However, we can see that the ECB is the main central bank in the smallest of the communities in Figure 3.2, which includes most of the Eurozone's central banks in the dataset. This is not surprising, as references to the Euro and to the monetary union are important drivers of the similarities of speeches between these central banks, as we will show ahead. It is important to note, though, that these institutions are not responsible for their monetary policy, as the ECB is the monetary authority for the Euro area. We therefore run a complementary network construction ruling out all of the central banks of the Eurozone but the ECB. The result, presented in Figure 3.3, shows that the number of identified communities dropped to

[16] G10 currencies are the most actively traded and liquid currencies, responsible for the large majority of turnover in global foreign exchange markets (Greenwood-Nimmo; Nguyen; Rafferty, 2016).

two, with the ECB migrating to the G10 community, leaving it with all of the central banks responsible for the main currencies in global trades. This result suggests our coordination measure is successful in capturing the main fundamental traits of institutions in the long-term similarity network. Moreover, the two groups in Figure 3.3 might suggest a leader-follower setup with support from both the theoretical and empirical economic literature (Figueroa; Padilla, 2022).

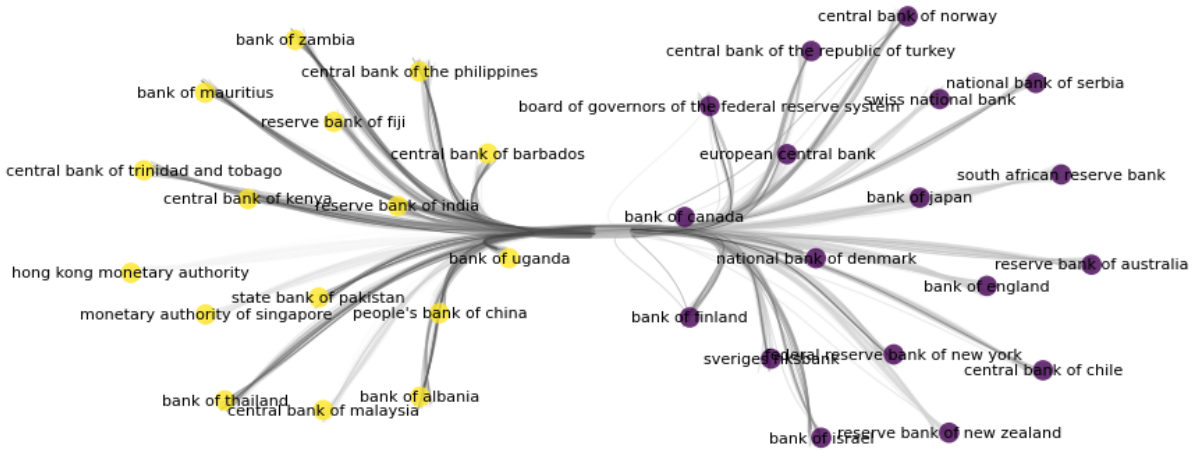


Figure 3.3 – Network representation for long-term speech similarity, from 2002 to 2023, without Eurozone central banks.

To understand what determines the links in the speech similarity network depicted in Figure 3.2, we must analyze the words that drive the connections between central bank speeches. Figure 3.4 shows how words connect the most relevant central banks in the European cluster illustrated in Figure 3.2. This figure has central banks on the left-hand side, with different colors attributed to each of them. The right-hand side has the most relevant terms in their speeches. Relevance is defined by Eq. (3.3), which determines how large are the links between left and right-hand sides. In other words, given central bank k and word w_i, the size of the link in Figure 3.4 is set by $f^k(w_i)$. Therefore, the overall relevance of each word for the subset of central banks illustrated is represented by the height of the rectangle associated to it. This figure thus shows that among the most relevant terms to the speeches of these central banks are *euro*, *area* and *european*. Despite having multiple degrees of relevance to each institution, we can see from the links of different colors arriving at their respective rectangle that the mentioned words are relevant to every one of the central banks represented here.

We can also extend the word analysis to a set of terms that should be able to capture the long-term relations we expect to see in this network. These should be terms related to policy frameworks adopted in each country. To do that, we focused on three of the most relevant central banks for the global economy, each of which adopting a different monetary policy

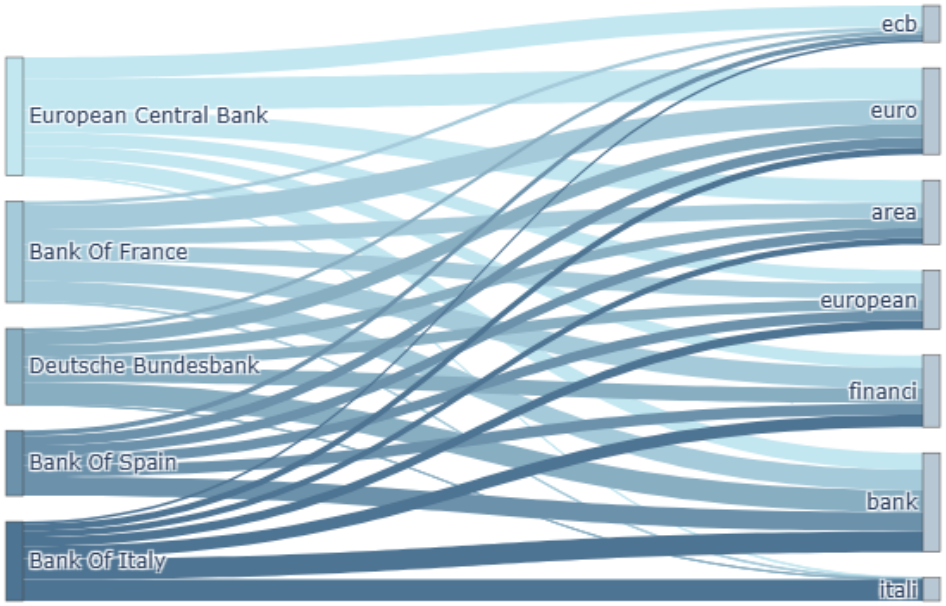


Figure 3.4 – Long-term word relations for European central banks.

framework: the Federal Reserve System of the United States, identified in the database as the Board of Governors of the Federal Reserve System, the European Central Bank and the Bank of England. The selected terms for this analysis are *price*, *stabil*, *inflat*, *labour*, *labor*, *employ*, *unemploy* and *target*[17]. Figure 3.5 illustrates the assessment, and should be interpreted the same way as Figure 3.4, with the only difference being that the set of terms are exogenously provided in Figure 3.5, whereas Figure 3.4 uses terms endogenously extracted from the most relevant occurrences for the analyzed central banks.

Figure 3.5 thus provides interesting results. First, we can see that the most relevant terms for the three most influential central banks are *price*, *stabil* and *inflat*, reinforcing the orthodox approach to monetary policy by those institutions, which is in fact a common feature for the entire G10 cluster depicted in Figure 3.3. The relevance of these terms show that inflation and price stability are the priority long-term goals for this set of central banks. Noteworthy, the term *stabil* also appears related to financial stability in speeches, without

[17] One should note that these are the resulting terms after the pre-processing step described in Section 3.3. Therefore, they should be understood as representatives of mentions to broad sets of words that capture specific concepts. For example, *stabil* represents the concept of "stability", also including "stabilization" and "stabilizing"; the term *inflat* summarizes the mentions to "inflation", "inflationary" and "inflating"; and *unemploy* stands for "unemployment" or "unemployed".

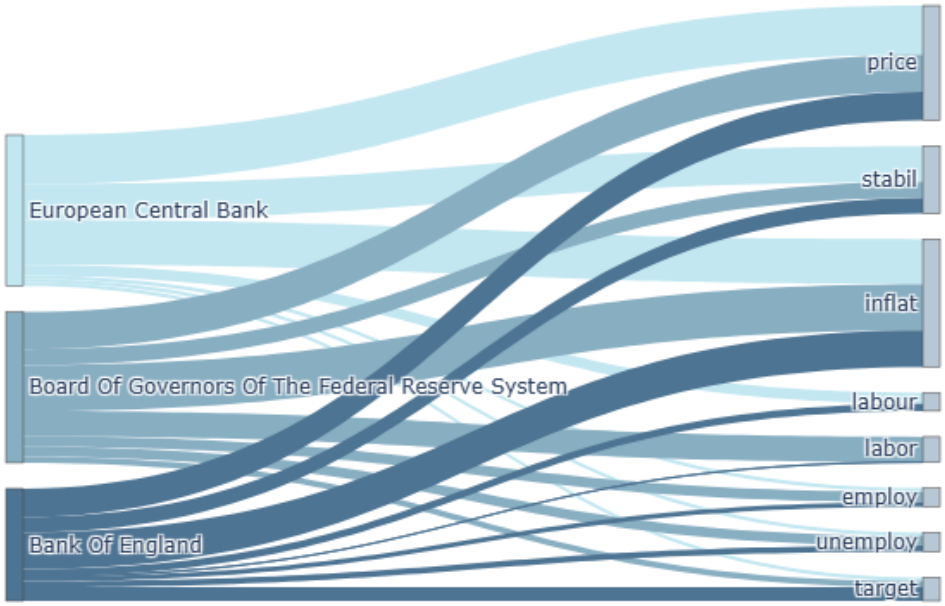


Figure 3.5 – Long-term word relations for monetary policy framework assessment of main central banks.

prejudice for the previous conclusion. Furthermore, we can see that the terms *labor*, *employ* and *unemploy* are much more relevant in the speeches of the American monetary authority when compared to the other two. This captures the dual mandate that the Federal Reserve System (Fed) pursues, of price stability and full employment. The dual mandate is specific to the Fed — whereas the ECB and the Bank of England (BoE) are guided by inflation targeting frameworks — and explains the increased relevance of related terms in speeches proffered by the Fed's governors. Again noteworthy, we can see from these terms that alternative spellings for the same word are present in speeches of different central banks, depending on their countries. We can see that *labor*, the American English version, has great relevance for the Fed's speeches, whereas *labour*, the British English version, is not even mentioned. Likewise, *labour* is present in the speeches of the ECB and the BoE, but with much smaller relative importance than the one for their American peer. This highlights that a complete analysis should include both spellings to lead to correct conclusions. Finally, we can see the term *target* is much more relevant to the BoE in this set of central banks. One should recall that the Bank of England was one of the pioneers in adopting an explicit inflation targeting framework, a trait that this term captures by presenting increased relevance in the communication of the British monetary authority.

3.5.2 Policy coordination in speech similarity through time

Our strategy to analyze the evolution of policy coordination in central bankers' speeches is to run the methodology described on Section 3.4.1 on disjoint time-windows that span the studied interval. In practice, we use two-year windows to run the evolution analysis from 2002 to 2023. This is a period when communication as an instrument in policymakers' toolbox has achieved a certain degree of maturity on most of the advanced economies. Furthermore, the two-year windows are enough for us to have a sufficient number of speeches for most of central banks in the database and are not long enough to include a broad range of stages of policy cycles, as this length matches roughly the average time it takes for central banks to shift their policy stance and perceive the complete effects of monetary policy.

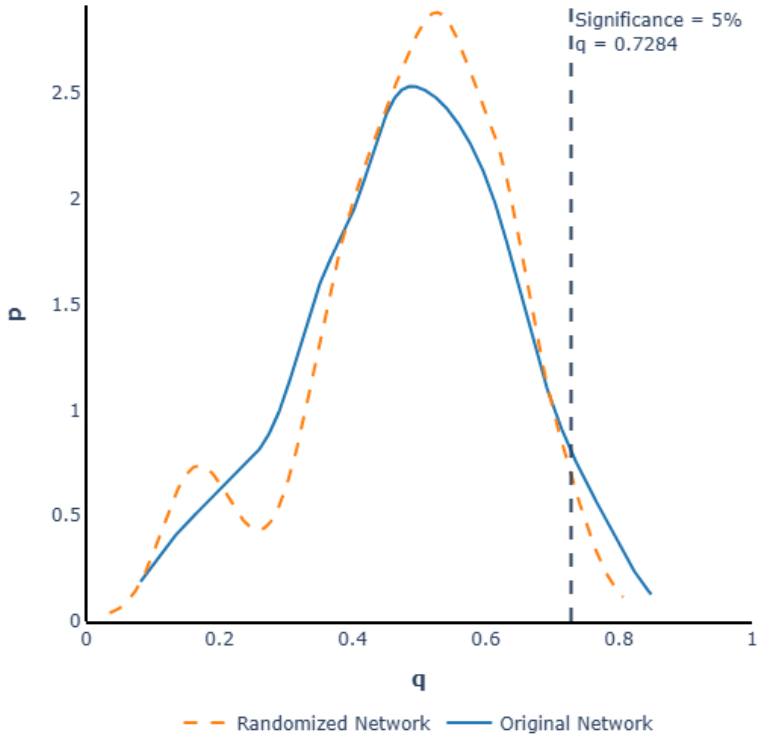


Figure 3.6 – Distribution of speech similarity links for 2020–2021. Original distribution in blue solid line, random network null distribution in the orange dashed line. Vertical black dashed line for 5% significant threshold.

For each of the time windows, we begin by calculating the speech similarity network with edges given by Eq. (3.2). Using the algorithm described in Section 3.4.1, we then analyze whether similarities in the network have statistical significance. In order to do that, we resort to the randomization algorithm described in Section 3.4.1 to calculate the distribution of links under the hypothesis of similarities driven by random choice of words

in central bankers' speeches. This is the null distribution against which we will measure significance of our estimated network's edges. Figure 3.6, using the Covid-19 pandemic biennium of 2020–2021 as example, shows the distributions of link strength for the originally estimated speech similarity network, in the solid blue line, and for the randomized null distribution, in the dashed orange line. Furthermore, the vertical black dashed line shows the 5% significance level for the null hypothesis, that sits around $q = 0.7284$. This level is used as threshold for separating links in the original network that can be attributed to random similarities, located left from the significance threshold, from those presenting evidence for strong relation between central bankers' speeches. We thus eliminate the links that did not meet the significance threshold from the original network. Moreover, since we are interested in measuring coordination of monetary policy, we exclude from our analysis central banks that remained with no links after this step. With this procedure, we guarantee our results are based on only significant links in each periods' speech similarity network. Figure 3.7 thus shows how connections in the speech similarity network has evolved through time, measured by the distribution of significant link strength, i.e., the distributions of the links whose strength exceeded the 5% significance threshold calculated based on randomized networks from their own time periods.

In addition to that, serving as benchmark for our speech similarity network, we perform a similar analysis on the two policy stance measures described in Section 3.4.3 and illustrated in Figure 3.1, namely monetary policy interest rate and quantitative easing or quantitative tightening. As previously mentioned, the two benchmark coordination networks are formed from the corelation of the respective policy stance measures across countries for the time windows matching the speech similarity evolution study. Relevant links are measured by the positive correlation pairs with 5% significance in the one-tailed t-test in Eq. (3.8). Figures 3.8 and 3.9 illustrate the evolution of the distribution of significant links for the two benchmark policy coordination networks.

Beginning our analysis by the first years depicted in Figure 3.7, we see relatively low coordination, captured by the position of the first two distributions, for the periods of 2002–2003 and 2004–2005. However, we can also see great dispersion and specially pronounced positive skewness in the two distributions. One must recall that the first years of our analysis are the years that communication was gaining traction as a tool of monetary policy, with different policymakers adopting instruments related to it at different paces. Illustrating that point with arguably the most straightforward use of communication to support conventional monetary policy tools, these years were marked by the early experiences with forward guidance in central bank speech. Forward guidance is central banks' disclosure to educate the public about its intended next policy decisions, in an explicit attempt to drive market expectations. Indeed, Campbell *et al.* (2012) point out that it was only a few years earlier, in 1999, that the Federal Open Market Committee (FOMC) began including explicit language

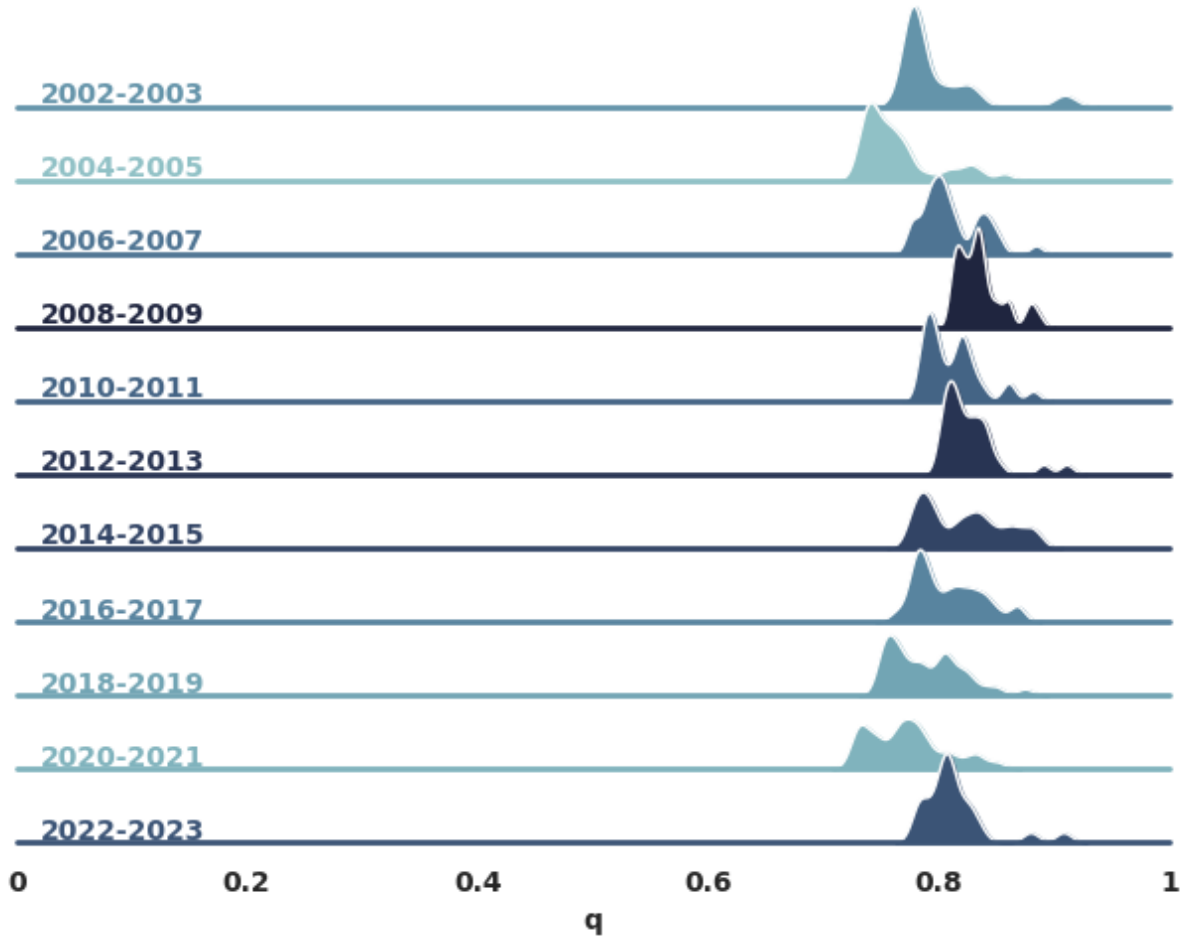


Figure 3.7 – Evolution of speech similarity network distribution over time.

about the future stance of policy in its meetings' decision statements, marking the early stages of its forward guidance adoption. Moreover, Svensson (2014) highlights that it was in 2012 that a complete forward guidance was adopted by the FOMC in the form of the "dot plot", reflecting individual committee members' expectations about the appropriate level of future policy rate. He also argues that variations of forward guidance have been adopted by other central banks through the years, with the Reserve Bank of New Zealand in 1997, the Norges Bank in 2005, the Sveriges Riksbank and the Bank of Israel in 2007, and the Czech National Bank in 2008. Therefore, the reduced coordination seems to reflect the greater dispersion of monetary policy instruments addressed in central bankers' speeches, with a range of maturity levels of communication and the policy instruments associated to it.

Moving our analysis to the next time windows, although the 2006–2007 period marked the beginning of the coordination trend that peak the following biennium, the story is much different between these two periods. Whereas in the 2006–2007 window the coordination was driven by the synchronous monetary tightening that began years earlier, as discussed by Forbes, Ha, and Kose (2024), the coordination in 2008–2009 reflect the stimulus with which central banks responded to the breakthrough of the Great Financial Crisis. This

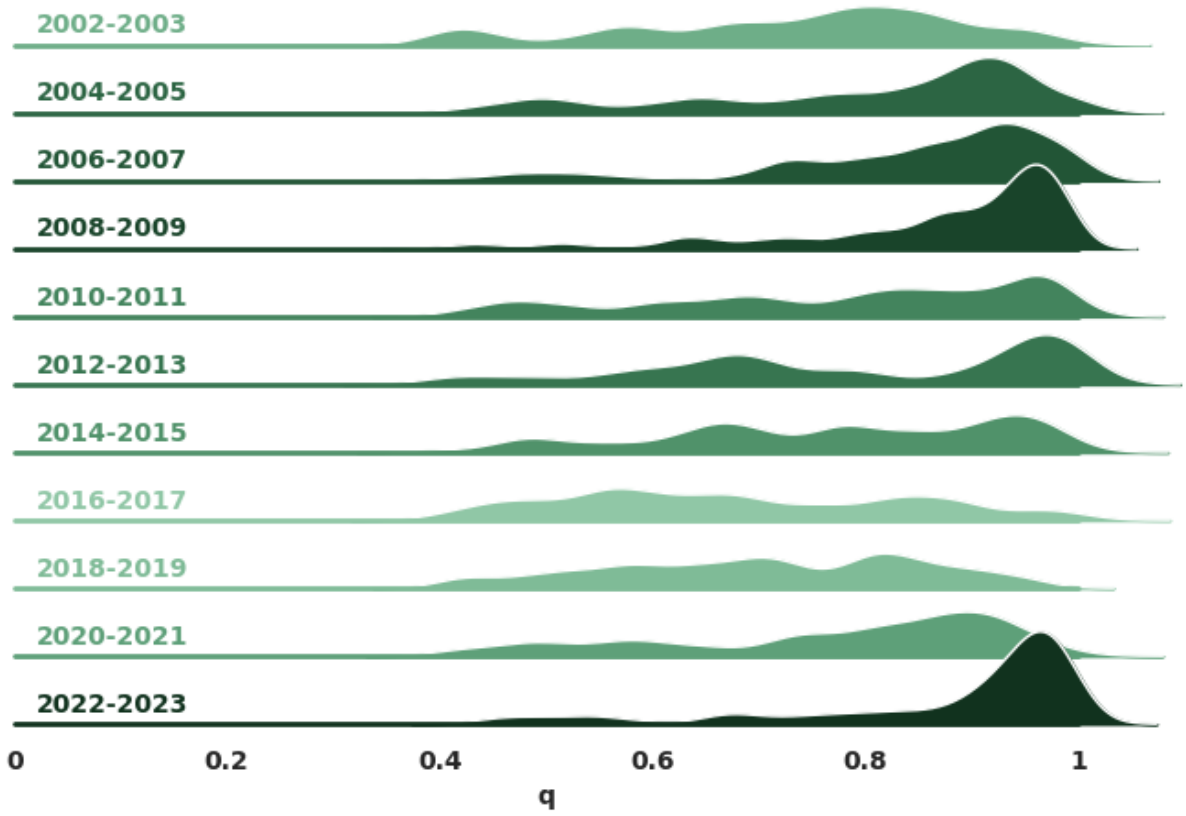


Figure 3.8 – Evolution of monetary stance network distribution over time.

patterns can be clearly observed when we complement the speech similarity coordination analysis with conventional monetary policy tools, specially from the stance derived from policy rates illustrated in the left-hand side of Figure 3.1. There, we can see that there has been a widespread shift from a contractionary policy stance, represented in red areas, to an expansionary stance, represented in blue. The coordination network formed from that policy rate stance measure, with evolution of significant links shown in Figure 3.8, tells a similar story as our proposed measure based on central bankers' speeches for the time interval between 2006 and 2009. As a matter of fact, both Figures 3.7 and 3.8 show that coordination peaked in the 2008–2009 biennium. This result is expected in light of the developments of the Great Financial Crisis of 2007–2008, in line with Blanchard, Ostry, and Ghosh (2013)'s observations that policy coordination tends to occur spontaneously on turbulent periods.

The following years were marked by high degrees of stimulus from monetary policy, with many central economies operating their policy rates near the zero lower bound. During this period, alternative instruments were proposed and became even more relevant than policy rates on central bankers' toolbox, given how limited interest rates were to provide additional stimulus[18]. Thus, central banks' asset purchasing programs, i.e., quantitative

[18] For a discussion on the zero lower bound and the alternative instruments of monetary policy, see Hamilton

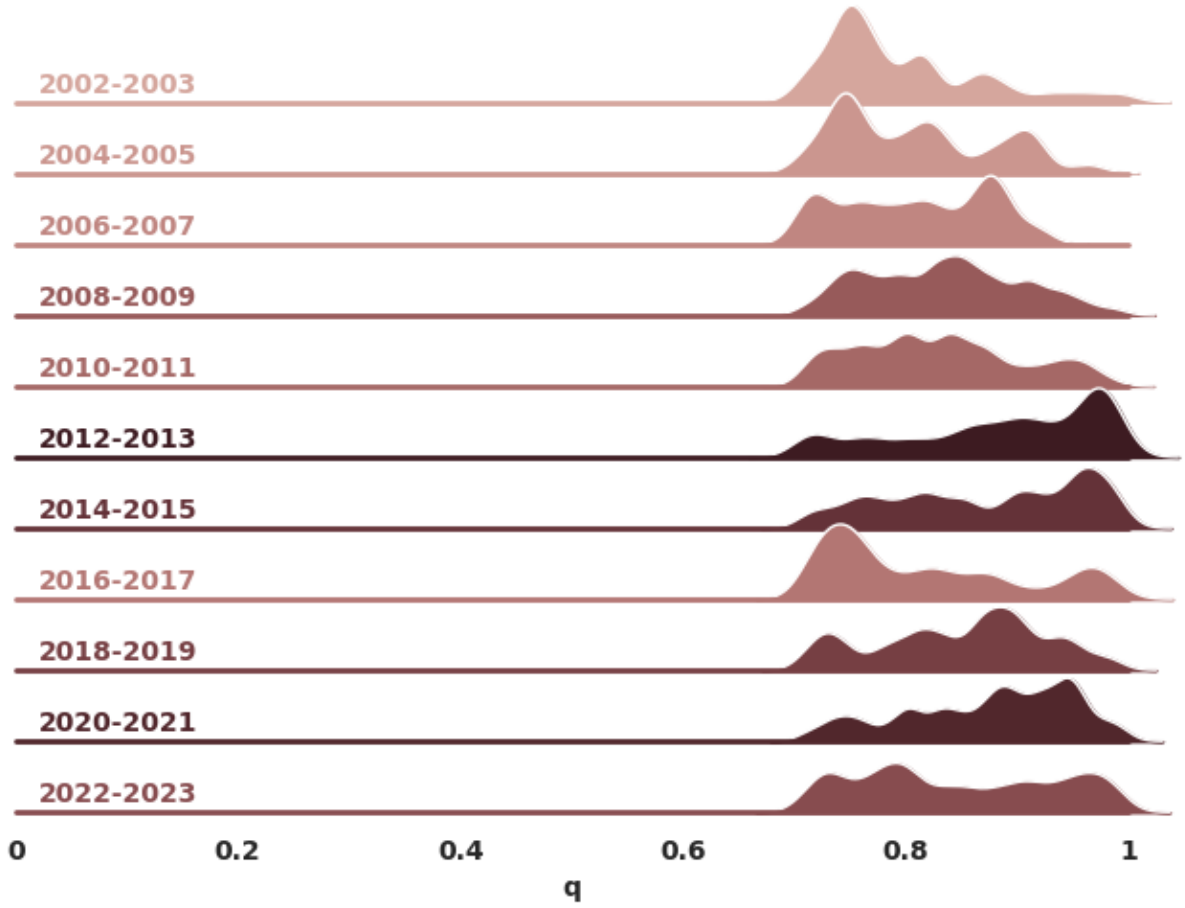


Figure 3.9 – Evolution of QE/QT network distribution over time.

easing, were arguably the most widely adopted alternative. Hence, turning our attention to the evolution of the coordination network formed from QE/QT policy stance, represented by the distributions of significant links in Figure 3.9, we can see that the years between 2012 and 2015 were years of increased coordination by this measure, with the first of these two bienniums, 2012–2013, presenting the peak for the period. These are exactly the same dynamics observed for our proposed speech similarity measure, with the same pattern of high coordination in the period, with 2012–2013 presenting itself as the peak, as can be seen in Figure 3.7. This result suggests that our speech similarity network was successful in capturing coordination in more than one policy instrument, using communication as the source of a multi dimension assessment.

The subsequent interval was marked by a reduction in coordination, a trend that lasted until after the Covid-19 pandemic by our speech similarity measure. Indeed, this period was characterized by the normalization of policy stance by the monetary authorities of many countries, but at different paces. However, during the pandemic, specially in the 2020–2021 biennium, there has been great simultaneous effort across countries, who resorted to both

and Wu (2012).

monetary and fiscal policies to provide stimulus to demand and oppose the economic risks associated with lockdown policies that took place to face the sanitary crisis. Despite being present in the QE/QT coordination network of Figure 3.9, the coordinated stimulus was only barely observable in the policy rate stance measure of Figure 3.8. In fact, during this period, speech coordination seems to have again responded more closely to the policy rate stance. This is specially true for the biennium right after the pandemic, 2022–2023, when the world turned to face the inflationary pressures that resulted from the combined economic stimulus from previous years. This period is reported as one of the highly synchronized tightening periods studied by Forbes, Ha, and Kose (2024), and is still a developing topic for policymakers and academics as the writing of this paper, motivating rich debate on the effects of such coordinated movements over perceived monetary policy results (Obstfeld, 2022).

3.5.3 Term relevance evolution

To conclude the assessment of our measure of monetary policy coordination, we now turn back to word-level analysis, in order to understand what drives the links between countries. Despite looking at specific terms as we did when understanding the long-term relations in the speech similarity network, this time we will focus on an evolution perspective. Here, we will explore how two effects discussed in the previous section can be explained from the co-occurrence of specific terms related to them across central bankers' speeches. The first effect is the observation by Blanchard, Ostry, and Ghosh (2013) that policy coordination usually occurs spontaneously in times of turmoil and the second is the relation between different central banks through unconventional monetary policy instruments, with special attention to the one with closest relation to central bank communication: forward guidance. In this Section, we focus on the four central banks responsible for the most liquid currencies, that respond for the greatest share of global trade: the Federal Reserve System, the European Central Bank, the Bank of England and the Bank of Japan.

Beginning by the analysis of periods of economic stress, we highlight the main term used by central banks to describe such scenarios: crisis, represented by the token *crisi*[19]. We also complement this analysis by including the token *pandem*, to understand how the mentions related to the Covid-19 pandemic appear in speeches through time. Figure 3.10 thus shows how relevant these terms were to the most influential central banks in the global economy, as measured by Eq. (3.3). We can see that mentions to *crisi* increased considerably in the two bienniums from 2006 to 2009, specially in the latter of the two time windows, for the four analyzed institutions. This evidence is in line with the conclusion we discussed in

[19] Once again one should note that we are analyzing the resulting terms after pre-processing as described in Section 3.3.

Figure 3.10 – Evolution of term related speeches - *crisis* and *pandemic*.

the previous Section for this time interval, that coordination reached a peak in the 2008–2009 due to the response of monetary authorities to the Global Financial Crisis. Additionally, we can also see that mentions to the term *pandem* appeared from the 2020–2021 period on. In fact, the 2020–2021 window also marked an increase in the relevance of token *crisi* for the European Central Bank and for the Bank of England, with also a marginal increase for the Bank of Japan.

Finally, we conclude our study with the terms related to forward guidance as an alternative instrument of monetary policy. Figure 3.11 shows how $f^k(w_i)$ from Eq. (3.3) evolved for the terms *forward* and *guidanc* among the four institutions from the previous analysis. We can see that, despite being present in previous periods in the appearances

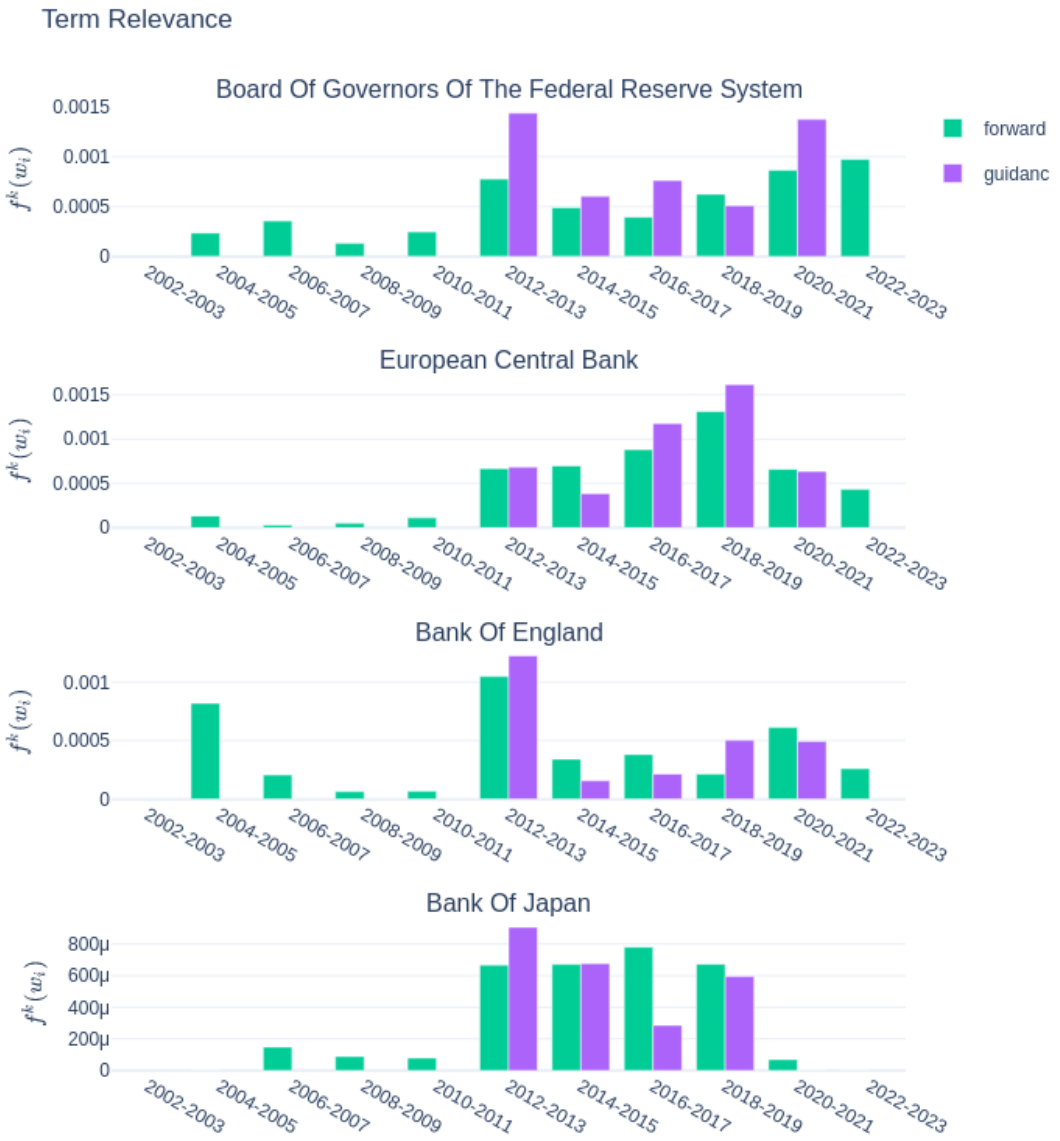


Figure 3.11 – Evolution of term related speeches - *forward* and *guidanc*.

of *forward* — which suggests that communications had some dimension of prospective assessment by central banks — the relevance of these two terms had a sharp increase during the 2012-2013 biennium. It is also interesting to note that they became highly correlated in speeches of every one of these four central banks, suggesting that most mentions relate to co-occurrences of the two terms. This evidence is in line with the observations of Svensson (2014), that the complete adoption of forward guidance by the most influential central bank in the global economy, the Federal Reserve System, happened the first year of that biennium, taking form in the "dot plot". Furthermore, this assessment suggests that, during the years in which the zero lower bound was a binding limit to policy rates, the adoption of alternative instruments went beyond the asset purchase programs captured by the measure in Figure

3.9. This complements the conclusions from the previous Section for the 2012–2013 increase in coordination, also providing qualitative insights on how it took place.

3.6 Conclusion

In this paper we provided a novel measure of international monetary policy coordination, leveraging the information present in central bankers' communication to draw a speech similarity network in the fashion of Cajueiro *et al.* (2021)'s companies' news similarity network. We showed how this measure responded to the main global economic events from 2002 to 2022, with results aligned to economic theory and empirical evidence, thus proving its practical value. Our exercises suggest that monetary policy coordination gets higher in times of economic turmoil. Furthermore, we provided insight on how long-term similarity between central banks' communication relate to economic fundamentals.

As a direct contribution of our work to the related literature, our novel measure could enhance the assessment of monetary policy coordination on empirical works. A straightforward example relates to Caldara *et al.* (2024), who propose their model can be extended to incorporate unconventional monetary policies, as the ones discussed in Gertler and Karadi (2011). Since our measure is able to capture coordination on monetary policy instruments other than conventional short-term interest rates, it should prove valuable to assess coordination effects in a more comprehensive fashion when combined with such empirical approaches.

We also see some directions to which our study could be further extended. First, through a methodological road, additional efforts to explain the dimensions by which similarities are established, augmenting the word-level assessment we provided, can shed light on other coordination drivers. We can achieve this by resorting to other natural language processing techniques, such as topic modeling, for instance, to assess what subjects determine central bankers' speech similarity. Second, in an application-directed way, our network can also be used to enhance diversification of financial portfolios, since it captures forward-looking sources of similarities between economies not present in common correlation measures. This information is of great value to foreign exchange portfolio managers, as well as for international investment strategies with other asset classes, such as stocks or bonds. Finally, our method can also contribute to the discussion about leader-follower monetary policy coordination (Figueroa; Padilla, 2022). For that, we may analyze whether the formed networks or communities exhibit some kind of behavioral contagion, where certain leaders emerge and begin to influence or drive concerns about a given monetary policy outlook, prompting other central banks to follow their lead.

4 A Multi-Dimensional Sentiment Index from Brazilian Central Bank Communication

4.1 Introduction

Effective monetary policy should not be just about influencing the short end of the economy's yield curve through traditional policy rates instruments. Both the investment and the asset prices monetary channels depend on long-term interest rates, be it by the effects over real investment decisions or by serving as benchmark for asset pricing[1]. Hence, monetary policy communication finds itself as one of the main instruments in modern central bankers' toolbox. Going beyond the short-term nature of policy rates, it has the potential of steering market participants' expectations and affecting longer-term horizons. In practical terms, such expectations shape the economy's yield curve, driving not only the level of future interest rates, but mostly yield curve's slope, as well as its curvature. Thus, when measuring the sentiment transmitted to economic agents through central bank communication, we should account for all these dimensions.

With that in mind, we propose a framework for estimating expectation-embedded multi-dimensional sentiment from monetary policy communication, combining economic fundamentals and state-of-the-art deep learning neural networks. We begin by building from the Litterman and Scheinkman (1991) approach to yield curve modeling. As we are interested in measuring monetary policy communication's effect over the entire term structure of interest rates, their approaches' three factors — understood as level, slope and curvature — provide economically meaningful measures of agents' expectations priced in financial assets, while capturing the vast majority of yield curve's variance. We then build a deep learning neural network model designed to extract sentiment from such factors' movements on events of information disclosure by the monetary authority. Furthermore, considering that we are modeling policy communication, it is important to have a robust way of handling textual data. For text modeling we incorporate in our framework the Bidirectional Encoder Representations from Transformers, BERT (Devlin *et al.*, 2019). BERT is based on the groundbreaking transformer architecture (Vaswani *et al.*, 2017), which currently provides state-of-the-art results to natural language processing problems.

As for the data used in our empirical analysis, we resort to Brazilian monetary policy communication. The reason for this choice is twofold. First, the Brazilian Central Bank's Monetary Policy Committee (COPOM) has a stable and well structured set of official communication tools, mainly in the form of its decision statements and meeting minutes. These instruments have evolved through time into a format that has an explicit goal to simplify and clearly transmit the monetary authority's messages to economic agents[2]. Second, and more important, COPOM statements and minutes are published with only six days apart from

[1] For a seminal discussion on monetary policy transmission channels, see Mishkin (1996).

[2] See Fasolo, Graminho, and Bastos (2022) for a long-term assessment of the Brazilian Central Bank communication.

each other, with three market sessions between them. This aspect, along with the fact that both documents relate to the same policy decision event, lets our framework understand what is already known by market participants and what is actually novel information from communication, not anticipated and thus expected to move asset prices.

We thus build our neural network model with two stages, designed to fit the different instances of central bank communication. The first stage focuses specifically on the disclosure of policy decision rationale in monetary committee statements, and also incorporates the new information provided by the interest rate decision itself. Besides the sentiment extracted from the document, this stage also outputs the representation of the information that market participants assimilate, which stands for the state of policy communication. As for the second stage, it is a more general setup, designed to update the information provided by new communication. It thus receives the previous state of communication as generated by first stage network, along with new central bank disclosures to extract the sentiment present in the novel information issued by the policymaker. Given the data at hand, this stage focuses on monetary policy meeting minutes, released few days after policy decision.

Our main contribution to the economic literature is to provide a framework that delivers and updates a multi-dimension monetary policy communication sentiment index. In line with this contribution we point to Niţoi, Pochea, and Radu (2023), who also develop a sentiment index for monetary policy communication. Furthermore, Araci (2019) provides a domain-specific fine-tuned version of BERT, named FinBERT. This model is also a ready for use BERT-based model targeted towards sentiment analysis, but their scope is broader, with the financial domain in mind. And in a similar fashion, but targeted specifically at central bank communication, Pfeifer and Marohl (2023) propose a fine-tuned version of RoBERTa (Liu *et al.*, 2019a), named CentralBankRoBERTa. However, these models provide single dimension sentiment index, while we increase dimension for complete and adequate market reaction incorporation as measured by yield curve factor shifts.

Our proposed model also makes some relevant methodological contributions to the branch of machine learning methods designed for economics and finance. The discussions we bring forth from our modeling strategies should help reduce the costs of designing, training and deploying models that apply state-of-the-art deep learning techniques for domain-specific financial questions. First and foremost, we enrich the literature that relies on market data to infer economic agents' sentiment in response to specific events. This is a strategy that supplants the need for manually labeling data for supervised leaning tasks, which depends on large volumes of data for the model to infer the patterns that lead to the expected outcomes. The conventional manual approach resorts to specialists' evaluation of each observation, in our case in the form of textual data, in a work that potentially takes many hours to produce the necessary volumes of data. Furthermore, it is usually necessary for more than one analyst to score data, in order for the results to be validated. On the other hand,

by resorting to market data, we assume that prices reflect the set of available information that agents have in any given moment, and the price response to meaningful events should represent the novel information being transmitted to asset prices. This way, one only has to take care for correctly handling the sets of information embedded in prices in order to systematically produce the data used for model training. Therefore, our model design and data treatments to deal with this kind of labeling strategy should enrich the discussion on such methods. Second, our training strategy and the resultant discussion on small samples further enables the reduction of costs for model training, once it allows less data to be fed into the model in this phase. And even more important, as some practical problems simply do not have enough data available for a usual deep learning model estimation, making small sample estimation available is crucial to make the deep learning approach feasible to tackle some questions in economic domains.

Our results reinforce the need for more than one sentiment dimension to comprehensively capture the relevant nuances of communication for monetary policy assessment. Indeed, we see that the first dimension of our sentiment index relates to the hawk/dove gauge usually adopted by conventional monetary policy sentiment analysis. This suggests that these approaches are a special case of our framework. Furthermore, we show that second and third dimensions of our sentiment index can at times be more relevant in setting the overall tone of policy communication. As a matter of fact, our results show that these new dimensions can even precede shifts in monetary policy stance.

We thus structure the remainder of this paper as follows. Section 4.2 reviews the related literature both on yield curve modeling and on economic textual data analysis. Section 4.3 describes our yield curve data and modeling choices, discussing how they incorporate economic fundamentals from market expectations extracted from pricing reactions to monetary authority communication release events. Section 4.4 details the textual data used in our framework, also discussing how this specific dataset was specially selected for fitting one of the main premises of our model. Section 4.5 guides the reader through the model's design decisions that try to capture the economic fundamentals from previous steps. Section 4.6 then discusses empirical results, showing our multi-dimensional sentiment index in practice, assessing its performance and exploring the insights it provides from actual central bank communication. Next, Section 4.7 brings our concluding remarks. Finally, the computational details of our framework are delegated to the appendices. How we incorporate BERT for textual modeling is discussed in Appendix A and our neural network framework is detailed in Appendix B. Then, Appendix C addresses the data labeling process from yield curve factor movements and Appendix D describes our training strategy, showing how deep learning hyperparameter decisions were made and the training and out-of-sample validation performance of our framework's various alternative setups

4.2 Literature Review

The well established term structure of interest rates decomposition literature that sets the economic foundations to our analysis has in Nelson and Siegel (1987) one of the most relevant early contributions. They proposed a parametric approach to decompose the yield curve in three factors that, when added up, presented good fit to empirical data. These three factors were written as exponential functions of a shape parameter. Given their share in the form of the final curve, they were interpreted as level, slope and curvature. Each of them was associated with a weighting term, defining the yield curve described by the model. This was the cornerstone of the methods that we build our framework upon.

In order to further improve empirical adherence of the Nelson and Siegel (1987) approach, resulting in one of the currently most adopted yield curve representation methods by both researchers and market participants, Svensson (1994) improves the baseline approach by adding a second shape parameter, along with an additional weight term for the extra curvature introduced. This added degree of freedom makes the overall resulting term structure of interest rates representation more flexible, allowing even better fit to observed yields. Practical term structure representations from this model are computed and updated in regular basis for various economies, including for United States Treasuries (Gürkaynak; Sack; Wright, 2007) and for the Brazilian sovereign yield curve. The latter stands as the starting point for developing our framework, as we will discuss ahead when we get into details about data and methodology.

Furthermore, in maybe one of the most important works derived from Nelson and Siegel (1987), Diebold and Li (2006) developed a framework known as the Dynamic Nelson-Siegel. Their model used the same parametric functional form as the original approach, but allowing its parameters to vary in time. Furthermore, Diebold and Li (2006) proposed that factors' weighting parameters followed a first-order autoregressive process, AR(1). This way, forecasting became more natural, once factors' dynamics were explicitly modeled. This approach is still a popular one for describing yield curve dynamics and performing forecasts.

Parallel to those developments, the seminal work of Litterman and Scheinkman (1991) proposed an alternative view over the factors that shape the term structure of interest rates. The authors resorted to principal components analysis to decompose the covariance matrix of interest rates, effectively rewriting empirical data in a new set of coordinates. The orthogonal vectors defining the dimensions of the new coordinate system shaped the factors in which the curve would be decomposed. Given the loads of such vectors in the original 'maturity coordinate system', they were also interpreted as level, slope and curvature, drawing a parallel with the Nelson and Siegel (1987) approach. As matter of fact, Almeida, Jr, and Fernandes (2003) show the relation between Nelson and Siegel (1987)'s parametric factors and Litterman and Scheinkman (1991)'s principal components.

There is also an extensive literature that, similar to our work, have its foundation on the mentioned yield curve models. Among them, we highlight two strands of research that are closely related to this paper. On the one hand, we find works that focus on studying the insights provided by these decomposition methods to economically relevant concepts, such as interest rates expectation and term premium (Cochrane; Piazzesi, 2009; Adrian; Crump; Moench, 2013), or even more broad relations to macroeconomic concepts as inflation expectations and output gap (Rudebusch; Wu, 2008). These works, along with the factor-oriented discussions of Litterman and Scheinkman (1991), Svensson (1994), Gürkaynak, Sack, and Wright (2007), are of special relevance to our modeling decisions, as they provide the bases of the economic meaning embedded in our yield curve factors-driven sentiment measures. These evidences are also backed by more general discussion on yield curve relation to economic agent's expectations and risk perception, explored in Fisher (2001), Boeck and Feldkircher (2021). On the other hand, we find an extensive line of works that try to explore yield curve model's dynamics, mostly focused on interest rates forecast (Ang; Piazzesi, 2003; Diebold; Rudebusch; Aruoba, 2006; Cajueiro; Divino; Tabak, 2009; Matsumura; Moreira; Vicente, 2011; Tabak *et al.*, 2012).

We now move from the well established literature of yield curve modeling to the fast-paced economic textual analysis research (Mishev *et al.*, 2020; Fasolo; Graminho; Bastos, 2022; Shapiro; Sudhof; Wilson, 2022; Shapiro; Wilson, 2022; Apel; Grimaldi; Hull, 2022), recently fueled by frontier natural language processing advances such as Vaswani *et al.* (2017). This literature in its current phase had its first contributions with Boukus and Rosenberg (2006), who analyzed the Federal Open Market Committee (FOMC) minutes with a machine learning technique to quantify the qualitative information in monetary policy communication. In a result that we resonate in our modeling decisions, they find evidence that market participants can extract complex, multifaceted signal from policy committee meeting minutes that affect yield changes.

Moreover, analyzing statements and minutes of Brazilian monetary policy communication, Fasolo, Graminho, and Bastos (2022) build from an inherently multi-dimension method with topic modeling. They then build sentiment indexes based on positive and negative terms according to a predefined dictionary. Among their main findings, which support our framework design and training strategy, they suggest both documents, statements and minutes, are coherent in terms of information transmitted. Furthermore, moving back to U.S. FOMC, Shapiro and Wilson (2022) use full meeting transcripts, made publicly available few years after monetary policy meetings. They also perform sentiment analysis based on dictionary methods — one of the still most common approaches — to estimate central bank preferences. Among their results, their approach suggests that financial variables, such as asset prices, are weighted by the monetary authority in policy decisions.

Going beyond central bank communication to other applications of language processing

in economics, it is worth noting that the sentiment analysis approach we resort to is one of the most adopted by economic researchers (Mishev *et al.*, 2020; Shapiro; Sudhof; Wilson, 2022). From the usually supervised learning nature of this approach, it becomes necessary to provide annotated data to the machine learning model, essentially letting it infer what outcomes are expected from the training dataset and allowing it to find patterns that lead to correct results in more general out-of-sample exercises. For data labeling strategies, there are two main alternatives followed by the literature. On the one hand, some works engage in producing manually labeled data, usually by specialists' annotation of documents (Araci, 2019; Pfeifer; Marohl, 2023). On the other hand, some works rely on market price movements to define whether documents have positive tone, usually related to asset value increase, or negative tone, which commonly result in asset depreciation. This approach can be further separated into two categories: (i) works that use up or down price directions to define binary labels for classification (Liu *et al.*, 2018); and (ii) works that use the magnitude of price movements as labels for textual data (Yang *et al.*, 2018; Chen *et al.*, 2019; Petropoulos; Siakoulis, 2021). In this paper, we experiment with both types of market-driven factor movements to set monetary policy communication labels for our supervised learning framework.

This strategy of training data labeling leads to one of the nuances that sets our work apart from other papers that combine textual analysis with yield curve modeling. Our approach is based on inferring sentiment from expectations priced in the yield curve, while previous contributions usually make the inverse path, computing sentiment from monetary policy communication independently from market prices and then using it to perform analysis over the term structure of interest rates. It is the case of Chague *et al.* (2015), that use a dictionary method to assess the impact of the Brazilian Central Bank communication over the yield curve. Despite the differences in methodology, their work is close to ours in other dimensions. First, it focuses on the same textual data universe, the Brazilian monetary authority communication. But more important, they find that communication successfully affects long-term interest rates by driving markets expectations, one of the main premises of our approach.

Furthermore, we find in Alves, Abraham, and Laurini (2023) another approach close to ours. Once again in a work targeted at the Brazilian monetary policy communication, this paper also uses factor decomposition to study the impacts of communication over the yield curve. However, their approach is different from ours in many ways. First, they resort to the Nelson and Siegel (1987) model for factor estimation. As a matter of fact, in the same fashion as Gotthelf and Uhl (2019), they extend this model by incorporating a new sentiment factor, in order to predict movements of interest rates across different maturities. Therefore, they also begin by measuring sentiment from textual data, again by dictionary method, and only then assess the effects over the term structure. However, they as well find relevant contribution of communication sentiment to price formation on long-term maturities of the

yield curve.

In terms of textual analysis methodology, we are not the first to use BERT or other transformer-based models to assess monetary policy communication or financial markets in general. Indeed, other works have tested them before with the resulting textual analysis comparable to experts' opinions in the financial domain (Mishev *et al.*, 2020). As mentioned, examples in line with our work are Niţoi, Pochea, and Radu (2023), who develop a sentiment index for monetary policy communication, Araci (2019), with a domain-specific fine-tuned version of BERT, and Pfeifer and Marohl (2023), with a fine-tuned version of RoBERTa (Liu *et al.*, 2019a), a model derived from BERT, designed for central bank communication.

4.3 Yield Curve Modeling

We organized this Section in four subsections. Section 4.3.1 describes the Nelson Siegel Svensson (NSS) model, which supports the term structure of interest rates data we use. Section 4.3.2 discusses the method adopted to estimate latent yield curve factors, which represent the dimensions on which our sentiment index is measured. Section 4.3.3 then discusses the economic fundamentals behind each of the factors included in our framework, providing the basis to extract meaning from each of the dimensions of our proposed sentiment gauge. Finally, Section 4.3.4 details how we calculate the expected monetary policy rate decision from yield curve data, necessary to measure the market surprise for each policy meeting outcome.

4.3.1 Nelson Siegel Svensson Model

In our study, yield curve data consists of parameters of the Nelson Siegel Svensson (NSS) model (Nelson; Siegel, 1987; Svensson, 1994), estimated from Brazilian Treasury bonds by ANBIMA, the Brazilian Financial and Capital Markets Association, and published on their website on a daily basis[3]. We also use time series of the CDI, the one-day interbank interest rate. It is the Brazilian risk free rate and follows closely the monetary policy interest rate, Selic. Both series start on September 21, 2009 and go until December 31, 2024.

The Nelson Siegel Svensson model is a parametric yield curve model widely adopted for term structure of interest rates representation for its ability to summarize in few parameters its entire dynamic. To be more specific, following Svensson (1994)'s notation, the set of parameters $\{\beta_0, \beta_1, \beta_2, \beta_3, \tau_1, \tau_2\}$ capture yield curves' level, slope and curvatures, along with slope and curvatures' location across maturities. Eq. 4.1 describes the model for spot zero coupon yields:

[3] https://www.anbima.com.br/pt_br/informar/curvas-de-juros-fechamento.htm.

$$
\begin{aligned}
i(m) = \beta_0 & \\
& + \beta_1 \frac{1 - \exp\left(\frac{-m}{\tau_1}\right)}{\frac{m}{\tau_1}} \\
& + \beta_2 \left(\frac{1 - \exp\left(\frac{-m}{\tau_1}\right)}{\frac{m}{\tau_1}} - \exp\left(\frac{-m}{\tau_1}\right) \right) \\
& + \beta_3 \left(\frac{1 - \exp\left(\frac{-m}{\tau_2}\right)}{\frac{m}{\tau_2}} - \exp\left(\frac{-m}{\tau_2}\right) \right),
\end{aligned}
\tag{4.1}
$$

where i is spot interest rate, m is the maturity in years and $\{\beta_0, \beta_1, \beta_2, \beta_3, \tau_1, \tau_2\}$ is the set of parameters as mentioned before.

This yield curve representation also provides forward rates closed form equation, which we use to calculate policy decision surprises. Eq. 4.2 is the NSS expression for forward rates:

$$
f(m) = \beta_0 + \beta_1 \exp\left(\frac{-m}{\tau_1}\right) + \beta_2 \frac{m}{\tau_1} \exp\left(\frac{-m}{\tau_1}\right) + \beta_3 \frac{m}{\tau_2} \exp\left(\frac{-m}{\tau_2}\right). \tag{4.2}
$$

4.3.2 Principal Component Factors

From yield curve series, we estimate latent factors according to Litterman and Scheinkman (1991), by applying principal component decomposition. In order to do that, we proceed as follows:

- First we calculate yield curve series from ANBIMA's estimated Nelson Siegel Svensson daily parameters according to Eq. 4.1. For vertices selection, to get a complete representation of the term structure, we define a ten-year horizon with quarter time step. This provides a long enough time horizon and granular enough vertices dispersion with quarter spacing between them. As a matter of fact, quarter spacing is also the minimum spacing between Brazilian Treasury Bonds' maturities;
- Second we standardize each vertex' series by subtracting the mean and dividing by the standard deviation. This step guarantees that the different variance profile across vertices does not result in biases in factor loadings;
- Then we calculate the variance-covariance matrix for the standardized yield curve series from the previous step;
- Finally we calculate latent factors from the eigendecomposition of the variance-covariance matrix, according to Eq. 4.3:

$$
\mathbf{C} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^T, \tag{4.3}
$$

where **C** represents the variance-covariance matrix, **V** is a matrix composed of eigenvectors in its columns, representing yield curve latent factors' loadings across vertices, and **Λ** is a diagonal matrix with eigenvalues associated with each eigenvector, disposed along its main diagonal. Eigenvalues represent each factor's contribution to the overall variance of the yield curve series, providing a mean of ordination of factors' relevance. Factor weights are then calculated from normalized eigenvalues. Once with factor loadings, factor scores are calculated by Eq. 4.4:

$$\mathbf{Z} = \mathbf{XV}, \tag{4.4}$$

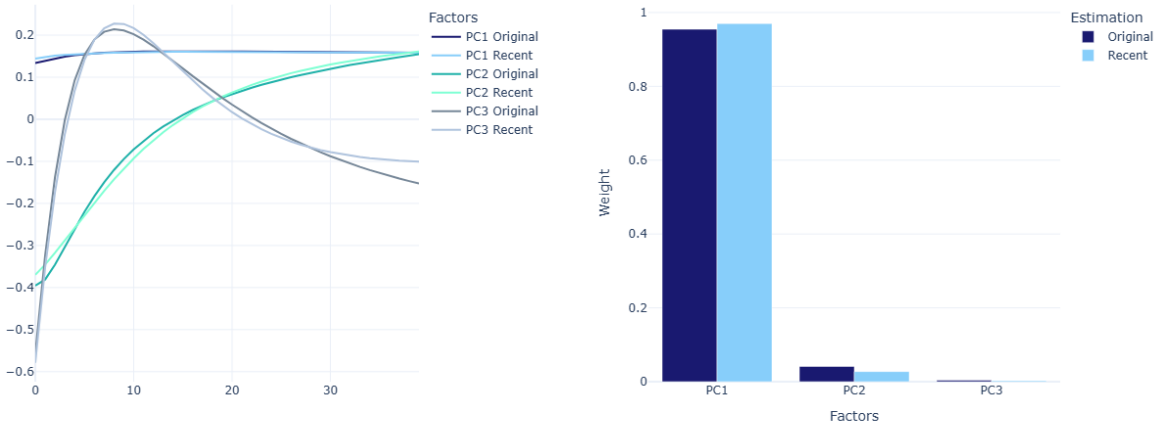
with **X** as stacked time series of yield curves, **V** as factor loadings and **Z** as stacked time series of factor scores.

Given that eigenvectors are all orthogonal to each other, principal component decomposition stands as a linear transformation into a new coordinate system. This new system is based on directions that capture the largest variation among original data. Therefore, by selecting a subset of dimensions that comprise most of the variance, this method stands as a method for dimensionality reduction, keeping most of information from original data. In our yield curve factor application, we follow Litterman and Scheinkman (1991) and keep the three first components, as they are reportedly enough to describe most of yield curve's dynamics. These will be the dimensions along which our sentiment analysis will be performed on central bank communication.

For factor estimation, we split the yield curve time series into training and scoring periods, so we do not incorporate future information in our model at any given moment. We do that by restricting the factor train period to the beginning of the series, on September 21, 2009, until the last day before the first central bank communication we assess, as we will discuss ahead, on July 19, 2016. This approach guarantees that factors are constant across our entire communication assessment period and scores are comparable. It also does not significantly diverges from the results we would have if we estimated factors on a recurrent basis, a result that we verify empirically.

Summarizing all that has been discussed thus far, Figure 4.1 shows factors' shapes and weights, along with their evolution across training and scoring periods. Factor shapes are drawn from factor loadings, which in turn are derived from variance-covariance eigenvectors, while factor weights are estimated from variance-covariance eigenvalues.

From factor shapes, in Figure 4.1a we can see two important information. First, in the evolution analysis, we can observe the empirical confirmation that factors' loadings across vertices are rather stable. Here, factors labeled as "Original" are factors estimated during the training period, used through all of our study; whereas factors labeled as "Recent" are estimated on scoring period and used to empirically validate time stability. This analysis shows little change between the two periods.



(a) Factors' shapes (loadings)

(b) Factors' weights

Figure 4.1 – Factors' Evolution

The second important information we can get from Figure 4.1a concerns factors' shapes. As discussed by Litterman and Scheinkman (1991), the three first principal components of yield curve decomposition provide interpretable factors capturing level, slope and curvature of yield curves. Recalling that these factors represent the dimensions on which our sentiment scores will be calculated, this insight brings additional interpretation to what we can expect our multi-dimensional sentiment analysis to incorporate.

We must also take a moment with Figure 4.1a to understand the coordinate system in which we will measure our sentiment index. Beginning by the first factor, named PC1[4] and representing the level factor, we see that its loadings are almost constant along every yield curve's vertices. Moreover, all the loadings are positive. This means that a positive shift along this dimension will result in an almost constant yield increase along the entire curve, which gives this factor the level meaning. Note that, since sentiment is measured in principal component dimensions, a positive realization in the first dimension for our sentiment index represents negative tone, usually related to contractionary indications by the central bank, as we will discuss in the next section. Furthermore, Figure 4.1a shows similar interpretation for the second factor, named PC2. We can see that this factor has loadings that increase with yield curve maturity. However, differently from what we have seen for the level factor, the slope factor has the short-end of the curve with negative loadings and the remaining term structure of interest rates in positive territory. We can thus conclude that a positive shift along this dimension represents an increase in slope of the yield curve, which in turn is related to higher long term inflationary risks, as we will see ahead. Therefore, once again positive

[4] PC1 stands for "Principal Component 1", or the one with the biggest contribution to the overall yield curve variance. The Principal Components that follow use the same naming convention: PC2 for the second most representative and PC3 for the third.

realizations of our sentiment index along this dimension represent negative sentiment. And finally, for the third dimension, given by PC3 and representing the curvature factor, we can see that loadings are negative for the short-term maturities of the yield curve, become positive up until mid-term vertices and become negative again further on. Thus, positive sentiment along this dimension result in an increase in mid-term yields, and decrease in the rest of the curve, a behavior that we will see to be associated with augmented perception of yield volatility. Hence, positive realizations of our index along the third dimension represent the perception that volatility will increase, again resulting in negative sentiment interpretation. Thus, positive shifts in any of the dimensions of our index represent negative sentiment, each for a different and meaningful reason. It is also straightforward to see that negative realizations for our index represent positive sentiment for every dimension. We decided to leave this behavior unchanged so we have our sentiment gauge closely related to yield curve dynamics, in which increases represent negative sentiment.

Moreover, to understand exactly how our multi-dimensional sentiment index captures market sentiment from monetary policy communication along the dimensions of yield curve's level, slope and curvature, we must start from the process of price formation in the bonds market. From the efficient-market hypothesis (Fama, 1970), we can expect prices to reflect the available information at any given point in time. Furthermore, communication as a monetary policy instrument works by providing new information to market participants, either about the policymakers' economic outlook or the policy rules that drive interest rates decisions[5]. This way, in order to reflect the sentiment from novel communication, we must consider how the yield curve changes in response to the new information made publicly available. We therefore must work with yield curve shifts. Translating this into our factor approach, we see that our sentiment index assume the form of expected factor score shift, which in turn can be expressed by Eq. 4.5:

$$\Delta z_{k,t} = z_{k,t} - z_{k,t-1}, \forall k \in \{1, \ldots, K\}, \tag{4.5}$$

where $z_{k,t}$ is factor k score in time t, and K is the number of factors used for the multi-dimensional sentiment index. In our current implementation we use $K = 3$.

Additionally, from factors' level, slope and curvature interpretation, we see that we have on Nelson Siegel Svensson model's parameters a straightforward alternative to our approach, since they carry similar meaning. However, the reason for our choice of principal components are twofold:

First, despite the valid parallel between β_i parameters and our factors' scores, the NSS model allows for factor shapes to vary according to τ_1 and τ_2. One must recall that shapes

[5] One should note that this holds true even in the presence of rational expectations, since the communication policy instrument addresses an information asymmetry problem (Blinder *et al.*, 2008).

are given by factor loadings in our principal component approach and present adequate time stability, allowing us to safely keep them constant along our study. In order to achieve similar results in a NSS framework, one would have to reestimate the model imposing further restrictions, a process much more complex than our principal component decomposition.

Second, principal components for yield curve modeling brings flexibility to our framework, allowing for the inclusion of a higher number of dimensions than those originally proposed. Furthermore, we do not need any parametric or previous assumption of higher order factors for their inclusion, once principal components decomposition estimates factors' shapes from data. Despite our parsimonious approach following Litterman and Scheinkman (1991)'s choice limited to three principal components, backed by the economic meaning of such factors, some authors suggest that a higher number of factors could improve representation of yield curves (Adrian; Crump; Moench, 2013).

Besides factor loadings, it is also interesting to see how weights stand across factors and how they evolved between training and scoring periods. Figure 4.1b shows these information. It shows how much of total variance is explained by each principal component, with the first one, representing the level factor, responding for over 95% of the overall yield curve variance. By using the three, we account for over 99.99% of total variance. Furthermore, in line with what we saw in Figure 4.1a, we can see that factor weights are also stable over time.

4.3.3 Factors' Economic Fundamentals

We should also take a moment to understand the fundamentals embedded in each of the three factors, as they will be the economic drivers captured by our multi-dimension sentiment index. Several works address this issue, trying to draw a bridge between macroeconomic variables and yield movements (Ang; Piazzesi, 2003; Dewachter; Lyrio, 2006; Diebold; Rudebusch; Aruoba, 2006; Evans; Marshall, 2007). Beginning by the level factor, which we have shown to respond for the largest amount of yield curve variance, it stands as the factor closest related to monetary policy and inflation expectation variables. We can see that when we look through the lens of the expectation theory, that postulates that current and the anticipated path of future short interest rates define longer-term maturities (Lutz, 1940; Rudebusch, 1995; Crump; Eusepi; Moench, 2024). Indeed, this theory builds from the fact that spot yields that form the term structure are compositions of short rates, which are set by the policymaker. A shift in the anticipated stance of monetary policy thus affects the entire term structure of interest rates. This way, the dimension of our sentiment index resulting from the level factor is the one usually captured by single dimension dove/hawk sentiment measures.

In spite of being an important part of yield curve formation, the expectation theory in its pure form is usually refuted by the empirical literature (Cook; Hahn, 1990; Campbell; Shiller, 1991; Sarno; Thornton; Valente, 2007), which finds that observed data deviates from

the predicted theoretical results by a time-varying risk premium, also referred to as the term premium. The premium component is usually related to the slope factor (Kim; Orphanides, 2007; Joslin; Priebsch; Singleton, 2014), with some evidence pointing to the connection thorough economic cycle conditions (Dewachter; Lyrio, 2006)[6]. Recalling that the term structure is referenced in nominal yields, it is straightforward to note that inflation risk is one of the main drivers of the premium priced in longer maturities, hence standing as an important source of shifts along the slope dimension. Therefore, we have in the second dimension of our sentiment index a measure commonly associated with the risk perceived by market participants.

Lastly, the curvature factor is related to implied volatility of interest rates along the term structure. To understand this relation, we must first understand how bond duration and convexity influence its price and the response to an increase in volatility, ultimately affecting the shape of the yield curve. On the one hand, duration stands as a measure of sensibility of bond price to variations on its negotiated yield. Duration gives the magnitude of a relation that is always negative: as the yield negotiated by market participants rise (fall), the bond price depreciates (appreciates). On the other hand, convexity measures the sensibility of duration to shifts in yields. As a second order derivative that is always positive, convexity makes bond prices more (less) sensible to yield changes as they fall (rise)[7]. Furthermore, one should note that, while duration increases linearly with maturities along the yield curve, convexity increases non-linearly as we move towards long end of the term structure. These combined effects make the overall response of bond price to an increase in volatility assume the shape of our third principal component — noting that principal components are measured along yield dimensions so the negative relation of factor loadings to bond price have to be accounted for. Gilles (1996) provides in depth explanation of this relation, showing how stochastic yield curve models capture the stylized fact. Moreover, Fisher (2001) also explores these effects on yield curve formation. Thus, the third dimension of our sentiment index relate to implied volatility priced by the market.

4.3.4 Monetary Policy Decision Surprises

Finally, still on yield curve modeling and in line with our goal of incorporating agents' expectations to our analysis, we must calculate the deviations of actual monetary policy decision from what was implied by market consensus embedded in interest rates. In order to

[6] One should note, thought, that despite being more closely related to the term premium, the slope of the term structure of interest rates is also influenced by the expectation of future monetary policy. This is specially relevant in times that the monetary authority assumes an extraordinary dovish stance, which usually makes the yield curve steeper, or when the policy stance in severely hawkish, which can even invert the yield curve, making long-term maturities price lower interest than their short-term counterparts.

[7] To put it differently, convexity reduces the module of the negative relation between price and yield as yields rise and increases its module as yields fall.

do that, we estimate interest rate decision implied by the yield curve by calculating the one month forward rate according to Eq. 4.2 and subtracting the current CDI at each monetary policy decision date. Our choice for the one month vertex comes from: (i) the fact that monetary policy meetings are usually spaced by 42 to 49 days intervals, thus guaranteeing that we are looking at forward rates expectations implied only in the next meeting; and (ii) despite not guaranteed that one month interval will comprehend treasury bonds, by not resorting to too short intervals we prevent distortions caused by the used vertex being much closer to current one day interest rates than to bonds yields. Monetary policy surprises are then calculated by subtracting the one month implied expectations from actual decisions, known after markets close, according to Eq. 4.6:

$$\Delta i_{t+1,0} - E[\Delta i_{t+1,0}] = (i_{t+1,0} - i_{t,0}) - (f_{t,1/12} - i_{t,0}) = i_{t+1,0} - f_{t,1/12}, \tag{4.6}$$

where $i_{t,0}$ is the monetary policy interest rate at time t, $\Delta i_{t+1,0}$ is the rate decision for meeting date t, which becomes effective in $t+1$, and $f_{t,1/12}$ is the one month forward rate priced by the yield curve at meeting date t. In our work, $i_{t,0}$ is given by the CDI series and $f_{t,1/12}$ is calculated according to Eq. 4.2. Surprise evolution by Copom meeting date is shown in Figure 4.2.

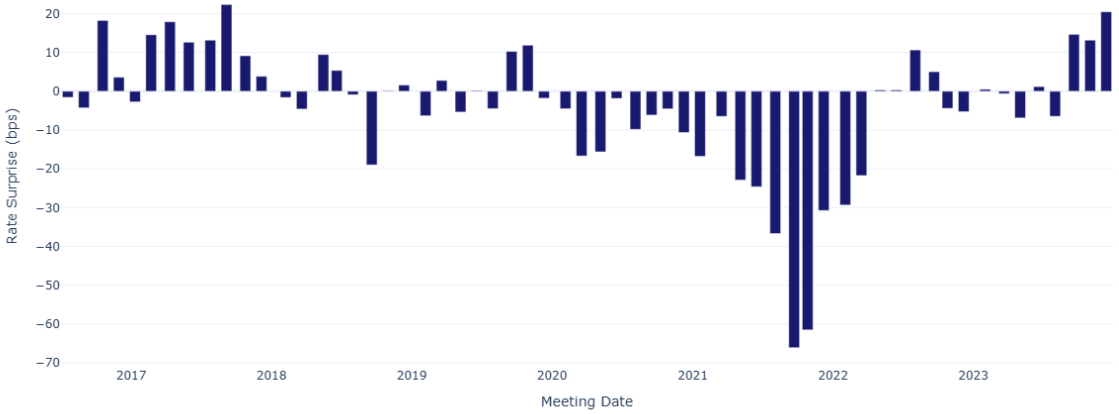


Figure 4.2 – Copom Monetary Decision Surprises

4.4 Monetary Policy Communication

For textual data we use English versions of the two documents issued by the Brazilian Central Bank for official monetary policy communication: decision statements[8] and com-

[8] https://www.bcb.gov.br/en/monetarypolicy/copomstatements/cronologicos.

mittee meeting minutes[9]. First, one should note that the use of English versions relates to the modeling choice, particularly regarding the transfer learning trait of transformer models. Models pre-trained on English datasets are the conventional approach in scientific research for many different fields, counting with more extensive investigation than their Portuguese counterparts. Therefore, using English versions of the documents allows us to leverage pre-trained models that have proven state-of-the-art language representation capabilities.

These documents are also the most usual choice when trying to assess the impacts of the Brazilian Central Bank communication over financial markets or economic variables. Chague *et al.* (2015) use Brazilian monetary policy meeting minutes to study the impact of communication over future interest rates and conclude that market participants incorporate the information disclosed by the monetary authority in these documents. Alves, Abraham, and Laurini (2023) use both statements and minutes to find that central bank communication improves forecasts of the term structure of interest rates. And Fasolo, Graminho, and Bastos (2022) also use both documents to investigate central bank tone related to different topics representing economic variables such as inflation and economic activity.

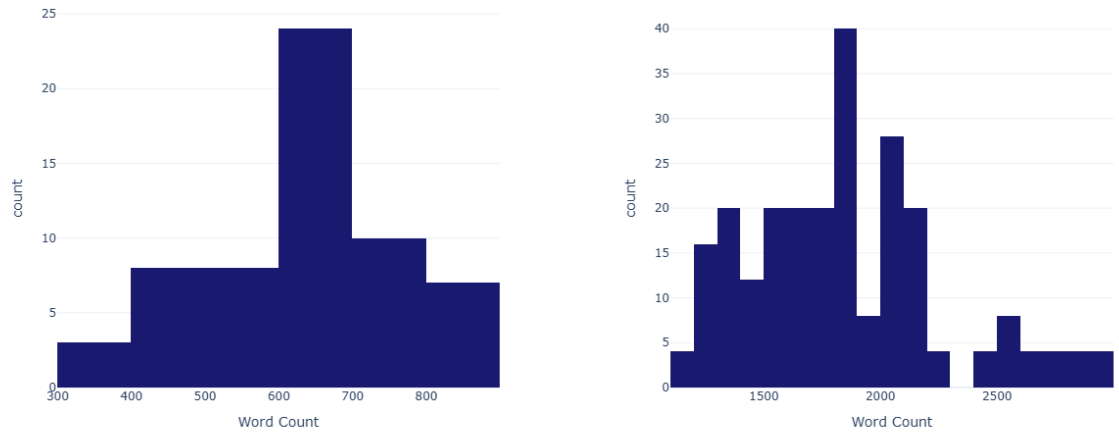
Our choice for Brazilian Central Bank communication for developing our framework is mainly due to the fact that meeting minute release takes place less than one week after monetary policy decision. It is an aspect crucial for measuring the state of communication expected by markets on minute publication date. This is a short period of time when compared to the practice of other monetary authorities, such as US Federal Reserve, which publishes minutes three weeks after the FOMC meeting. The shorter time span allows us to infer communication state without significant noise from discourse innovations between meeting date and minute release. Furthermore, both publications take place with markets closed: decision statements after trading hours and meeting minutes before markets open.

It is also interesting to note that, as Fasolo, Graminho, and Bastos (2022) point out, there has recently been a structural change in the extension and role of the two documents discussed here. In fact, the Brazilian Central Bank has modernized its communication starting from the 200th Copom meeting, on July 2016, in order to make it more effective in guiding markets' expectations. This remodeling affected both (i) the Copom decision statement, which became more comprehensive about the board's rationale in monetary decision making, including even information about Copom's inflation projections; and (ii) meeting minute, with in depth details of meetings discussions. Minutes also became organized in four sections: A) Update of economic outlook and Copom's scenario; B) Scenarios and risk analysis; C) Discussion about the conduct of monetary policy; D) Monetary policy decision.

Accounting for all of this, our textual dataset thus consists of the English versions of Brazilian Central Bank's monetary policy decision statements and meeting minutes from July

[9] https://www.bcb.gov.br/en/publications/copomminutes/cronologicos.

2016 to December 2024. We take statements as whole documents and minutes segregated by section. Figure 4.3 shows the distributions of text length for both document types.



(a) Statement Length Distribution

(b) Minute Length Distribution

Figure 4.3 – Document Length Distributions

4.5 Our Model Outline

In this Section we will explore the most relevant aspects of our framework, specifically the ones that let us understand the fundamentals that support each of the modeling choices. Formal specification of each step of our methodology is present in the Appendices.

Our model consists of a deep learning framework used to extract sentiment from central bank communication. This framework is unique as it explicitly incorporates economic fundamentals that usually lack on direct applications of standard "out of the box" deep learning techniques to monetary policy sentiment analysis. It does that by resorting to yield curve factors, thus resulting in an increased number of sentiment dimensions that respond to economic agents' reactions to monetary policy communication in terms of level, slope and curvature of the term structure of interest rates. Each of the dimensions are designed to capture different nuances of market participants' expectations embedded in asset prices. From the fundamentals of yield curve modeling discussed in Section 4.3.3, we see that sentiment measured along the dimension given by the level factor is related to monetary policy implementation concepts such as forward guidance and interest rate expectations for policy horizon. These are the concept usually captured by single-dimension approaches, common in the economic literature (Niţoi; Pochea; Radu, 2023)[10]. However, the two new

[10] As a matter of fact, the conventional hawk/dove gauge for monetary policy sentiment is a special case of our framework, achieved by making the number of factors $K = 1$.

factors further incorporate aspects derived from economically relevant concepts such as long-term inflation risk, closely connected to term premium and yield curve slope, and implied volatility, which by bond convexity effect shape the yield curve's curvature. Finally, we should also recall from Section 4.3.2 that our sentiment index is related to factor score shifts, as expressed by Eq. 4.5.

By building from yield curve factors, our framework also makes it straightforward to generate the necessary data for model estimation. Since factors are latent variables extracted from observable market prices, we can calculate the response of such factors to specific events, whose novel information is assumed to be incorporated by market participants in assets' values. This way, we can use the shifts in factors as a response to the disclosure of monetary policy communication in order to measure how economic agents reacted to it. In turn, this aspect of our approach addresses one of the main challenges of model estimation for sentiment analysis: data labeling. This challenge is one of special relevance to the monetary policy sentiment domain (Pfeifer; Marohl, 2023). Compared to the usual solution, which involves estimating model's parameters with textual data manually annotated with a sentiment gauge based on individual perceptions (Araci, 2019; Niţoi; Pochea; Radu, 2023; Pfeifer; Marohl, 2023), the price-driven approach makes the process more objective and economically coherent. Furthermore, it also contributes to cost reduction in the data generation process. Since the datasets needed for deep learning models' estimation are considerably large, manual annotation requires several working hours of one or more experts, as it is usually interesting to somehow validate the results. Therefore, having a systematic solution can significantly reduce the cost of data for model training. We discuss the details of our data labeling strategy in Appendix C.

However, working with the extraction of market sentiment from shifts in factor pricing requires some special attention. On the one hand, market participants trading yield curve assets are indeed expected to respond to monetary policy communication. On the other hand, they are also expected to respond to other information that may influence their perceptions on such communication. Consequently, this additional information has to be explicitly modeled in order not to bias the model's results. Accounting for that, we designed our framework as a two-stage neural network setup, with each stage designed to extract sentiment from a specific monetary policy communication instrument, namely, policy decision statements and meeting minutes.

The first stage neural network was designed for policy decision statements. In this stage, we also incorporate the most important information for yield curve pricing in the event of monetary policy decisions: the rate decision itself. More than that, it is the deviation from what market participants anticipated for the decision that should move prices. Thus, the first stage's inputs are the document we want to score the sentiment for, i.e., policy decision statements; and the policy rate decision surprise as described in Section 4.3.4. This

way, we make sure that the neural network model incorporates the most relevant pieces of information and learns patterns from data to extract the sentiment in policy communication that moves markets. The main output of the first stage is the multi-dimensional sentiment extracted from monetary policy decision statement. Detailed neural network design for the first stage is discussed in Appendix B.

Before moving on to the second stage network, which focuses on policy meeting minutes, one should note that there are no specific policy events that are recurrently associated with minute publication the way rates decisions were associated with statements. However, the short time interval between statement and minute releases for the Brazilian case makes it possible to draw a link between them. The Brazilian central bank publishes the two documents only six days apart from each other, with only three market sessions in between. Furthermore, both documents provide different levels of information about the same event, the monetary policy interest rate decision. It is thus reasonable to assume that market participants expect an a priori similar tone from minute as the one from statement. As a result, yield curve factors should move in response to surprises in the tone of novel information contained in the minute. As a matter of fact, by making release dates of statements and minutes that close, the Brazilian policy framework provides an interesting opportunity to inform the second stage network what market participants expect in terms of communication. And again, this is also one of the main reasons for choosing this economy's data to support our model development.

In order to incorporate that into the model, we resort to the assessment that the first stage model performed on decision statements. The set of information contained in these documents, assumed to be what market participants may anticipate for minute and which we will refer to as communication state, is then extracted as an additional output from first stage network. The communication state is hence passed as input to the second stage, alongside with the policy meeting minute we want to extract sentiment from. In turn, these inputs, associated with the neural network design detailed in Appendix B, are expected to make the model learn how to identify the differences in speeches that drive factor shifts, in order to better measure sentiment in the form of economic agents' reactions. The output of the second stage is again our multi-dimensional sentiment index, but now extracted from the novel in monetary policy meeting minutes.

One should also note that this setup makes the second stage neural network a general model to update communication state and score novel information, independent of whether it is an official monetary meeting minute or some manifestation of the monetary authority board members in the form of speeches or interviews. In turn, this aspect highlights that our framework can be used in an iterative setup to continuously update the communication state and extract sentiment from novel monetary policy communication, independent of its format. However, in spite of this flexibility of our proposed framework, we focus this paper

on official communication in the form of statements and minutes.

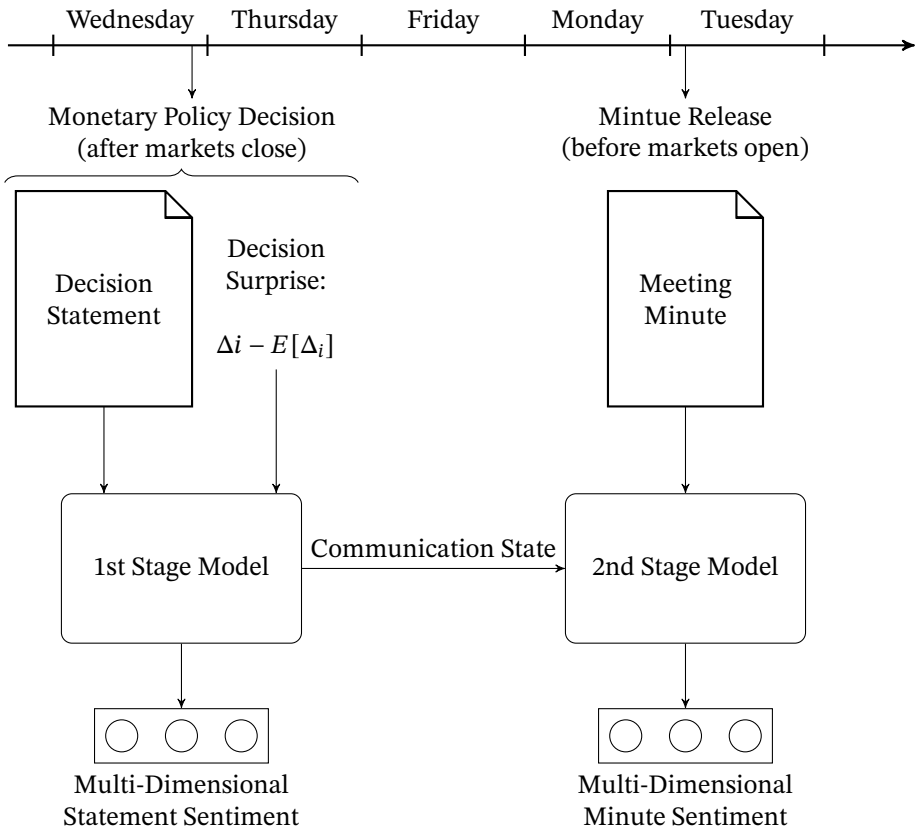


Figure 4.4 – Multi-Dimensional Sentiment Framework

Figure 4.4 illustrates our framework, accounting for all the effects discussed thus far. It details the one week interval which encompasses monetary policy decision and meeting minute release. On its left-hand side, first stage model receives both inputs from monetary policy decision — the statement and the surprise as defined by Eq. 4.6 — and outputs our multi-dimensional sentiment index for decision statement. It further outputs the communication state, passed along to the second stage. Second stage also receives monetary policy meeting minute as input, in order to output again our multi-dimensional sentiment index.

Finally, still on the discussion about our monetary policy sentiment framework, it is interesting to understand the implications of our model of choice for textual data representation, the state-of-the-art transformer model: BERT. BERT works by representing each word in the analyzed document as a vector in a high dimension vector space. We expect these dimensions to capture different aspects of the word's semantic or grammatical properties, resulting from a process known as word embedding (Gentzkow; Kelly; Taddy, 2019). Furthermore, to be more precise, BERT works with contextual embeddings, which depend on the overall context of communication. This approach provides improved language representation, delivering state-of-the-art performance in many natural language processing tasks, with sentiment analysis being one of them. BERT also provides a unique embedding vector

representation for the meaning of the entire document. Analogous to word embeddings, this high dimensional vector captures various aspects of the documents, many of which are expected to help the deep learning network extract sentiment along our three yield curve factor dimensions. Hence, BERT's document embeddings are the perfect candidates for our measure of communication state, which is passed from one stage network to the next, as Figure 4.4 shows. This vector, representing what market participants already expect the tone of communication to be upon minute release, helps the second stage network to infer what has altered in communication so it better understands what caused yield curve shifts[11] [12]. Details of how BERT integrates our model are provided in Appendix A.

4.6 Results

We now get to the analysis of the results of our multi-dimensional sentiment index. Given that we experimented with several alternative model setups (detailed in Appendix D), and considering the training metrics and how the model assimilates information, we focus our attention on the sentiment index generated by our two stage framework in a regression setup. The two stage approach delivers better training statistics, whilst the regression configuration, by leveraging continuous variables to capture the magnitude of yield curve movement represented by its three factors, generates sentiment more responsive to market movements. In spite of the regression approach being a more sophisticated problem for the deep learning model to deal with, again the quality of training metrics discussed in Appendix D supports this setup.

For result evaluation, we turn to the still unused data from the eight Brazilian monetary policy meetings during 2024. This dataset was not employed in the training phase, not even as validation data split used to calculate out-of-sample statistics. This way, we can assess our multi-dimensional sentiment index behavior when faced with new data, as its intended use. Moreover, as an additional characteristic of the proposed application of our deep learning framework to score sentiment from monetary policy communication on a promptly fashion, we resort to our latest iteration of the language model, BERT, to score new textual data. In practice, we replace the BERT neural network used in each of our framework's stages

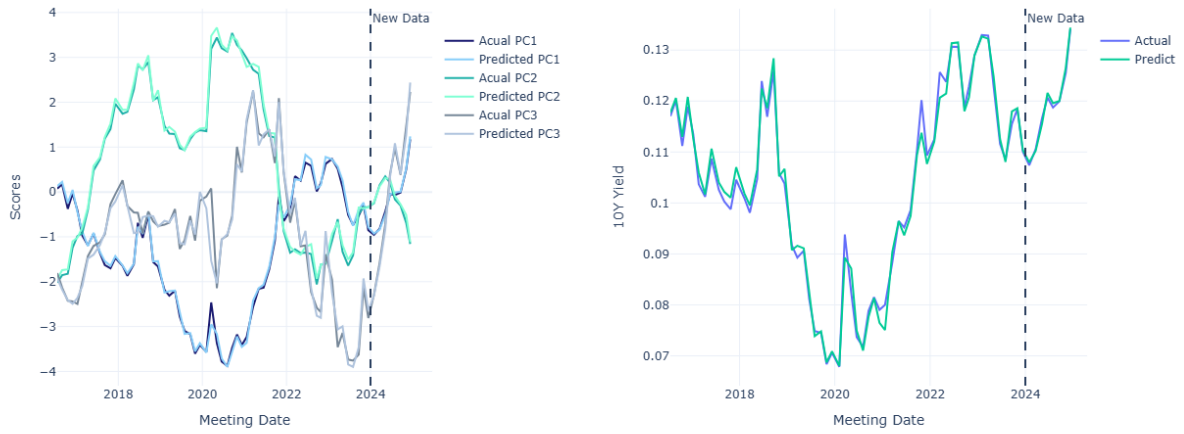
[11] To illustrate the reason for incorporating this into our framework, imagine the case that the overall communication, independent of having positive or negative tone, remains the same between first and second stage. Then, we expect market prices also to remain the same, as there is no novel information to move them. Passing communication state between the two stages allows the deep neural network to learn this.

[12] Interesting to note that this modeling decision had its inspirations on recurrent neural networks (Jordan, 1986; Elman, 1990) that, besides generating an output at each stage of the recurrent unit, also passes the latent state to the next iteration. The difference is that we work only with two stages for data restrictions (despite being straightforward to further extend the framework) and use this exclusively to better capture fundamental changes in communication tone. From this perspective, our use of communication state is analogous to time series latent state space models (Hamilton, 2020).

by the final version of BERT for the second stage, fine-tuned during training step. This assures that the best performant version of the model is used, while still taking care for not using future information, not even to train the model's parameters. Finally, we do not further fine-tune our deep learning framework on new data for the 2024 meetings as they come to pass. This would probably make the model even more up-to-date with the language used by the monetary authority, incorporating new aspects regarding factors such as policy committee board composition or novel economic outlook, to name a few. However, as this would possibly alter the baseline result assessment, we decided to leave it to be explored by future works.

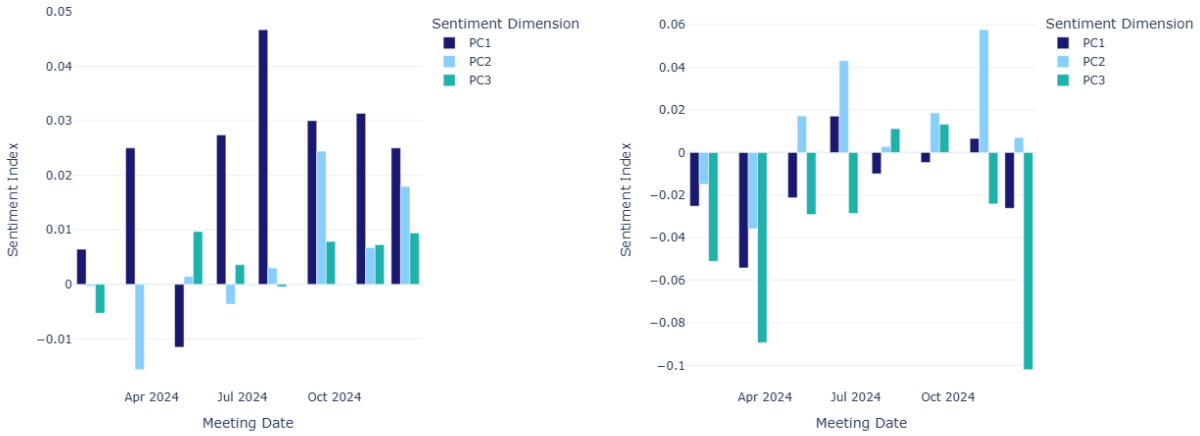
Before looking into the sentiment index itself, Figure 4.5 shows the main results in terms of regression predictions, which illustrate our strategy to infer sentiment from policy communication. One should recall from Section 4.3.2 and Eq. 4.5 that our sentiment index is related to factor score shifts. We can thus construct an expected factor score series, derived from our sentiment measure, by rearranging Eq. 4.5 and using our sentiment index output as expected $\Delta z_{k,t}$, for t in communication release dates. Panel 4.5a draws the evolution of the scores for each of the yield curve's principal component factors, highlighting the period with new data and comparing actual to predicted series. Important to note that actual factor scores series are filtered to match only the dates to which we have predictions, consequently not displaying their evolution between monetary policy meetings. We can see that the model's outputs follows observed data fairly close. Additionally, in order take the assessment closer to the intuition of market prices, Panel 4.5b shows predictions and actual values of the ten-year vertex of the yield curve, a key maturity frequently monitored by market participants. The same adjustment is made to the yield series so it matches dates of policy communication. Furthermore, these yield predictions are yield reconstructions from factor scores estimated by our model, calculated by rearranging Eq. 4.4. Again, our results follow realized data closely.

Figure 4.6 then shows the evolution of the three dimensions of our sentiment index for the eight 2024 Brazilian monetary policy meetings. Panel 4.6a focus on sentiment from policy decision statements, extracted using the first stage model, and Panel 4.6b is dedicated to the sentiment index from meeting minutes, calculated by the second stage. Again from the discussion on Section 4.3.2 and Eq. 4.5, we see that these series should be interpreted as sentiment increment from novel disclosed information by the monetary authority, given that these indexes are related to price shifts along the three dimensions of yield curve factors. Moreover, one should recall, from the nature of yield curve and how its factors capture its movements as also discussed in Section 4.3.2, that factor score increase represents negative sentiment. As being measured along yield dimensions, factor increase result in higher traded yield, in turn resulting in depreciation of financial assets. Hence, we can see that factor score reduction analogously represents positive sentiment.



(a) Factor Scores Prediction (b) 10 year yield prediction

Figure 4.5 – Regression Model Predictions



(a) Sentiment Index - Statements (b) Sentiment Index - Minutes

Figure 4.6 – Three dimensional sentiment index

First, we can see clearly from Panel 4.6a that the 2024 meetings were dominated by a hawkish tone along the first dimension of decision statements, which relates to overall sentiment about policy stance. All but one meeting had the level dimension assuming positive values, also presenting larger modules than other factors on all of such occasions. As a matter of fact, 2024 was marked by a relevant pivot in Brazilian monetary policy stance. Whilst the year began with the central bank conducting an easing cycle, which was then priced by the market to at least bring the interest rate back to around its neutral level, by the last meeting of the first half of the year the cycle had to be halted after a pace reduction in the previous decision. The monetary authority then started raising the Brazilian basic

interest rate, Selic, by the sixth decision of the year, engaging in a tightening cycle that gained momentum meeting after meeting, ultimately leading the Selic rate to close 2024 on higher levels than it has begun.

Comparing that behavior with minutes' sentiment, shown in Panel 4.6b, we can see two direct effects of our modeling choices. First, the sentiment along the first dimension did not show all the strength captured from statements by our first stage neural network. Indeed, the sentiment associated with the level factor for policy meeting minutes did not present clear direction in the analyzed period, slightly reinforcing negative tone at some occasions and partially correcting it at others. This shows how incorporating communication state affects sentiment extracted from novel information. Recalling that both documents are separated by a short period of time — not to mention that they are referenced in the same event: the monetary policy meeting and its rate decision — it is straightforward to note that novel information in meeting minutes should not regularly produce significant innovations in policy stance communication. Therefore, passing communication state as a feature for the second stage deep learning model seems to make it able to focus only on novel information, relevant to market sentiment increments. This leads us to the second interesting effect of our modeling choices.

Once able to focus on novel information for tone extraction from policy communication, our model results highlight the importance of the added dimensions to perform comprehensive sentiment analysis. Panel 4.6b shows that variations in level sentiment are usually less intense when compared with slope and curvature variations in the second stage model. Moreover, they are clearly less relevant in second stage than they were in the first stage, which does not receive previous communication state as input.

In order to understand what might have influenced these outcomes, we now turn to explore the actual documents that resulted in the sentiment illustrated in Figure 4.6. Important to note that actual effects from text input over sentiment cannot be precisely calculated given the nature of deep learning neural networks. We can, however, understand what economic intuition suggests for the tone of communication, and face it with model output. First, we will explore how monetary authority responded to one of the main challenges it face in 2024, and how its communication evolved accordingly. The challenge in question was the exhaustion of the desinflationary trend that backed the begging of the easing cycle in 2023 and the resurgence of considerable inflationary risks that forced the central bank to revert its policy stance. In order to analyze this, we will resort to the minutes of the second and the fourth monetary policy meetings of 2024.

Beginning by the March meeting, the second of the year[13], we find ourselves in a time when both the Brazilian Monetary Policy Committee (Copom) and market participants

[13] The minute for the March meeting is available at https://www.bcb.gov.br/en/publications/copomminutes/20032024.

believed that the desinflationary trend in prices would extend further ahead in time. Back then, we had an overall positive tone in policy communication, specially from minute sentiment. One of the paragraphs in this minute's Section D, "Monetary policy decision", summarized the somewhat positive, despite cautious, tone delivered by this communication. This excerpt even provided forward guidance in line with the continuity of the easing cycle:

> "27. The committee judges that the baseline scenario has not changed substantially. Due to heightened uncertainty and the need for more flexibility in the conduct of monetary policy, the Committee members unanimously decided to communicate that, if the scenario evolves as expected, they anticipate a reduction of the same magnitude in the next meeting. The Committee judges that this monetary policy stance is appropriate to keep the necessary contractionary monetary policy for the disinflationary process."

Besides this passage, it is specially interesting to see how the communication in Section B evolved, since this Section is the one closest related to long term inflation risk, an usually important driver for the second dimension of our sentiment gauge. The following excerpt shows one of the concluding remarks for this section in the June meeting minute[14], in which the monetary authority highlights longer-term inflation perspectives deteriorating.

> "14. Copom concluded by assessing that the inflation outlook has become more challenging, with the increase of medium-term inflation projections, even conditioned on a higher interest rate. Benign surprises were noted in the recent period, but also the rise in projections for shorter periods, involving market prices. In the end, Copom unanimously concluded that a more contractionary and more cautious monetary policy was needed to reinforce the disinflationary dynamics."

These passages illustrate the dynamics seen in communication contained in meeting minutes for the first half of 2024, which went from a cautious positive tone to show some sharp increases in pessimism, specially captured by the second dimension of our sentiment measure. This is an interesting fact, considering that longer-term inflation expectations, along with its medium-term projections, were the main driver for the need of the policy stance pivot as seen by mid-year. These are economic variables that usually affect the slope of the yield curve, which in turn had led the increase in tone negativity, being followed closely by the first principal component measure. From Figure 4.6b we can see that this pattern was particularly present in the time interval between the March and June Copom meetings.

[14] The minute for the June meeting is available at https://www.bcb.gov.br/en/publications/copomminutes/19062024.

Further investigating actual policy communication documents and the intuition provided by them to our proposed sentiment measure leads us to the tightening cycle started in the second half of 2024. During this period, maybe the most relevant communication is in the last meeting of the year, when pace of interest rate hikes was significantly altered. The following passage comes from the concluding remarks of the December meeting decision statement[15]:

> "Copom therefore decided to increase the Selic rate by 1.00 p.p. to 12.25% p.a., and judges that this decision is consistent with the strategy for inflation convergence to a level around its target throughout the relevant horizon for monetary policy. Without compromising its fundamental objective of ensuring price stability, this decision also implies smoothing economic fluctuations and fostering full employment.
>
> In light of a more adverse scenario for inflation convergence, the Committee anticipates further adjustments of the same magnitude in the next two meetings, if the scenario evolves as expected. The total magnitude of the tightening cycle will be determined by the firm commitment of reaching the inflation target and will depend on the inflation dynamics, especially the components that are more sensitive to monetary policy and economic activity, on the inflation projections, on the inflation expectations, on the output gap, and on the balance of risks."

To be fair, we believe this passage should have been associated with the most hawkish sentiment along the first dimension for the entire 2024 communication dataset studied. In fact, not only did it double the pace of monetary tightening from 50 bps to 100 bps, it also anticipated two more hikes of the same magnitude in the following meetings. Nonetheless, Panel 4.6a shows it was still very hawkish, in line with most of the communication for the period. Besides, when adding the three dimensions' sentiment in response to this communication, we indeed see one of the most negative tones for the year, again reinforcing the importance of analyzing higher dimension sentiment.

Finally, concluding the analysis on communication from the last monetary policy meeting of 2024, we can see from Panel 4.6b that sentiment had a considerable decompression on the curvature dimension with this meeting's minute tone. We should then recall that the third principal component usually responds to interest rate implied volatility. We then turn to this minute's Section B, "Scenarios and risk analysis", which, among its final paragraphs, mentions the following[16]:

[15] The statement for the December meeting is available at https://www.bcb.gov.br/en/monetarypolicy/copomstatements/2584.

[16] The minute for the December meeting is available at https://www.bcb.gov.br/en/publications/copomminutes/11122024.

> "18. Due to the materialization of risks, the Committee judges that the scenario is less uncertain and more adverse than in the previous meeting. Upside inflation risks, such as the resilience of services inflation, the deanchoring of expectations, and exchange rate depreciation, have materialized. As a result, a scenario that until then had been quite uncertain has become more adverse."

The monetary authority thus recognized the materialization of risks, possibly contribution to the slight deterioration along the second dimension of our sentiment index. However, they also affirmed that they see a scenario "less uncertain". This aspect, associated with the use of forward guidance to anchor the following decisions, further reinforced by the highlighted consensus present in both statement and minute for this particular meeting, might have contributed to the significant positive perception related to implied volatility by market participants.

Therefore, all things considered, the empirical examples discussed in this Section suggest that our multi-dimensional sentiment index is successful in measuring some of the most relevant aspects of monetary policy communication and how market sentiment responds to it. Besides, we also found evidence that both of our most meaningful modeling decisions, namely to incorporate higher dimension sentiment and to adopt a tow-stage design in order to provide the model with means to update communication state, contribute to the enhanced performance.

4.7 Conclusion

In this paper, we proposed a novel multi-dimensional measure of sentiment extracted from the Brazilian Central Bank communication. We designed this measure to incorporate relevant aspects of economic agents' expectation shifts in response to innovations in monetary authority communication that are not incorporated in usual single-dimension sentiment analysis. We thus develop a deep learning neural network model to calculate this sentiment index in a practical and iterative fashion. Its empirical assessment shows that additional dimensions are indeed necessary for a comprehensive gauge of monetary policy sentiment.

In addition to economic agents interested in objectively measuring the tone of central bank communication, one interesting practical use of our index is by policymakers themselves. Our framework can be used prior to communication release to provide an objective, expectation-driven measure of how market participants tend to interpret the novel communication. This is specially useful for incorporating higher dimensional sentiment, as the added dimensions might hide communication side effects that are not obvious at first sight. For example, in times of need for expansionary policy, the monetary authority could use our framework to calibrate the discourse not to let a too lenient tone result in increased inflationary risks, resulting in pressures over the yield curve's slope and possibly the unwanted side

effect of monetary tightening by rising long yields. Note that our sentiment index's second dimension captures yield curve's slope, making that analysis possible while single dimension sentiment indexes would probably neglect it. Therefore, knowing objectively a priori how market participants will react to communication can help the policymaker calibrate the tone of the message exactly to what is intended, making communication a more effective and less noisy policy instrument.

Moreover, despite not being the main goal of our work, it is straightforward from our training strategy, derived from market prices, to also see practical application of our model to short-term yield curve prediction. As a matter of fact, we can see that our framework delivers very good out-of-sample performance to one-day term structure factor forecasts. Following that path, future work could further explore this aspect, adapting the model for longer-term predictions. Such adaptations would probably need to involve the input of other macroeconomic fundamentals, as inflation expectations and output gap projections, among other economic indicators associated with anticipated monetary policy and yield curve term premium.

We also see space for improving our communication state measuring mechanism. Our proposed discrete Markovian approach of updating communication and passing only the previous state to forward iterations could be replaced by a more sophisticated recurrent neural network model, such as long short-term memory neural networks (LSTM). This alternative approach would be able to infer long-term dependencies by design, in an unified neural network framework which could be trained to capture long lasting influences of monetary authority communication in agents' expectations and market pricing reactions. However, as such new development would significantly increase complexity of the framework, potentially shifting the core of discussions to deep learning neural network design, we leave that improvement for future works.

References

ACHARYA, S.; PESENTI, P. A. **Spillovers and Spillbacks**. [*S.l.*], 2024. Cit. on p. 67.

ADRIAN, T.; CRUMP, R. K.; MOENCH, E. Pricing the term structure with linear regressions. **Journal of Financial Economics**, Elsevier, v. 110, n. 1, p. 110–138, 2013. Cit. on pp. 96 and 103.

AGARWAL, S. Stock market response to information diffusion through internet sources_ A literature review. **International Journal of Information Management**, 2019. Cit. on pp. 24 and 54.

AHERN, K. R.; SOSYURA, D. Who Writes the News? Corporate Press Releases during Merger Negotiations. **The Journal of Finance**, v. 69, n. 1, p. 241–291, Feb. 2014. ISSN 0022-1082, 1540-6261. DOI 10.1111/jofi.12109. Cit. on p. 57.

ALGABA, A.; ARDIA, D.; BLUTEAU, K.; BORMS, S.; BOUDT, K. Econometrics meets sentiment: An overview of methodology and applications. **Journal of Economic Surveys**, v. 34, n. 3, p. 512–547, 2020. DOI 10.1111/joes.12370. Cit. on pp. 19 and 23.

ALMEIDA, C. I. R. D.; JR, A. M. D.; FERNANDES, C. A. C. A generalization of principal component analysis for non-observable term structures in emerging markets. **International Journal of Theoretical and Applied Finance**, World Scientific, v. 6, n. 08, p. 885–903, 2003. Cit. on p. 95.

ALVES, C. R. D. A.; ABRAHAM, K. J.; LAURINI, M. P. Can Brazilian Central Bank communication help to predict the yield curve? **Journal of Forecasting**, v. 42, n. 6, p. 1429–1444, Sep. 2023. ISSN 0277-6693, 1099-131X. DOI 10.1002/for.2964. Cit. on pp. 97 and 106.

AMARAL, P. Monetary Policy Tightening and Long-Term Interest Rates. **Economic Commentary (Federal Reserve Bank of Cleveland)**, p. 1–6, Jul. 2013. ISSN 2163-3738, 0428-1276. DOI 10.26509/frbc-ec-201308. Cit. on p. 58.

ANG, A.; PIAZZESI, M. A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables. **Journal of Monetary Economics**, v. 50, n. 4, p. 745–787, May 2003. ISSN 03043932. DOI 10.1016/S0304-3932(03)00032-1. Cit. on pp. 96 and 103.

ANGELICO, C.; MARCUCCI, J.; MICCOLI, M.; QUARTA, F. Can we measure inflation expectations using Twitter? **Journal of Econometrics**, Elsevier, v. 228, n. 2, p. 259–277, 2022. Cit. on p. 28.

APEL, M.; GRIMALDI, M. B.; HULL, I. How Much Information Do Monetary Policy Committees Disclose? Evidence from the FOMC's Minutes and Transcripts. **Journal of**

Money, Credit and Banking, v. 54, n. 5, p. 1459–1490, Aug. 2022. ISSN 0022-2879, 1538-4616. DOI 10.1111/jmcb.12885. Cit. on pp. 33 and 96.

ARACI, D. **FinBERT: Financial Sentiment Analysis with Pre-trained Language Models**. [*S.l.*]: arXiv, 2019. Cit. on pp. 48, 49, 55, 93, 97, 98, and 108.

ARMELIUS, H.; BERTSCH, C.; HULL, I.; ZHANG, X. Spread the Word: International spillovers from central bank communication. **Journal of International Money and Finance**, 2020. Cit. on pp. 63 and 69.

ARRIETA, A. B.; Díaz-Rodríguez, N.; SER, J. D.; BENNETOT, A.; TABIK, S.; BARBADO, A.; GARCIA, S.; Gil-Lopez, S.; MOLINA, D.; BENJAMINS, R.; CHATILA, R.; HERRERA, F. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. **Information Fusion**, v. 58, p. 82–115, Jun. 2020. ISSN 15662535. DOI 10.1016/j.inffus.2019.12.012. Cit. on p. 29.

ASH, E.; HANSEN, S. Text Algorithms in Economics. **Annual Review of Economics**, v. 15, n. 1, p. 659–688, 2023. DOI 10.1146/annurev-economics-082222-074352. Cit. on pp. 18 and 19.

Baeza-Yates, R.; Ribeiro-Neto, B. **Modern Information Retrieval**. 2. ed. USA: Addison-Wesley Publishing Company, 2008. ISBN 978-0-321-41691-9. Cit. on p. 34.

BAHDANAU, D.; CHO, K.; BENGIO, Y. **Neural Machine Translation by Jointly Learning to Align and Translate**. [*S.l.*]: arXiv, 2014. Cit. on pp. 22 and 43.

BAKER, S. R.; BLOOM, N.; DAVIS, S. J. Measuring Economic Policy Uncertainty. **The Quarterly Journal of Economics**, v. 131, n. 4, p. 1593–1636, Nov. 2016. ISSN 1531-4650, 0033-5533. DOI 10.1093/qje/qjw024. Cit. on p. 57.

BALL, L. The performance of alternative monetary regimes. *In*: **Handbook of Monetary Economics**. [*S.l.*]: Elsevier, 2010. v. 3, p. 1303–1343. Cit. on p. 69.

Bank for International Settlements. **Central Bank Policy Rates, BIS WS_CBPOL 1.0**. 2024. Cit. on p. 69.

Bank for International Settlements. **Central Bank Total Assets, BIS WS_CBTA 1.0**. 2024. Cit. on p. 70.

BATTISTON, STEFANO.; WEISBUCH, GÉRARD.; BONABEAU, ERIC. Decision spread in the corporate board network. **Advances in Complex Systems**, v. 06, n. 04, p. 631–644, 2003. Cit. on p. 64.

BELTAGY, I.; PETERS, M. E.; COHAN, A. **Longformer: The Long-Document Transformer**. [*S.l.*]: arXiv, 2020. Cit. on p. 48.

BENGIO, Y.; FRASCONI, P.; SIMARD, P. The problem of learning long-term dependencies in recurrent networks. *In*: **IEEE International Conference on Neural Networks**. [*S.l.*]: IEEE, 1993. p. 1183–1188. Cit. on p. 41.

BENIGNO, G.; BENIGNO, P. Price stability in open economies. **The Review of Economic Studies**, Wiley-Blackwell, v. 70, n. 4, p. 743–764, 2003. Cit. on pp. 65 and 66.

BENIGNO, G.; BENIGNO, P. Designing targeting rules for international monetary policy cooperation. **Journal of Monetary Economics**, Elsevier, v. 53, n. 3, p. 473–506, 2006. Cit. on pp. 65 and 66.

BENIGNO, G.; BENIGNO, P. Implementing international monetary cooperation through inflation targeting. **Macroeconomic Dynamics**, Cambridge University Press, v. 12, n. S1, p. 45–59, 2008. Cit. on pp. 65 and 66.

BERKSON, J. Application of the logistic function to bio-assay. **Journal of the American statistical association**, Taylor & Francis, v. 39, n. 227, p. 357–365, 1944. Cit. on pp. 22 and 31.

BERKSON, J. Why I prefer logits to probits. **Biometrics. Journal of the International Biometric Society**, JSTOR, v. 7, n. 4, p. 327–339, 1951. Cit. on pp. 22 and 31.

BIANCHI, J.; COULIBALY, L. **Financial Integration and Monetary Policy Coordination**. [*S.l.*], 2024. Cit. on p. 67.

BISHOP, C. M.; NASRABADI, N. M. **Pattern Recognition and Machine Learning**. [*S.l.*]: Springer, 2006. v. 4. Cit. on p. 32.

BLANCHARD, O.; OSTRY, J. D.; GHOSH, A. R. International policy coordination: The Loch Ness monster. **International Monetary Fund, IMFDirect**, 2013. Cit. on pp. 84 and 86.

BLEI, D. M.; NG, A. Y.; JORDAN, M. I. Latent Dirichlet Allocation. **Journal of machine Learning research**, v. 3, n. Jan, p. 993–1022, 2003. Cit. on pp. 24 and 34.

BLINDER, A. S.; EHRMANN, M.; FRATZSCHER, M.; HAAN, J. D.; JANSEN, D.-J. Central bank communication and monetary policy: A survey of theory and evidence. **Journal of economic literature**, American Economic Association, v. 46, n. 4, p. 910–945, 2008. Cit. on pp. 68 and 102.

BLINDER, A. S.; EHRMANN, M.; HAAN, J. D.; JANSEN, D.-J. Central bank communication with the general public: Promise or false hope? **Journal of Economic Literature**, American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203-2425, v. 62, n. 2, p. 425–457, 2024. Cit. on p. 68.

BLONDEL, V. D.; GUILLAUME, J.-L.; LAMBIOTTE, R.; LEFEBVRE, E. Fast unfolding of communities in large networks. **Journal of Statistical Mechanics: Theory and**

Experiment, v. 2008, n. 10, p. P10008, Oct. 2008. ISSN 1742-5468. DOI 10.1088/1742-5468/2008/10/P10008. Cit. on pp. 62, 70, 73, and 77.

BODENSTEIN, M.; CORSETTI, G.; GUERRIERI, L. The elusive gains from nationally oriented monetary policy. **Review of Economic Studies**, Oxford University Press, p. rdae046, 2024. Cit. on p. 67.

BOECK, M.; FELDKIRCHER, M. The impact of monetary policy on yield curve expectations. **Journal of Economic Behavior & Organization**, Elsevier, v. 191, p. 887–901, 2021. Cit. on p. 96.

BOLLEN, J.; MAO, H.; ZENG, X. Twitter mood predicts the stock market. **Journal of Computational Science**, v. 2, n. 1, p. 1–8, Mar. 2011. ISSN 18777503. DOI 10.1016/j.jocs.2010.12.007. Cit. on p. 53.

BONANNO, G.; VANDEWALLE, N.; MANTEGNA, R. N. Taxonomy of stock market indices. **Physical Review E: Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics**, American Physical Society, v. 62, n. 6, p. R7615–R7618, Dec. 2000. DOI 10.1103/PhysRevE.62.R7615. Cit. on p. 64.

BORDO, M. D. Monetary policy cooperation/coordination and global financial crises in historical perspective. **Open economies review**, Springer, v. 32, n. 3, p. 587–611, 2021. Cit. on p. 64.

BORUP, D.; HANSEN, J. W.; LIENGAARD, B. D.; SCHUETTE, E. C. M. Quantifying investor narratives and their role during COVID-19. **Journal of Applied Econometrics**, Wiley Online Library, v. 38, n. 4, p. 512–532, 2023. Cit. on pp. 24 and 57.

BOUKUS, E.; ROSENBERG, J. V. The Information Content of FOMC Minutes. **SSRN Electronic Journal**, 2006. ISSN 1556-5068. DOI 10.2139/ssrn.922312. Cit. on pp. 24, 28, 56, and 96.

BREIMAN, L. Random forests. **Machine learning**, Springer, v. 45, n. 1, p. 5–32, 2001. Cit. on p. 31.

CABOT, P.-L. H.; NAVIGLI, R. REBEL: Relation Extraction By End-to-end Language generation. *In*: MOENS, M.-F.; HUANG, X.; SPECIA, L.; YIH, S. W.-t. (Ed.). **Findings of the Association for Computational Linguistics: EMNLP 2021**. Punta Cana, Dominican Republic: Association for Computational Linguistics, 2021. p. 2370–2381. DOI 10.18653/v1/2021.findings-emnlp.204. Cit. on pp. 29 and 30.

CAJUEIRO, D. O.; BASTOS, S. B.; PEREIRA, C. C.; ANDRADE, R. F. S. A model of indirect contagion based on a news similarity network. **Journal of Complex Networks**, v. 9, n. 5, p. cnab035, Sep. 2021. ISSN 2051-1310, 2051-1329. DOI 10.1093/comnet/cnab035. Cit. on pp. 5, 7, 15, 31, 56, 62, 70, 72, and 89.

CAJUEIRO, D. O.; DIVINO, J. A.; TABAK, B. M. Forecasting the Yield Curve for Brazil. **Central Bank of Brazil Working Paper Series**, v. 197, 2009. Cit. on p. 96.

CAJUEIRO, D. O.; NERY, A. G.; TAVARES, I.; MELO, M. K. D.; REIS, S. A. dos; WEIGANG, L.; CELESTINO, V. R. R. **A Comprehensive Review of Automatic Text Summarization Techniques: Method, Data, Evaluation and Coding**. [*S.l.*]: arXiv, 2023. Cit. on pp. 30 and 52.

CALDARA, D.; FERRANTE, F.; IACOVIELLO, M.; PRESTIPINO, A.; QUERALTO, A. The international spillovers of synchronous monetary tightening. **Journal of Monetary Economics**, Elsevier, v. 141, p. 127–152, 2024. Cit. on pp. 67 and 89.

CALDARELLI, G.; CHESSA, A. **Data Science and Complex Networks: Real Cases Studies with Python**. [*S.l.*]: Oxford University Press, 2016. Cit. on p. 64.

CALOMIRIS, C. W.; MAMAYSKY, H. How news and its context drive risk and returns around the world. **Journal of Financial Economics**, v. 133, n. 2, p. 299–336, Aug. 2019. ISSN 0304405X. DOI 10.1016/j.jfineco.2018.11.009. Cit. on p. 54.

CAMPBELL, J. R.; EVANS, C. L.; FISHER, J. D.; JUSTINIANO, A. *et al.* Macroeconomic effects of federal reserve forward guidance. **Brookings Papers on Economic Activity**, Economic Studies Program, The Brookings Institution, v. 43, n. 1 (Spring), p. 1–80, 2012. Cit. on p. 82.

CAMPBELL, J. Y.; SHILLER, R. J. Yield spreads and interest rate movements: A bird's eye view. **The Review of Economic Studies**, Wiley-Blackwell, v. 58, n. 3, p. 495–514, 1991. Cit. on p. 103.

CAN, U.; SALTIK, O.; CAN, Z. G.; DEGIRMEN, S. Evaluation of international monetary policy coordination: Evidence from machine learning algorithms. **Computational Economics**, Springer, p. 1–26, 2024. Cit. on p. 63.

CANZONERI, M. B.; CUMBY, R. E.; DIBA, B. T. The need for international policy coordination: What's old, what's new, what's yet to come? **Journal of International Economics**, Elsevier, v. 66, n. 2, p. 363–384, 2005. Cit. on pp. 64, 65, and 66.

CANZONERI, M. B.; GRAY, J. A. Monetary policy games and the consequences of noncooperative behavior. **International Economic Review**, JSTOR, p. 547–564, 1985. Cit. on pp. 64 and 65.

CANZONERI, M. B.; HENDERSON, D. W. **Monetary Policy in Interdependent Economies: A Game-Theoretic Approach**. [*S.l.*]: MIT press, 1991. Cit. on pp. 64 and 65.

CARMONA, P.; DWEKAT, A.; MARDAWI, Z. No more black boxes! Explaining the predictions of a machine learning XGBoost classifier algorithm in business failure. **Research**

in International Business and Finance, v. 61, p. 101649, Oct. 2022. ISSN 02755319. DOI 10.1016/j.ribaf.2022.101649. Cit. on p. 58.

CHAGUE, F.; De-Losso, R.; GIOVANNETTI, B.; MANOEL, P. Central Bank Communication Affects the Term-Structure of Interest Rates. **Revista Brasileira de Economia**, v. 69, n. 2, 2015. ISSN 0034-7140. DOI 10.5935/0034-7140.20150007. Cit. on pp. 56, 97, and 106.

CHAI, C. P. Comparison of text preprocessing methods. **Natural Language Engineering**, Cambridge University Press, v. 29, n. 3, p. 509–553, 2023. Cit. on p. 53.

CHANG, J.; GERRISH, S.; WANG, C.; Boyd-graber, J.; BLEI, D. Reading Tea Leaves: How Humans Interpret Topic Models. *In*: **Advances in Neural Information Processing Systems**. [*S.l.*]: Curran Associates, Inc., 2009. v. 22. Cit. on p. 28.

CHAUHAN, U.; SHAH, A. Topic modeling using latent Dirichlet allocation: A survey. **ACM Computing Surveys (CSUR)**, ACM New York, NY, USA, v. 54, n. 7, p. 1–35, 2021. Cit. on p. 35.

CHEN, D.; ma, S.; HARIMOTO, K.; BAO, R.; SU, Q.; SUN, X. **Group, Extract and Aggregate: Summarizing a Large Amount of Finance News for Forex Movement Prediction**. [*S.l.*]: arXiv, 2019. Cit. on pp. 30, 55, and 97.

CHEN, H.; DE, P.; HU, Y. J.; HWANG, B.-H. Wisdom of Crowds: The Value of Stock Opinions Transmitted Through Social Media. **Review of Financial Studies**, v. 27, n. 5, p. 1367–1403, May 2014. ISSN 0893-9454, 1465-7368. DOI 10.1093/rfs/hhu001. Cit. on p. 33.

CHEN, X.; MA, X.; WANG, H.; LI, X.; ZHANG, C. A hierarchical attention network for stock prediction based on attentive multi-view news learning. **Neurocomputing**, v. 504, p. 1–15, Sep. 2022. ISSN 09252312. DOI 10.1016/j.neucom.2022.06.106. Cit. on p. 54.

CHEN, X.-Q.; MA, C.-Q.; REN, Y.-S.; LEI, Y.-T.; HUYNH, N. Q. A.; NARAYAN, S. Explainable artificial intelligence in finance: A bibliometric review. **Finance Research Letters**, v. 56, p. 104145, Sep. 2023. ISSN 15446123. DOI 10.1016/j.frl.2023.104145. Cit. on pp. 20, 29, and 58.

CHENG, D.; YANG, F.; XIANG, S.; LIU, J. Financial time series forecasting with multi-modality graph neural network. **Pattern Recognition**, v. 121, p. 108218, Jan. 2022. ISSN 00313203. DOI 10.1016/j.patcog.2021.108218. Cit. on pp. 29 and 56.

CHO, K.; MERRIËNBOER, B. V.; GULCEHRE, C.; BAHDANAU, D.; BOUGARES, F.; SCHWENK, H.; BENGIO, Y. Learning phrase representations using RNN encoder-decoder for statistical machine translation. **arXiv preprint arXiv:1406.1078**, 2014. Cit. on p. 42.

CHUNG, J.; GULCEHRE, C.; CHO, K.; BENGIO, Y. Empirical evaluation of gated recurrent neural networks on sequence modeling. **arXiv preprint arXiv:1412.3555**, 2014. Cit. on p. 42.

CLARIDA, R.; GALI, J.; GERTLER, M. A simple framework for international monetary policy analysis. **Journal of monetary economics**, Elsevier, v. 49, n. 5, p. 879–904, 2002. Cit. on pp. 65 and 66.

COCHRANE, J. H.; PIAZZESI, M. Decomposing the Yield Curve. **SSRN Electronic Journal**, 2009. ISSN 1556-5068. DOI 10.2139/ssrn.1333274. Cit. on p. 96.

CONT, R.; SCHAANNING, E. Monitoring indirect contagion. **Journal of Banking & Finance**, v. 104, p. 85–102, 2019. Cit. on p. 64.

COOK, T.; HAHN, T. K. Interest rate expectations and the slope of the money market yield curve. **FRB Richmond Economic Review**, v. 76, n. 5, p. 3–26, 1990. Cit. on p. 103.

CORREIA, M. P. R.; MUELLER, B. Pattern Making and Pattern Breaking: Measuring Novelty in Brazilian Economics. **SSRN Electronic Journal**, 2021. ISSN 1556-5068. DOI 10.2139/ssrn.3846395. Cit. on pp. 24 and 28.

CORSETTI, G.; PESENTI, P. Welfare and macroeconomic interdependence. **The Quarterly Journal of Economics**, MIT Press, v. 116, n. 2, p. 421–445, 2001. Cit. on pp. 65 and 66.

CORSETTI, G.; PESENTI, P. International dimensions of optimal monetary policy. **Journal of Monetary economics**, Elsevier, v. 52, n. 2, p. 281–305, 2005. Cit. on p. 66.

CORSETTI, G.; PESENTI, P.; ROUBINI, N.; TILLE, C. Competitive devaluations: Toward a welfare-based approach. **Journal of International Economics**, Elsevier, v. 51, n. 1, p. 217–241, 2000. Cit. on p. 66.

CORTES, C.; VAPNIK, V. Support-vector networks. **Machine learning**, Springer, v. 20, n. 3, p. 273–297, 1995. Cit. on pp. 22 and 31.

COVER, T.; HART, P. Nearest neighbor pattern classification. **IEEE transactions on information theory**, IEEE, v. 13, n. 1, p. 21–27, 1967. Cit. on p. 31.

CRUMP, R. K.; EUSEPI, S.; MOENCH, E. Is there hope for the expectations hypothesis? **FRB of New York Staff Report**, n. 1098, 2024. Cit. on p. 103.

DAI, Z.; YANG, Z.; YANG, Y.; CARBONELL, J.; LE, Q. V.; SALAKHUTDINOV, R. **Transformer-XL: Attentive Language Models Beyond a Fixed-Length Context**. [*S.l.*]: arXiv, 2019. Cit. on p. 48.

DANIEL, M.; NEVES, R. F.; HORTA, N. Company event popularity for financial markets using Twitter and sentiment analysis. **Expert Systems with Applications**, v. 71, p. 111–124, Apr. 2017. ISSN 09574174. DOI 10.1016/j.eswa.2016.11.022. Cit. on pp. 23 and 53.

DAVIS, A. K.; GE, W.; MATSUMOTO, D.; ZHANG, J. L. The effect of manager-specific optimism on the tone of earnings conference calls. **Review of Accounting Studies**, v. 20, n. 2, p. 639–673, Jun. 2015. ISSN 1380-6653, 1573-7136. DOI 10.1007/s11142-014-9309-4. Cit. on pp. 28 and 32.

DAVIS, A. K.; PIGER, J. M.; SEDOR, L. M. Beyond the Numbers: Measuring the Information Content of Earnings Press Release Language*. **Contemporary Accounting Research**, v. 29, n. 3, p. 845–868, Sep. 2012. ISSN 0823-9150, 1911-3846. DOI 10.1111/j.1911-3846.2011.01130.x. Cit. on pp. 28 and 32.

DAVIS, A. K.; TAMA-SWEET, I. Managers' Use of Language Across Alternative Disclosure Outlets: Earnings Press Releases versus MD&A*. **Contemporary Accounting Research**, v. 29, n. 3, p. 804–837, Sep. 2012. ISSN 0823-9150, 1911-3846. DOI 10.1111/j.1911-3846.2011.01125.x. Cit. on p. 32.

DAVIS, G. F.; GREVE, H. R. Corporate elite networks and governance changes in the 1980s. **American Journal of Sociology**, The University of Chicago Press, v. 103, n. 1, p. 1–37, 1997. Cit. on p. 64.

DEDOLA, L.; KARADI, P.; LOMBARDO, G. Global implications of national unconventional policies. **Journal of Monetary Economics**, Elsevier, v. 60, n. 1, p. 66–85, 2013. Cit. on p. 67.

DENG, S.; ZHANG, N.; ZHANG, W.; CHEN, J.; PAN, J. Z.; CHEN, H. Knowledge-Driven Stock Trend Prediction and Explanation via Temporal Convolutional Network. *In*: **Companion Proceedings of The 2019 World Wide Web Conference**. San Francisco USA: ACM, 2019. p. 678–685. ISBN 978-1-4503-6675-5. DOI 10.1145/3308560.3317701. Cit. on pp. 29 and 56.

DEVEREUX, M. B.; ENGEL, C. Monetary policy in the open economy revisited: Price setting and exchange-rate flexibility. **The review of economic studies**, Wiley-Blackwell, v. 70, n. 4, p. 765–783, 2003. Cit. on pp. 65 and 66.

DEVLIN, J.; CHANG, M.-W.; LEE, K.; TOUTANOVA, K. **BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding**. [*S.l.*]: arXiv, 2019. Cit. on pp. 5, 7, 16, 22, 25, 47, 48, 49, 55, 92, 147, 148, 152, and 159.

DEWACHTER, H.; LYRIO, M. Macro factors and the term structure of interest rates. **Journal of Money, Credit and Banking**, JSTOR, p. 119–140, 2006. Cit. on pp. 103 and 104.

DIEBOLD, F. X.; LI, C. Forecasting the term structure of government bond yields. **Journal of Econometrics**, v. 130, n. 2, p. 337–364, Feb. 2006. ISSN 03044076. DOI 10.1016/j.jeconom.2005.03.005. Cit. on p. 95.

DIEBOLD, F. X.; RUDEBUSCH, G. D.; ARUOBA, S. B. The macroeconomy and the yield curve: A dynamic latent factor approach. **Journal of Econometrics**, v. 131, n. 1-2, p.

309–338, Mar. 2006. ISSN 03044076. DOI 10.1016/j.jeconom.2005.01.011. Cit. on pp. 96 and 103.

DOMINGOS, P.; PAZZANI, M. On the optimality of the simple Bayesian classifier under zero-one loss. **Machine learning**, Springer, v. 29, n. 2, p. 103–130, 1997. Cit. on pp. 22 and 32.

DONTHU, N.; KUMAR, S.; MUKHERJEE, D.; PANDEY, N.; LIM, W. M. How to conduct a bibliometric analysis: An overview and guidelines. **Journal of Business Research**, v. 133, p. 285–296, Sep. 2021. ISSN 01482963. DOI 10.1016/j.jbusres.2021.04.070. Cit. on pp. 19, 20, and 21.

DORAN, J. S.; PETERSON, D. R.; PRICE, S. M. Earnings Conference Call Content and Stock Price: The Case of REITs. **The Journal of Real Estate Finance and Economics**, v. 45, n. 2, p. 402–434, Aug. 2012. ISSN 0895-5638, 1573-045X. DOI 10.1007/s11146-010-9266-z. Cit. on pp. 28 and 32.

DOUGAL, C.; ENGELBERG, J.; GARCÍA, D.; PARSONS, C. A. Journalists and the Stock Market. **Review of Financial Studies**, v. 25, n. 3, p. 639–679, Mar. 2012. ISSN 0893-9454, 1465-7368. DOI 10.1093/rfs/hhr133. Cit. on p. 33.

DUDA, R. O.; HART, P. E. *et al.* **Pattern Classification and Scene Analysis**. [*S.l.*]: Wiley New York, 1973. v. 3. Cit. on pp. 22 and 32.

DUMAIS, S. T. Improving the retrieval of information from external sources. **Behavior Research Methods, Instruments, & Computers**, v. 23, n. 2, p. 229–236, Jun. 1991. ISSN 1532-5970. DOI 10.3758/BF03203370. Cit. on p. 34.

ECK, N. J. V.; WALTMAN, L. VOS: A New Method for Visualizing Similarities Between Objects. *In*: DECKER, R.; LENZ, H. J. (Ed.). **Advances in Data Analysis**. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007. p. 299–306. ISBN 978-3-540-70981-7. Cit. on p. 21.

EHRMANN, M.; HOLTON, S.; KEDAN, D.; PHELAN, G. Monetary policy communication: Perspectives from former policymakers at the ECB. **Journal of Money, Credit and Banking**, Wiley Online Library, v. 56, n. 4, p. 837–864, 2024. Cit. on p. 68.

ELMAN, J. L. Finding structure in time. **Cognitive science**, v. 14, n. 2, p. 179–211, 1990. Cit. on pp. 22, 41, 43, and 111.

EVANS, C. L.; MARSHALL, D. A. Economic determinants of the nominal treasury yield curve. **Journal of Monetary Economics**, Elsevier, v. 54, n. 7, p. 1986–2003, 2007. Cit. on p. 103.

FAMA, E. F. Efficient capital markets. **Journal of finance**, v. 25, n. 2, p. 383–417, 1970. Cit. on p. 102.

FARIMANI, S. A.; JAHAN, M. V.; FARD, A. M. From Text Representation to Financial Market Prediction: A Literature Review. **Information**, Multidisciplinary Digital Publishing Institute, v. 13, n. 10, p. 466, Oct. 2022. ISSN 2078-2489. DOI 10.3390/info13100466. Cit. on pp. 24, 50, 53, and 55.

FASOLO, A. M.; GRAMINHO, F. M.; BASTOS, S. B. Seeing the forest for the trees: Using hLDA models to evaluate communication in banco central do brasil. BIS Working Papers, n. 1021, 2022. Cit. on pp. 63, 92, 96, and 106.

FELDMAN, R.; GOVINDARAJ, S.; LIVNAT, J.; SEGAL, B. Management's tone change, post earnings announcement drift and accruals. **Review of Accounting Studies**, v. 15, n. 4, p. 915–953, Dec. 2010. ISSN 1380-6653, 1573-7136. DOI 10.1007/s11142-009-9111-x. Cit. on p. 33.

FERREIRA, T. R.; SHOUSHA, S. Determinants of global neutral interest rates. **Journal of International Economics**, Elsevier, v. 145, p. 103833, 2023. Cit. on p. 74.

FERRIS, S. P.; HAO, G. Q.; LIAO, S. M.-Y. The Effect of Issuer Conservatism on IPO Pricing and Performance*. **Review of Finance**, v. 17, n. 3, p. 993–1027, Jul. 2013. ISSN 1573-692X, 1572-3097. DOI 10.1093/rof/rfs018. Cit. on p. 57.

FIGUEREDO, F. C. D.; MUELLER, B.; CAJUEIRO, D. O. A natural language measure of ideology in the Brazilian Senate. **Revista Brasileira de Ciência Política**, n. 37, p. e246618, 2022. ISSN 2178-4884, 0103-3352. DOI 10.1590/0103-3352.2022.37.246618. Cit. on pp. 27 and 57.

FIGUEROA, J. G.; PADILLA, F. I. V. Coordination and monetary policy among central banks. The NAFTA Case. **Revista Economía y Política**, n. 36, p. 33–50, 2022. Cit. on pp. 64, 78, and 89.

FISHER, M. Forces that shape the yield curve. **Economic Review**, Federal Reserve Bank of Atlanta, v. 86, n. Q1, p. 1–15, 2001. Cit. on pp. 96 and 104.

FORBES, K.; HA, J.; KOSE, M. A. Rate cycles. **World Bank Working Paper**, World Bank, 2024. Cit. on pp. 69, 83, and 86.

FORNARO, L.; ROMEI, F. Monetary policy during unbalanced global recoveries. CEPR Discussion Paper No. DP16971, 2022. Cit. on p. 67.

FRIEDMAN, J. H. Greedy function approximation: A gradient boosting machine. **Annals of statistics**, JSTOR, p. 1189–1232, 2001. Cit. on p. 31.

GAO, J.; YING, X.; XU, C.; WANG, J.; ZHANG, S.; LI, Z. Graph-Based Stock Recommendation by Time-Aware Relational Attention Network. **ACM Transactions on Knowledge Discovery from Data**, v. 16, n. 1, p. 1–21, Feb. 2022. ISSN 1556-4681, 1556-472X. DOI 10.1145/3451397. Cit. on p. 56.

GARCÍA, D. Sentiment during Recessions. **The Journal of Finance**, v. 68, n. 3, p. 1267–1300, Jun. 2013. ISSN 0022-1082, 1540-6261. DOI 10.1111/jofi.12027. Cit. on p. 33.

GARCÍA, D.; HU, X.; ROHRER, M. The colour of finance words. **Journal of Financial Economics**, v. 147, n. 3, p. 525–549, Mar. 2023. ISSN 0304405X. DOI 10.1016/j.jfineco.2022.11.006. Cit. on p. 33.

GENTZKOW, M.; KELLY, B.; TADDY, M. Text as Data. **Journal of Economic Literature**, v. 57, n. 3, p. 535–574, Sep. 2019. ISSN 0022-0515. DOI 10.1257/jel.20181020. Cit. on pp. 18, 33, and 110.

GENTZKOW, M.; SHAPIRO, J. M. What Drives Media Slant? Evidence From U.S. Daily Newspapers. **Econometrica**, v. 78, n. 1, p. 35–71, 2010. ISSN 0012-9682. DOI 10.3982/ECTA7195. Cit. on p. 57.

GERS, F. A.; SCHMIDHUBER, J.; CUMMINS, F. Learning to forget: Continual prediction with LSTM. **Neural computation**, v. 12, n. 10, p. 2451–2471, 2000. Cit. on p. 42.

GERTLER, M.; KARADI, P. A model of unconventional monetary policy. **Journal of monetary Economics**, Elsevier, v. 58, n. 1, p. 17–34, 2011. Cit. on p. 89.

GETMANSKY, M.; LO, A. W.; PELIZZON, L. Econometric measures of connectedness and systemic risk in the finance and insurance sectors. **Journal of Financial Economics**, v. 104, p. 535–559, 2012. Cit. on p. 64.

GILLES, C. Volatility and the treasury yield curve. *In*: **Financial Market Volatility: Measurement, Causes, and Consequences, BIS Conference Proceedings**. [*S.l.*: *s.n.*], 1996. p. 228–242. Cit. on p. 104.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep Learning**. [*S.l.*]: MIT press, 2016. Cit. on pp. 22 and 37.

GORODNICHENKO, Y.; PHAM, T.; TALAVERA, O. The voice of monetary policy. **American Economic Review**, American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203, v. 113, n. 2, p. 548–584, 2023. Cit. on p. 68.

GOTTHELF, N.; UHL, M. W. News Sentiment: A New Yield Curve Factor. **Journal of Behavioral Finance**, v. 20, n. 1, p. 31–41, Jan. 2019. ISSN 1542-7560, 1542-7579. DOI 10.1080/15427560.2018.1432620. Cit. on p. 97.

Greenwood-Nimmo, M.; NGUYEN, V. H.; RAFFERTY, B. Risk and return spillovers among the G10 currencies. **Journal of Financial Markets**, v. 31, p. 43–62, Nov. 2016. ISSN 13864181. DOI 10.1016/j.finmar.2016.05.001. Cit. on p. 77.

GRISHMAN, R.; WESTBROOK, D.; MEYERS, A. Our English extraction system, developed over the course of the last several ACE evaluations, includes facilities for the EDR (entity), RDR (relation), and VDR (event) tasks; TIMEX and value detection have not yet been implemented. **ACE**, v. 5, p. 2, 2005. Cit. on p. 29.

GÜRKAYNAK, R. S.; SACK, B.; WRIGHT, J. H. The U.S. Treasury yield curve: 1961 to the present. **Journal of Monetary Economics**, v. 54, n. 8, p. 2291–2304, Nov. 2007. ISSN 03043932. DOI 10.1016/j.jmoneco.2007.06.029. Cit. on pp. 95 and 96.

GUTHRIE, L.; PUSTEJOVSKY, J.; WILKS, Y.; SLATOR, B. M. The role of lexicons in natural language processing. **Communications of the ACM**, ACM New York, NY, USA, v. 39, n. 1, p. 63–72, 1996. Cit. on p. 21.

HAMADA, K. A strategic analysis of monetary interdependence. **Journal of Political Economy**, The University of Chicago Press, v. 84, n. 4, Part 1, p. 677–700, 1976. Cit. on p. 64.

HAMILTON, J. D. **Time Series Analysis**. [*S.l.*]: Princeton university press, 2020. Cit. on pp. 41 and 111.

HAMILTON, J. D.; WU, J. C. The effectiveness of alternative monetary policy tools in a zero lower bound environment. **Journal of Money, Credit and Banking**, Wiley Online Library, v. 44, p. 3–46, 2012. Cit. on p. 85.

HANSEN, S.; MCMAHON, M. Shocking language: Understanding the macroeconomic effects of central bank communication. **Journal of International Economics**, Elsevier, v. 99, p. S114–S133, 2016. Cit. on p. 63.

HANSSON, M. **Evolution of Topics in Central Bank Speech Communication**. [*S.l.*]: arXiv, 2021. Cit. on pp. 51 and 69.

HARRIS, Z. S. Distributional Structure. ***WORD***, v. 10, n. 2-3, p. 146–162, Aug. 1954. ISSN 0043-7956, 2373-5112. DOI 10.1080/00437956.1954.11659520. Cit. on p. 32.

HAYKIN, S. **Neural Networks and Learning Machines, 3/E**. [*S.l.*]: Pearson Education India, 2009. Cit. on p. 37.

HENRY, E. Are Investors Influenced By How Earnings Press Releases Are Written? **Journal of Business Communication**, v. 45, n. 4, p. 363–407, Oct. 2008. ISSN 0021-9436. DOI 10.1177/0021943608319388. Cit. on p. 32.

HERMANN, K. M.; KOCISKY, T.; GREFENSTETTE, E.; ESPEHOLT, L.; KAY, W.; SULEYMAN, M.; BLUNSOM, P. Teaching Machines to Read and Comprehend. *In*: **Advances in Neural Information Processing Systems**. [*S.l.*]: Curran Associates, Inc., 2015. v. 28. Cit. on p. 52.

HESTON, S. L.; SINHA, N. R. News vs. Sentiment: Predicting Stock Returns from News Stories. **Financial Analysts Journal**, v. 73, n. 3, p. 67–83, Jul. 2017. ISSN 0015-198X, 1938-3312. DOI 10.2469/faj.v73.n3.3. Cit. on pp. 28 and 32.

HILLERT, A.; Niessen-Ruenzi, A.; RUENZI, S. Mutual Fund Shareholder Letter Tone Do Investors Listen? **SSRN Electronic Journal**, 2014. ISSN 1556-5068. DOI 10.2139/ssrn.2524610. Cit. on p. 57.

HOBERG, G.; PHILLIPS, G. Text-based network industries and endogenous product differentiation. **Journal of Political Economy**, University of Chicago Press Chicago, IL, v. 124, n. 5, p. 1423–1465, 2016. Cit. on p. 31.

HOCHREITER, S.; SCHMIDHUBER, J. Long Short-Term Memory. **Neural Computation**, v. 9, n. 8, p. 1735–1780, Nov. 1997. ISSN 0899-7667, 1530-888X. DOI 10.1162/neco.1997.9.8.1735. Cit. on pp. 22, 25, and 42.

HUANG, X.; TEOH, S. H.; ZHANG, Y. Tone management. **The Accounting Review**, American Accounting Association, v. 89, n. 3, p. 1083–1113, 2014. Cit. on p. 33.

IMISIKER, S. International Monetary Policy Coordination through Communication: Chasing the Loch Ness Monster. **SSRN Electronic Journal**, 2016. ISSN 1556-5068. DOI 10.2139/ssrn.2888087. Cit. on p. 63.

IZENMAN, A. J. **Modern Multivariate Statistical Techniques: Regression, Classification and Manifold Learning**. [*S.l.*]: Springer, 2008. Cit. on p. 32.

JENSEN, H. Monetary policy cooperation and multiple equilibria. **Journal of Economic Dynamics and Control**, Elsevier, v. 23, n. 8, p. 1133–1153, 1999. Cit. on p. 64.

JIANG, W. Applications of deep learning in stock market prediction: Recent progress. **Expert Systems with Applications**, v. 184, p. 115537, Dec. 2021. ISSN 09574174. DOI 10.1016/j.eswa.2021.115537. Cit. on pp. 24 and 43.

JIN, Z.; YANG, Y.; LIU, Y. Stock closing price prediction based on sentiment analysis and LSTM. **Neural Computing and Applications**, v. 32, n. 13, p. 9713–9729, Jul. 2020. ISSN 0941-0643, 1433-3058. DOI 10.1007/s00521-019-04504-2. Cit. on p. 53.

JITMANEEROJ, B.; LAMLA, M. J.; WOOD, A. The implications of central bank transparency for uncertainty and disagreement. **Journal of International Money and Finance**, Elsevier, v. 90, p. 222–240, 2019. Cit. on p. 63.

JORDÀ, Ò.; TAYLOR, A. M. **Riders on the Storm**. [*S.l.*], 2019. Cit. on p. 74.

JORDAN, MI. **Serial Order: A Parallel Distributed Processing Approach. Technical Report, June 1985-March 1986**. [*S.l.*], 1986. Cit. on pp. 22, 41, 43, and 111.

JOSLIN, S.; PRIEBSCH, M.; SINGLETON, K. J. Risk premiums in dynamic term structure models with unspanned macro risks. **The Journal of Finance**, Wiley Online Library, v. 69, n. 3, p. 1197–1233, 2014. Cit. on p. 104.

Kalemli-Özcan, S. **US Monetary Policy and International Risk Spillovers**. [*S.l.*], 2019. Cit. on p. 67.

KAPUR, M.; MOHAN, R. **Monetary Policy Coordination and the Role of Central Banks**. [*S.l.*], 2014. Cit. on p. 64.

KATHAROPOULOS, A.; VYAS, A.; PAPPAS, N.; FLEURET, F. **Transformers Are RNNs: Fast Autoregressive Transformers with Linear Attention**. [*S.l.*]: arXiv, 2020. Cit. on p. 48.

KEARNEY, C.; LIU, S. Textual sentiment in finance: A survey of methods and models. **International Review of Financial Analysis**, v. 33, p. 171–185, May 2014. ISSN 10575219. DOI 10.1016/j.irfa.2014.02.006. Cit. on p. 32.

KEARNS, J.; SCHRIMPF, A.; XIA, F. D. Explaining monetary spillovers: The matrix reloaded. **Journal of Money, Credit and Banking**, Wiley Online Library, v. 55, n. 6, p. 1535–1568, 2023. Cit. on p. 67.

KHAN, A.; GOODELL, J. W.; HASSAN, M. K.; PALTRINIERI, A. A bibliometric review of finance bibliometric papers. **Finance Research Letters**, v. 47, p. 102520, Jun. 2022. ISSN 15446123. DOI 10.1016/j.frl.2021.102520. Cit. on pp. 20 and 21.

KILEY, M. T. The global equilibrium real interest rate: Concepts, estimates, and challenges. **Annual Review of Financial Economics**, Annual Reviews, v. 12, n. 1, p. 305–326, 2020. Cit. on p. 74.

KIM, D. H.; ORPHANIDES, A. The bond market term premium: What is it, and how can we measure it? **BIS Quarterly Review, June**, 2007. Cit. on p. 104.

KIM, R.; SO, C. H.; JEONG, M.; LEE, S.; KIM, J.; KANG, J. **HATS: A Hierarchical Graph Attention Network for Stock Movement Prediction**. [*S.l.*]: arXiv, 2019. Cit. on p. 56.

KIM, Y. **Convolutional Neural Networks for Sentence Classification**. [*S.l.*]: arXiv, 2014. Cit. on p. 53.

KING, M. Evaluating natural language processing systems. **Communications of the ACM**, ACM New York, NY, USA, v. 39, n. 1, p. 73–79, 1996. Cit. on p. 23.

KINGMA, D. P.; BA, J. **Adam: A Method for Stochastic Optimization**. [*S.l.*]: arXiv, 2017. DOI 10.48550/arXiv.1412.6980. Cit. on p. 159.

KUMBURE, M. M.; LOHRMANN, C.; LUUKKA, P.; PORRAS, J. Machine learning techniques and data for stock market forecasting: A literature review. **Expert Systems with Applications**, v. 197, p. 116659, Jul. 2022. ISSN 09574174. DOI 10.1016/j.eswa.2022.116659. Cit. on pp. 34, 43, 53, and 54.

LABONDANCE, F.; HUBERT, P. Central Bank sentiment and policy expectations. 2017. Cit. on p. 63.

LAMPLE, G.; CONNEAU, A. **Cross-Lingual Language Model Pretraining**. [*S.l.*]: arXiv, 2019. Cit. on p. 48.

LAN, Z.; CHEN, M.; GOODMAN, S.; GIMPEL, K.; SHARMA, P.; SORICUT, R. **ALBERT: A Lite BERT for Self-supervised Learning of Language Representations**. [*S.l.*]: arXiv, 2020. Cit. on p. 48.

LANE, P. R. The new open economy macroeconomics: A survey. **Journal of international economics**, Elsevier, v. 54, n. 2, p. 235–266, 2001. Cit. on p. 66.

LARSEN, V. H.; THORSRUD, L. A. The value of news for economic developments. **Journal of Econometrics**, v. 210, n. 1, p. 203–218, May 2019. ISSN 03044076. DOI 10.1016/j.jeconom.2018.11.013. Cit. on p. 57.

LATORA, V.; NICOSIA, V.; RUSSO, G. **Complex Networks: Principles, Methods and Applications**. [*S.l.*]: Cambridge University Press, 2017. Cit. on p. 64.

LE, Q.; MIKOLOV, T. Distributed representations of sentences and documents. *In*: **International Conference on Machine Learning**. [*S.l.*]: PMLR, 2014. p. 1188–1196. Cit. on p. 40.

LECUN, Y.; BOSER, B.; DENKER, J. S.; HENDERSON, D.; HOWARD, R. E.; HUBBARD, W.; JACKEL, L. D. Backpropagation applied to handwritten zip code recognition. **Neural computation**, MIT Press, v. 1, n. 4, p. 541–551, 1989. Cit. on p. 22.

LEWIS, M.; LIU, Y.; GOYAL, N.; GHAZVININEJAD, M.; MOHAMED, A.; LEVY, O.; STOYANOV, V.; ZETTLEMOYER, L. **BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension**. [*S.l.*]: arXiv, 2019. Cit. on p. 48.

LI, F. Textual Analysis of Corporate Disclosures: A Survey of the Literature. **Journal of Accounting Literature**, v. 29, p. 143, 2010. Cit. on p. 58.

LI, R.; ZHAO, X.; MOENS, M.-F. A Brief Overview of Universal Sentence Representation Methods: A Linguistic View. **ACM Computing Surveys**, v. 55, n. 3, p. 1–42, Mar. 2023. ISSN 0360-0300, 1557-7341. DOI 10.1145/3482853. Cit. on p. 40.

LI, X.; WU, P.; WANG, W. Incorporating stock prices and news sentiments for stock market prediction: A case of Hong Kong. **Information Processing & Management**, v. 57, n. 5, p. 102212, Sep. 2020. ISSN 03064573. DOI 10.1016/j.ipm.2020.102212. Cit. on p. 53.

LI, Z.; LIU, F.; YANG, W.; PENG, S.; ZHOU, J. A survey of convolutional neural networks: Analysis, applications, and prospects. **IEEE transactions on neural networks and learning systems**, IEEE, 2021. Cit. on p. 22.

LIANG, P.; BOMMASANI, R.; LEE, T.; TSIPRAS, D.; SOYLU, D.; YASUNAGA, M.; ZHANG, Y.; NARAYANAN, D.; WU, Y.; KUMAR, A. *et al.* Holistic evaluation of language models. **arXiv preprint arXiv:2211.09110**, 2022. Cit. on p. 23.

LIN, T.; WANG, Y.; LIU, X.; QIU, X. **A Survey of Transformers**. [*S.l.*]: arXiv, 2021. Cit. on p. 48.

LITTERMAN, R. B.; SCHEINKMAN, J. Common factors affecting bond returns. **The journal of fixed income**, Institutional Investor Journals Umbrella, v. 1, n. 1, p. 54–61, 1991. Cit. on pp. 5, 7, 16, 92, 95, 96, 99, 100, 101, and 103.

LIU, B.; MCCONNELL, J. J. The role of the media in corporate governance: Do the media influence managers' capital allocation decisions? **Journal of Financial Economics**, v. 110, n. 1, p. 1–17, Oct. 2013. ISSN 0304405X. DOI 10.1016/j.jfineco.2013.06.003. Cit. on p. 33.

LIU, Q.; CHENG, X.; SU, S.; ZHU, S. Hierarchical Complementary Attention Network for Predicting Stock Price Movements with News. *In*: **Proceedings of the 27th ACM International Conference on Information and Knowledge Management**. Torino Italy: ACM, 2018. p. 1603–1606. ISBN 978-1-4503-6014-2. DOI 10.1145/3269206.3269286. Cit. on pp. 55 and 97.

LIU, X.; HUANG, H.; ZHANG, Y.; YUAN, C. News-Driven Stock Prediction With Attention-Based Noisy Recurrent State Transition. **Neurocomputing**, v. 470, p. 66–75, Jan. 2022. ISSN 09252312. DOI 10.1016/j.neucom.2021.10.092. Cit. on p. 56.

LIU, X.; LUO, Z.; HUANG, H. Jointly Multiple Events Extraction via Attention-based Graph Information Aggregation. *In*: **Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing**. [*S.l.*: *s.n.*], 2018. p. 1247–1256. DOI 10.18653/v1/D18-1156. Cit. on p. 29.

LIU, Y. **Fine-Tune BERT for Extractive Summarization**. [*S.l.*]: arXiv, 2019. DOI 10.48550/arXiv.1903.10318. Cit. on p. 30.

LIU, Y.; OTT, M.; GOYAL, N.; DU, J.; JOSHI, M.; CHEN, D.; LEVY, O.; LEWIS, M.; ZETTLEMOYER, L.; STOYANOV, V. **RoBERTa: A Robustly Optimized BERT Pretraining Approach**. [*S.l.*]: arXiv, 2019. Cit. on pp. 48, 93, and 98.

LIU, Y.; ZENG, Q.; MERÉ, J. O.; YANG, H. Anticipating Stock Market of the Renowned Companies: A Knowledge Graph Approach. **Complexity**, v. 2019, p. 1–15, Aug. 2019. ISSN 1076-2787, 1099-0526. DOI 10.1155/2019/9202457. Cit. on pp. 29 and 30.

LIU, Z.; PAPPA, E. Gains from international monetary policy coordination: Does it pay to be different? **Journal of economic Dynamics and control**, Elsevier, v. 32, n. 7, p. 2085–2117, 2008. Cit. on p. 64.

LOSHCHILOV, I.; HUTTER, F. **Decoupled Weight Decay Regularization**. [*S.l.*]: arXiv, 2019. DOI 10.48550/arXiv.1711.05101. Cit. on p. 159.

LOUGHRAN, T.; MCDONALD, B. When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks. **The Journal of Finance**, v. 66, n. 1, p. 35–65, Feb. 2011. ISSN 0022-1082, 1540-6261. DOI 10.1111/j.1540-6261.2010.01625.x. Cit. on pp. 32, 33, and 34.

LOUGHRAN, T.; MCDONALD, B. IPO first-day returns, offer price revisions, volatility, and form S-1 language. **Journal of Financial Economics**, v. 109, n. 2, p. 307–326, Aug. 2013. ISSN 0304405X. DOI 10.1016/j.jfineco.2013.02.017. Cit. on p. 57.

LOUGHRAN, T.; MCDONALD, B. Textual Analysis in Accounting and Finance: A Survey. **Journal of Accounting Research**, v. 54, n. 4, p. 1187–1230, Sep. 2016. ISSN 0021-8456, 1475-679X. DOI 10.1111/1475-679X.12123. Cit. on pp. 32, 51, 58, and 59.

LU, J.; GONG, P.; YE, J.; ZHANG, J.; ZHANG, C. A survey on machine learning from few samples. **Pattern Recognition**, Elsevier, v. 139, p. 109480, 2023. Cit. on p. 50.

LUTZ, F. A. The structure of interest rates. **The Quarterly Journal of Economics**, MIT Press, v. 55, n. 1, p. 36–63, 1940. Cit. on p. 103.

MAIA, M.; SALES, J. E.; FREITAS, A.; HANDSCHUH, S.; ENDRES, M. A comparative study of deep neural network models on multi-label text classification in finance. *In*: **2021 IEEE 15th International Conference on Semantic Computing (ICSC)**. [*S.l.*]: IEEE, 2021. p. 183–190. Cit. on p. 43.

MALO, P.; SINHA, A.; KORHONEN, P.; WALLENIUS, J.; TAKALA, P. Good debt or bad debt: Detecting semantic orientations in economic texts. **Journal of the Association for Information Science and Technology**, v. 65, n. 4, p. 782–796, Apr. 2014. ISSN 2330-1635, 2330-1643. DOI 10.1002/asi.23062. Cit. on p. 51.

MALTE, A.; RATADIYA, P. **Evolution of Transfer Learning in Natural Language Processing**. [*S.l.*]: arXiv, 2019. Cit. on p. 40.

MANNING, C. D. Human Language Understanding & Reasoning. **Daedalus**, v. 151, n. 2, p. 127–138, May 2022. ISSN 0011-5266, 1548-6192. Cit. on p. 18.

MANNING, C. D.; RAGHAVAN, P.; SCHÜTZE, H. **Introduction to Information Retrieval**. New York, NY, USA: Cambridge University Press, 2008. ISBN 0-521-86571-9 978-0-521-86571-5. Cit. on p. 34.

MANTEGNA, R. Hierarchical structure in financial markets. **European Physical Journal B**, v. 11, p. 193–197, 1999. Cit. on p. 64.

MATSUMURA, M.; MOREIRA, A.; VICENTE, J. Forecasting the yield curve with linear factor models. **International Review of Financial Analysis**, v. 20, n. 5, p. 237–243, Oct. 2011. ISSN 10575219. DOI 10.1016/j.irfa.2011.05.003. Cit. on p. 96.

MCAULIFFE, J.; BLEI, D. Supervised Topic Models. *In*: **Advances in Neural Information Processing Systems**. [*S.l.*]: Curran Associates, Inc., 2007. v. 20. Cit. on p. 28.

MIKOLOV, T.; CHEN, K.; CORRADO, G.; DEAN, J. **Efficient Estimation of Word Representations in Vector Space**. [*S.l.*]: arXiv, 2013. Cit. on pp. 22 and 38.

MIKOLOV, T.; SUTSKEVER, I.; CHEN, K.; CORRADO, G.; DEAN, J. **Distributed Representations of Words and Phrases and Their Compositionality**. [*S.l.*]: arXiv, 2013. Cit. on p. 38.

MINH, D. L.; Sadeghi-Niaraki, A.; HUY, H. D.; MIN, K.; MOON, H. Deep Learning Approach for Short-Term Stock Trends Prediction Based on Two-Stream Gated Recurrent Unit Network. **IEEE Access**, v. 6, p. 55392–55404, 2018. ISSN 2169-3536. DOI 10.1109/ACCESS.2018.2868970. Cit. on p. 40.

MISHEV, K.; GJORGJEVIKJ, A.; VODENSKA, I.; CHITKUSHEV, L. T.; TRAJANOV, D. Evaluation of Sentiment Analysis in Finance: From Lexicons to Transformers. **IEEE Access**, v. 8, p. 131662–131682, 2020. ISSN 2169-3536. DOI 10.1109/ACCESS.2020.3009626. Cit. on pp. 23, 36, 55, 96, 97, and 98.

MISHKIN, F. S. **The Channels of Monetary Transmission: Lessons for Monetary Policy**. [*S.l.*]: National Bureau of Economic Research Cambridge, Mass., USA, 1996. Cit. on p. 92.

NELSON, C. R.; SIEGEL, A. F. Parsimonious Modeling of Yield Curves. **The Journal of Business**, v. 60, n. 4, p. 473, Jan. 1987. ISSN 0021-9398, 1537-5374. DOI 10.1086/296409. Cit. on pp. 95, 97, and 98.

NEUENKIRCH, M. Managing financial market expectations: The role of central bank transparency and central bank communication. **European Journal of Political Economy**, Elsevier, v. 28, n. 1, p. 1–13, 2012. Cit. on p. 63.

NEWMAN, M. E.; GIRVAN, M. Finding and evaluating community structure in networks. **Physical review E**, APS, v. 69, n. 2, p. 026113, 2004. Cit. on p. 73.

NIŢOI, M.; POCHEA, M.-M.; RADU, Ş.-C. Unveiling the sentiment behind central bank narratives: A novel deep learning index. **Journal of Behavioral and Experimental Finance**, v. 38, p. 100809, Jun. 2023. ISSN 22146350. DOI 10.1016/j.jbef.2023.100809. Cit. on pp. 23, 51, 57, 93, 98, 107, and 108.

OBSTFELD, M. Uncoordinated monetary policies risk a historic global slowdown. **PIIE Realtime Economic Issues Watch, September**, v. 12, 2022. Cit. on p. 86.

OBSTFELD, M.; ROGOFF, K. Exchange rate dynamics redux. **Journal of political economy**, The University of Chicago Press, v. 103, n. 3, p. 624–660, 1995. Cit. on pp. 65 and 66.

OBSTFELD, M.; ROGOFF, K. Global implications of self-oriented national monetary rules. **The Quarterly journal of economics**, MIT Press, v. 117, n. 2, p. 503–535, 2002. Cit. on pp. 65 and 66.

OPENAI. **GPT-4 Technical Report**. [*S.l.*]: arXiv, 2023. Cit. on pp. 47 and 48.

ORPHANIDES, A.; NORDEN, S. van. The unreliability of output-gap estimates in real time. **Review of economics and statistics**, MIT Press 238 Main St., Suite 500, Cambridge, MA 02142-1046, USA journals . . . , v. 84, n. 4, p. 569–583, 2002. Cit. on p. 74.

OUDIZ, G.; SACHS, J.; BLANCHARD, O. J.; MARRIS, S. N.; WOO, W. T. Macroeconomic policy coordination among the industrial economies. **Brookings Papers on Economic Activity**, JSTOR, v. 1984, n. 1, p. 1–75, 1984. Cit. on pp. 64 and 65.

OVER, P.; DANG, H.; HARMAN, D. DUC in context. **Information Processing & Management**, v. 43, n. 6, p. 1506–1520, Nov. 2007. ISSN 03064573. DOI 10.1016/j.ipm.2007.01.019. Cit. on p. 52.

PASCANU, R.; MIKOLOV, T.; BENGIO, Y. On the difficulty of training recurrent neural networks. *In*: **International Conference on Machine Learning**. [*S.l.*]: PMLR, 2013. p. 1310–1318. Cit. on p. 41.

PASSALIS, N.; AVRAMELOU, L.; SEFICHA, S.; TSANTEKIDIS, A.; DOROPOULOS, S.; MAKRIS, G.; TEFAS, A. Multisource financial sentiment analysis for detecting Bitcoin price change indications using deep learning. **Neural Computing and Applications**, v. 34, n. 22, p. 19441–19452, Nov. 2022. ISSN 0941-0643, 1433-3058. DOI 10.1007/s00521-022-07509-6. Cit. on pp. 25 and 53.

PENNINGTON, J.; SOCHER, R.; MANNING, C. GloVe: Global Vectors for Word Representation. *In*: MOSCHITTI, A.; PANG, B.; DAELEMANS, W. (Ed.). **Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)**. Doha, Qatar: Association for Computational Linguistics, 2014. p. 1532–1543. DOI 10.3115/v1/D14-1162. Cit. on p. 40.

PETERS, M. E.; NEUMANN, M.; IYYER, M.; GARDNER, M.; CLARK, C.; LEE, K.; ZETTLEMOYER, L. **Deep Contextualized Word Representations**. [*S.l.*]: arXiv, 2018. Cit. on p. 40.

PETROPOULOS, A.; SIAKOULIS, V. Can central bank speeches predict financial market turbulence? Evidence from an adaptive NLP sentiment index analysis using XGBoost machine learning technique. **Central Bank Review**, v. 21, n. 4, p. 141–153, Dec. 2021. ISSN 13030701. DOI 10.1016/j.cbrev.2021.12.002. Cit. on pp. 23, 57, and 97.

PFEIFER, M.; MAROHL, V. P. CentralBankRoBERTa: A fine-tuned large language model for central bank communications. **The Journal of Finance and Data Science**, v. 9, p. 100114, Nov. 2023. ISSN 24059188. DOI 10.1016/j.jfds.2023.100114. Cit. on pp. 49, 93, 97, 98, and 108.

PINCOMBE, B. **Comparison of Human and Latent Semantic Analysis (LSA) Judgements of Pairwise Document Similarities for a News Corpus**. [*S.l.*]: Citeseer, 2004. Cit. on p. 71.

PRICE, S. M.; DORAN, J. S.; PETERSON, D. R.; BLISS, B. A. Earnings conference calls and stock returns: The incremental informativeness of textual tone. **Journal of Banking & Finance**, v. 36, n. 4, p. 992–1011, Apr. 2012. ISSN 03784266. DOI 10.1016/j.jbankfin.2011.10.013. Cit. on pp. 28 and 32.

PURDA, L.; SKILLICORN, D. Accounting Variables, Deception, and a Bag of Words: Assessing the Tools of Fraud Detection. **Contemporary Accounting Research**, v. 32, n. 3, p. 1193–1223, Sep. 2015. ISSN 0823-9150, 1911-3846. DOI 10.1111/1911-3846.12089. Cit. on p. 57.

QIU, X.; SUN, T.; XU, Y.; SHAO, Y.; DAI, N.; HUANG, X. Pre-trained models for natural language processing: A survey. **Science China Technological Sciences**, v. 63, n. 10, p. 1872–1897, Oct. 2020. ISSN 1674-7321, 1869-1900. DOI 10.1007/s11431-020-1647-3. Cit. on p. 49.

RADFORD, A.; WU, J.; CHILD, R.; LUAN, D.; AMODEI, D.; SUTSKEVER, I. *et al.* Language models are unsupervised multitask learners. **OpenAI blog**, v. 1, n. 8, p. 9, 2019. Cit. on pp. 47 and 48.

RAVASSI, S. Measuring Central Banks Communications with Machine Learning. 2020. Cit. on p. 33.

REY, H. **Dilemma Not Trilemma: The Global Financial Cycle and Monetary Policy Independence**. [*S.l.*], 2015. Cit. on p. 67.

RIBEIRO, A. H.; TIELS, K.; AGUIRRE, L. A.; SCHÖN, T. Beyond exploding and vanishing gradients: Analysing RNN training using attractors and smoothness. *In*: **International Conference on Artificial Intelligence and Statistics**. [*S.l.*]: PMLR, 2020. p. 2370–2380. Cit. on p. 41.

ROGOFF, K. Can international monetary policy cooperation be counterproductive? **Journal of International Economics**, v. 18, n. 3, p. 199–217, 1985. ISSN 0022-1996. DOI 10.1016/0022-1996(85)90052-2. Cit. on p. 65.

ROOS, M.; RECCIUS, M. Narratives in economics. **Journal of Economic Surveys**, p. joes.12576, Jun. 2023. ISSN 0950-0804, 1467-6419. DOI 10.1111/joes.12576. Cit. on pp. 24, 28, and 57.

RUDEBUSCH, G. D. Federal Reserve interest rate targeting, rational expectations, and the term structure. **Journal of monetary Economics**, Elsevier, v. 35, n. 2, p. 245–274, 1995. Cit. on p. 103.

RUDEBUSCH, G. D.; WU, T. A Macro-Finance Model of the Term Structure, Monetary Policy and the Economy. **The Economic Journal**, v. 118, n. 530, p. 906–926, Jul. 2008. ISSN 0013-0133, 1468-0297. DOI 10.1111/j.1468-0297.2008.02155.x. Cit. on p. 96.

SAHA, S.; GAO, J.; GERLACH, R. A survey of the application of graph-based approaches in stock market analysis and prediction. **International Journal of Data Science and Analytics**, v. 14, n. 1, p. 1–15, Jun. 2022. ISSN 2364-415X, 2364-4168. DOI 10.1007/s41060-021-00306-9. Cit. on p. 55.

SALTON, G.; BUCKLEY, C. Term-weighting approaches in automatic text retrieval. **Information Processing & Management**, v. 24, n. 5, p. 513–523, Jan. 1988. ISSN 03064573. DOI 10.1016/0306-4573(88)90021-0. Cit. on pp. 22 and 33.

SANH, V.; DEBUT, L.; CHAUMOND, J.; WOLF, T. **DistilBERT, a Distilled Version of BERT: Smaller, Faster, Cheaper and Lighter**. [*S.l.*]: arXiv, 2020. Cit. on pp. 48 and 55.

SARNO, L.; THORNTON, D. L.; VALENTE, G. The empirical failure of the expectations hypothesis of the term structure of bond yields. **Journal of Financial and Quantitative Analysis**, Cambridge University Press, v. 42, n. 1, p. 81–100, 2007. Cit. on p. 103.

SEBASTIANI, F. Machine learning in automated text categorization. **ACM computing surveys (CSUR)**, ACM New York, NY, USA, v. 34, n. 1, p. 1–47, 2002. Cit. on p. 23.

SHAKER, S. S.; ALHAJIM, D.; Al-Khazaali, A. A. T.; HUSSEIN, H. A.; ATHAB, A. F. Feature extraction based text classification: A review. **Journal of Algebraic Statistics**, v. 13, n. 1, p. 646–653, 2022. Cit. on p. 23.

SHAPIRO, A. H.; SUDHOF, M.; WILSON, D. J. Measuring news sentiment. **Journal of Econometrics**, v. 228, n. 2, p. 221–243, Jun. 2022. ISSN 03044076. DOI 10.1016/j.jeconom.2020.07.053. Cit. on pp. 23, 33, 55, 96, and 97.

SHAPIRO, A. H.; WILSON, D. J. Taking the fed at its word: A new approach to estimating central bank objectives using text analysis. **The Review of Economic Studies**, Oxford University Press, v. 89, n. 5, p. 2768–2805, 2022. Cit. on pp. 56, 63, and 96.

SHILLER, R. J. Narrative economics. **American economic review**, American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203, v. 107, n. 4, p. 967–1004, 2017. Cit. on p. 57.

SOHANGIR, S.; WANG, D.; POMERANETS, A.; KHOSHGOFTAAR, T. M. Big Data: Deep Learning for financial sentiment analysis. **Journal of Big Data**, v. 5, n. 1, p. 3, Dec. 2018. ISSN 2196-1115. DOI 10.1186/s40537-017-0111-6. Cit. on p. 23.

SOLOMON, D. H.; SOLTES, E.; SOSYURA, D. Winners in the spotlight: Media coverage of fund holdings as a driver of flows. **Journal of Financial Economics**, v. 113, n. 1, p. 53–72, Jul. 2014. ISSN 0304405X. DOI 10.1016/j.jfineco.2014.02.009. Cit. on p. 57.

SOUMA, W.; VODENSKA, I.; AOYAMA, H. Enhanced news sentiment analysis using deep learning methods. **Journal of Computational Social Science**, v. 2, n. 1, p. 33–46,

Jan. 2019. ISSN 2432-2717, 2432-2725. DOI 10.1007/s42001-019-00035-x. Cit. on pp. 24 and 54.

SUTHERLAND, A. International monetary policy coordination and financial market integration. **Available at SSRN 510183**, 2004. Cit. on p. 64.

SUTSKEVER, I.; VINYALS, O.; LE, Q. V. Sequence to sequence learning with neural networks. **Advances in neural information processing systems**, v. 27, 2014. Cit. on pp. 22, 25, and 42.

SVENSSON, L. E. **Estimating and Interpreting Forward Interest Rates: Sweden 1992-1994**. [*S.l.*]: National bureau of economic research Cambridge, Mass., USA, 1994. Cit. on pp. 95, 96, and 98.

SVENSSON, L. E. **Forward Guidance**. [*S.l.*], 2014. Cit. on pp. 83 and 88.

SVENSSON, L. E. O. What is wrong with Taylor rules? Using judgment in monetary policy through targeting rules. **Journal of Economic Literature**, American Economic Association, v. 41, n. 2, p. 426–477, 2003. Cit. on p. 68.

SWANSON, E. T. Have increases in Federal Reserve transparency improved private sector interest rate forecasts? **Journal of Money, Credit and Banking**, JSTOR, p. 791–819, 2006. Cit. on p. 63.

TABAK, B.; SOLLACI, A.; GOMES, G.; CAJUEIRO, D. Forecasting the yield curve for the Euro region. **Economics Letters**, v. 117, n. 2, p. 513–516, Nov. 2012. ISSN 01651765. DOI 10.1016/j.econlet.2012.05.056. Cit. on p. 96.

TABAK, B. M.; CAJUEIRO, D. O.; SERRA, T. R. Topological properties of bank networks: The case of Brazil. **International Journal of Modern Physics C**, v. 20, n. 08, p. 1121–1143, 2009. Cit. on p. 64.

TABAK, B. M.; SERRA, T. R.; CAJUEIRO, D. O. Topological properties of stock market networks: The case of Brazil. **Physica A: Statistical Mechanics and its Applications**, v. 389, n. 16, p. 3240–3249, 2010. Cit. on p. 64.

TAUSCH, F.; ZUMBUEHL, M. Stability of risk attitudes and media coverage of economic news. **Journal of Economic Behavior & Organization**, v. 150, p. 295–310, Jun. 2018. ISSN 01672681. DOI 10.1016/j.jebo.2018.01.013. Cit. on p. 34.

TAYLOR, J. B. International coordination in the design of macroeconomic policy rules. **European Economic Review**, Elsevier, v. 28, n. 1-2, p. 53–81, 1985. Cit. on p. 58.

TAYLOR, J. B. International monetary policy coordination: Past, present and future. BIS working paper, 2013. Cit. on pp. 58 and 64.

TEDESCHI, G.; IORI, G.; GALLEGATI, M. The role of communication and imitation in limit order markets. **The European Physical Journal B**, Springer, v. 71, p. 489–497, 2009. Cit. on p. 64.

TEDESCHI, G.; IORI, G.; GALLEGATI, M. Herding effects in order driven markets: The rise and fall of gurus. **Journal of Economic Behavior & Organization**, Elsevier, v. 81, n. 1, p. 82–96, 2012. Cit. on p. 64.

TETLOCK, P. C. Giving Content to Investor Sentiment: The Role of Media in the Stock Market. **The Journal of Finance**, v. 62, n. 3, p. 1139–1168, Jun. 2007. ISSN 0022-1082, 1540-6261. DOI 10.1111/j.1540-6261.2007.01232.x. Cit. on pp. 27, 32, and 34.

TETLOCK, P. C. Does Public Financial News Resolve Asymmetric Information? **Review of Financial Studies**, v. 23, n. 9, p. 3520–3557, Sep. 2010. ISSN 0893-9454, 1465-7368. DOI 10.1093/rfs/hhq052. Cit. on p. 34.

TETLOCK, P. C.; SAAR-TSECHANSKY, M.; MACSKASSY, S. More Than Words: Quantifying Language to Measure Firms' Fundamentals. **The Journal of Finance**, v. 63, n. 3, p. 1437–1467, Jun. 2008. ISSN 0022-1082, 1540-6261. DOI 10.1111/j.1540-6261.2008.01362.x. Cit. on pp. 28, 32, and 34.

TURC, I.; CHANG, M.-W.; LEE, K.; TOUTANOVA, K. Well-read students learn better: On the importance of pre-training compact models. **arXiv preprint arXiv:1908.08962v2**, 2019. Cit. on p. 49.

TURNEY, P. D.; PANTEL, P. From Frequency to Meaning: Vector Space Models of Semantics. **Journal of Artificial Intelligence Research**, v. 37, p. 141–188, Feb. 2010. ISSN 1076-9757. DOI 10.1613/jair.2934. Cit. on p. 40.

van Eck, N. J.; WALTMAN, L. Software survey: VOSviewer, a computer program for bibliometric mapping. **Scientometrics**, v. 84, n. 2, p. 523–538, Aug. 2010. ISSN 1588-2861. DOI 10.1007/s11192-009-0146-3. Cit. on p. 21.

van Eck, N. J.; WALTMAN, L. **Text Mining and Visualization Using VOSviewer**. [*S.l.*]: arXiv, 2011. DOI 10.48550/arXiv.1109.2058. Cit. on p. 21.

VASWANI, A.; SHAZEER, N.; PARMAR, N.; USZKOREIT, J.; JONES, L.; GOMEZ, A. N.; KAISER, L.; POLOSUKHIN, I. **Attention Is All You Need**. [*S.l.*]: arXiv, 2017. DOI 10.48550/arXiv.1706.03762. Cit. on pp. 16, 18, 21, 43, 44, 55, 92, 96, and 147.

VIEGAS, E.; GOTO, H.; TAKAYASU, M.; TAKAYASU, H.; JENSEN, H. J. Assembling real networks from synthetic and unstructured subsets: The corporate reporting case. **Scientific Reports**, v. 9, n. 1, p. 11075, Jul. 2019. ISSN 2045-2322. DOI 10.1038/s41598-019-47490-0. Cit. on p. 31.

VINYALS, O.; LE, Q. A neural conversational model. **arXiv preprint arXiv:1506.05869**, 2015. Cit. on p. 42.

WANG, C.; NULTY, P.; LILLIS, D. A comparative study on word embeddings in deep learning for text classification. *In*: **Proceedings of the 4th International Conference on Natural Language Processing and Information Retrieval**. [*S.l.*: *s.n.*], 2020. p. 37–46. Cit. on p. 43.

WANG, G.; WANG, T.; WANG, B.; SAMBASIVAN, D.; ZHANG, Z.; ZHENG, H.; ZHAO, B. Y. Crowds on Wall Street: Extracting Value from Collaborative Investing Platforms. *In*: **Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing**. Vancouver BC Canada: ACM, 2015. p. 17–30. ISBN 978-1-4503-2922-4. DOI 10.1145/2675133.2675144. Cit. on pp. 24 and 54.

WANG, T.; YUAN, C.; WANG, C. Does Applying Deep Learning in Financial Sentiment Analysis Lead to Better Classification Performance? **Economics Bulletin**, v. 40, n. 2, p. 1091–1105, 2020. Cit. on pp. 36 and 54.

WOLF, T.; DEBUT, L.; SANH, V.; CHAUMOND, J.; DELANGUE, C.; MOI, A.; CISTAC, P.; RAULT, T.; LOUF, R.; FUNTOWICZ, M.; DAVISON, J.; SHLEIFER, S.; von Platen, P.; MA, C.; JERNITE, Y.; PLU, J.; XU, C.; SCAO, T. L.; GUGGER, S.; DRAME, M.; LHOEST, Q.; RUSH, A. M. **HuggingFace's Transformers: State-of-the-art Natural Language Processing**. [*S.l.*]: arXiv, 2020. Cit. on pp. 49 and 147.

WRIGHT, J. H. What does Monetary Policy do to Long-term Interest Rates at the Zero Lower Bound? **The Economic Journal**, v. 122, n. 564, p. F447–F466, Nov. 2012. ISSN 0013-0133, 1468-0297. DOI 10.1111/j.1468-0297.2012.02556.x. Cit. on p. 58.

XING, F. Z.; CAMBRIA, E.; WELSCH, R. E. Natural language based financial forecasting: A survey. **Artificial Intelligence Review**, v. 50, n. 1, p. 49–73, Jun. 2018. ISSN 0269-2821, 1573-7462. DOI 10.1007/s10462-017-9588-9. Cit. on p. 53.

YANG, L.; ZHANG, Z.; XIONG, S.; WEI, L.; NG, J.; XU, L.; DONG, R. Explainable Text-Driven Neural Network for Stock Prediction. *In*: **2018 5th IEEE International Conference on Cloud Computing and Intelligence Systems (CCIS)**. Nanjing, China: IEEE, 2018. p. 441–445. ISBN 978-1-5386-6005-8. DOI 10.1109/CCIS.2018.8691233. Cit. on pp. 56 and 97.

ZENG, D.; LIU, K.; LAI, S.; ZHOU, G.; ZHAO, J. Relation Classification via Convolutional Deep Neural Network. *In*: TSUJII, J.; HAJIC, J. (Ed.). **Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers**. Dublin, Ireland: Dublin City University and Association for Computational Linguistics, 2014. p. 2335–2344. Cit. on pp. 29 and 30.

ZHANG, T.; WU, F.; KATIYAR, A.; WEINBERGER, K. Q.; ARTZI, Y. Revisiting few-sample BERT fine-tuning. **CoRR**, abs/2006.05987, 2020. Cit. on pp. 50 and 159.

ZHANG, W.; TANIDA, J.; ITOH, K.; ICHIOKA, Y. Shift-invariant pattern recognition neural network and its optical architecture. *In*: **Proceedings of Annual Conference of the Japan Society of Applied Physics**. [*S.l.*: *s.n.*], 1988. p. 2147–2151. Cit. on p. 22.

ZHAO, L.; LI, L.; ZHENG, X.; ZHANG, J. A BERT based Sentiment Analysis and Key Entity Detection Approach for Online Financial Texts. *In*: **2021 IEEE 24th International Conference on Computer Supported Cooperative Work in Design (CSCWD)**. Dalian, China: IEEE, 2021. p. 1233–1238. ISBN 978-1-72816-597-4. DOI 10.1109/CSCWD49262.2021.9437616. Cit. on pp. 23, 36, and 55.

ZUPIC, I.; ČATER, T. Bibliometric Methods in Management and Organization. **Organizational Research Methods**, v. 18, n. 3, p. 429–472, Jul. 2015. ISSN 1094-4281, 1552-7425. DOI 10.1177/1094428114562629. Cit. on pp. 19 and 20.

Appendices

Appendix A – Text Representation: BERT

As discussed in the main text, for textual data representation we adopt the Bidirectional Encoder Representations from Transformers (BERT) model (Devlin *et al.*, 2019), one of the most popular models that build from the state-of-the-art transformer architecture (Vaswani *et al.*, 2017). Figures A.1 and A.2 show how BERT integrates our neural networks' designs. Both of them show BERT incorporated into the networks the same way, but applied to each stage-specific monetary policy document: statements for the first stage and minutes for the second stage. Therefore, in both Figures we see that the node named "BERT Tokenizer" receives the entire documents as streams of words, with almost no previous processing, which is usually required in other natural language processing models. In fact, the tokenizer is responsible for all the preprocessing BERT needs, including breaking the text into smaller units for the model to work with — a step known as tokenization. The tokenizer also adds special tokens to the text, allowing the model to interact with them during training phase. One such token of great relevance for our framework is the [CLS] token depicted in our networks' illustrations. This token, whose name stands for "classification", is the responsible for carrying the overall meaning of the entire document[1]. This way, the embedding output related to the [CLS] token is the one we will work with, as it contains the contextual vector representation of the meaning of the entire document, as calculated by BERT. As for its dimensions, they follow the dimensions of the BERT model used. For our case, as Figures A.1 and A.2 depict, these are vectors in a 768-dimension space. When we get to the overall network design, we will have more detailed discussion about the use of this vector and how it allows us to measure communication state passed through our framework's stages, as we will be able to explore how different layers relate to each other.

The rationale behind our text representation model choice is related to how BERT incorporates transfer learning to deliver contextual-rich text embeddings. In that sense, the readily available pre-trained version of BERT provided by huggingface[2] (Wolf *et al.*, 2020) leverages trained parameters from large text corpus and allows us to fine-tune the network for our specific task on much smaller datasets. We thus fine-tune the same BERT model instance beginning on the first stage of the model, using policy decision statements, and then passing it along for further fine-tuning on second stage, when we focus on meeting minutes.

In practice, fine-tuning works by allowing BERT's parameters to also be adjusted in

[1] One should recall that, while documents in the first stage refer to the entire monetary policy decision statements, for the second stage the meeting minutes are segregated by section, thus having each one of them individually passed to the neural network depicted in Figure A.2.

[2] https://github.com/huggingface/transformers.

neural networks' training phases, along with parameters in the remaining layers (Devlin *et al.*, 2019). This approach is useful to make the text representations from the model more adapted to domain-specific applications, such as ours. However, still inspired on recommended uses of the BERT model discussed in its original paper, fine-tuning is not the only way we incorporate BERT in our framework. From first to second stages, we provide document embeddings as features, capturing monetary policy communication state as mentioned before. The main difference between fine-tuning and feature-based approaches is that, in the former, model's parameters are allowed to be further trained, whereas in the latter, model's output is provided to other networks or subsequent layers without the original model's parameters being altered.

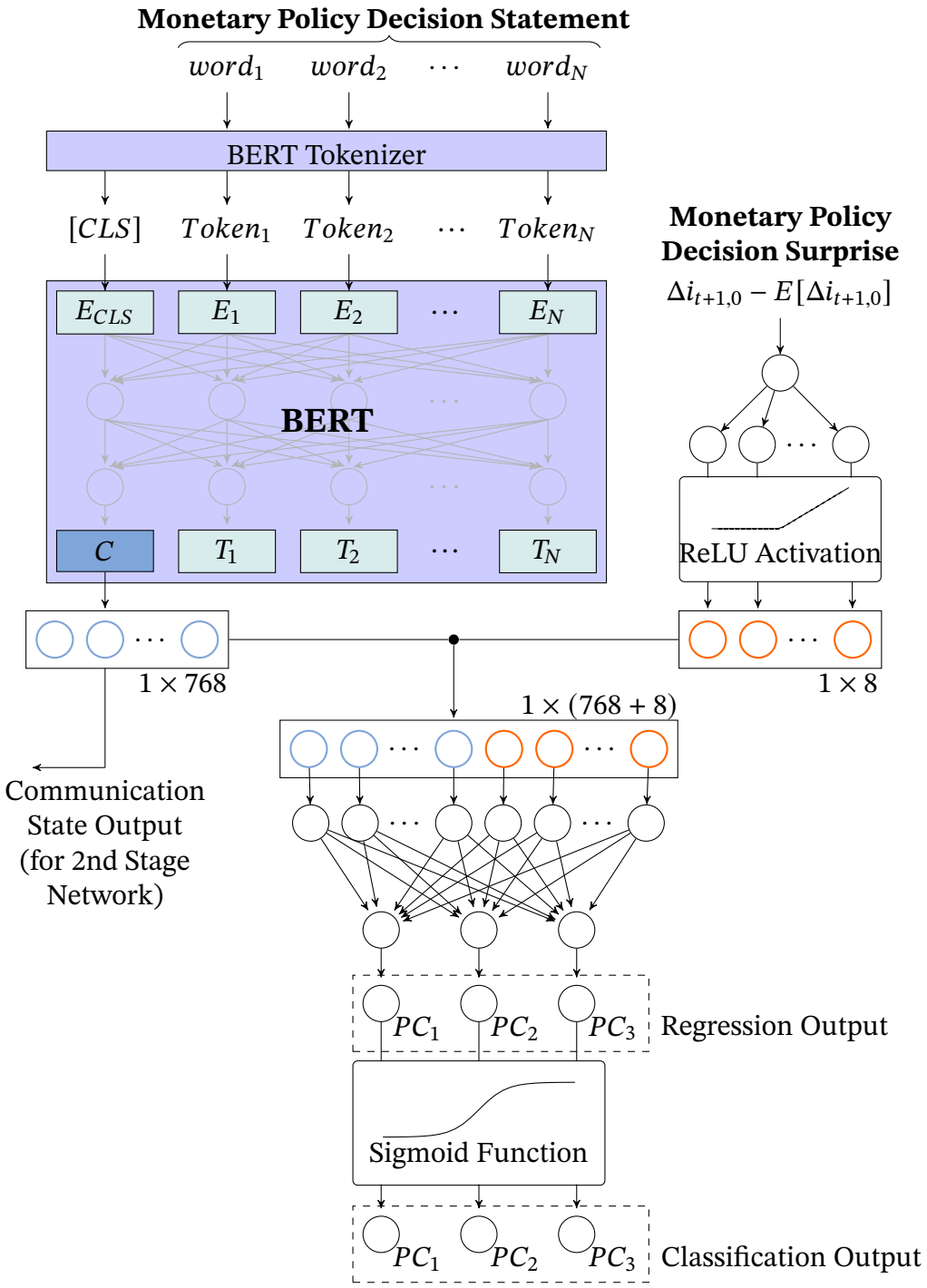


Figure A.1 – 1st Stage Model Neural Network

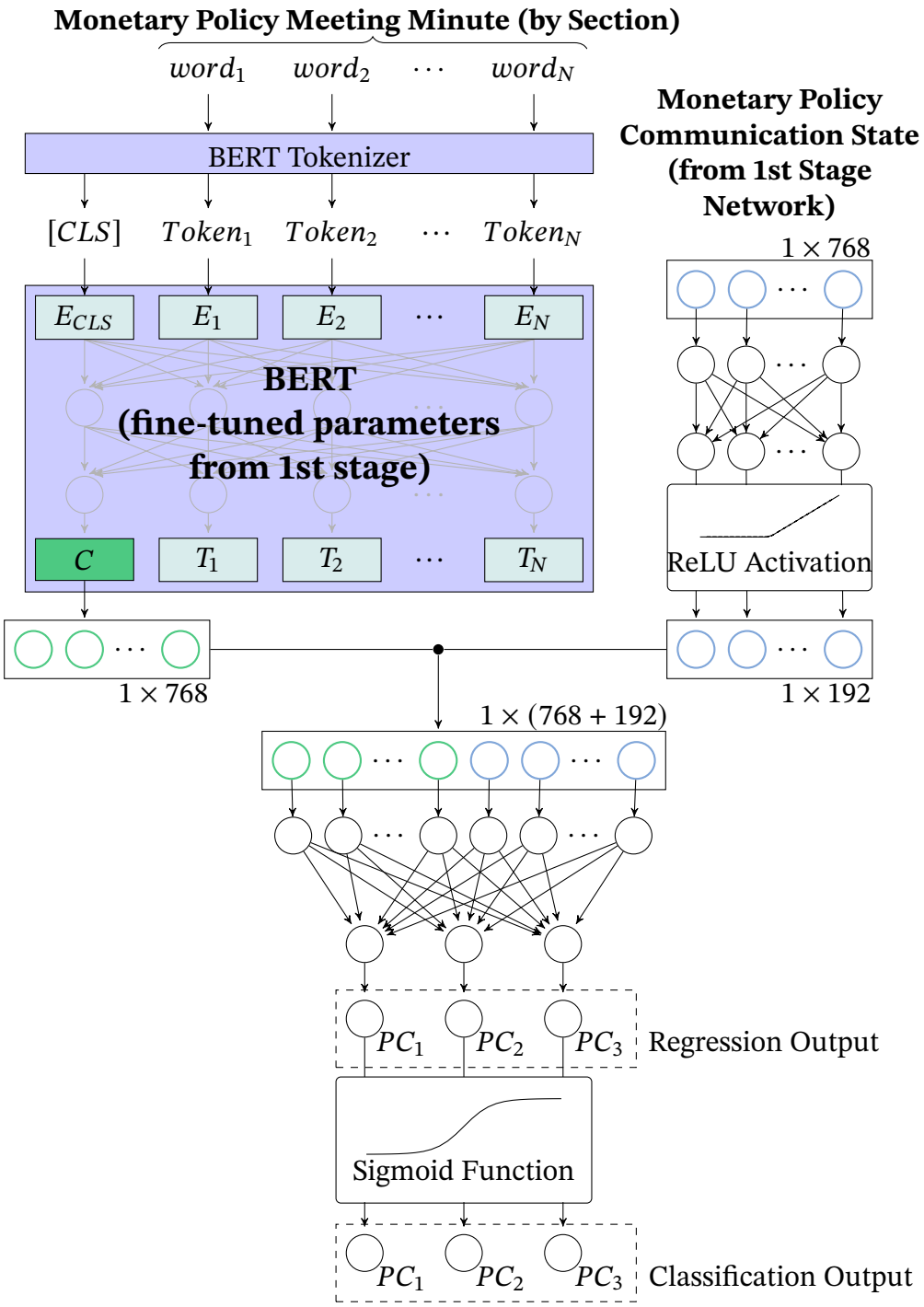


Figure A.2 – 2nd Stage Model Neural Network

Appendix B – Neural Networks' Designs

This Appendix is dedicated to detailing our model's neural networks. Networks for the two stages will be discussed simultaneously, as many modeling choices apply to both. Furthermore, by design our framework is intended to work either for classification or regression setups for sentiment extraction. Hence, we will also explore the differences in networks to address both of these approaches.

B.1 Input Layers

Based on our framework design from Figures A.1 and A.2, we will now discuss both stages' inputs. Beginning by the first stage, in which we model monetary policy decision statements — and recalling that our sentiment dimensions are associated with latent factors of interest rates as priced by markets — we note that monetary policy rate decision itself stands as an important information that should not be overlooked. More than that, decision surprises, as described in Section 4.3.4, are maybe the most important price movers for short-term yield curve adjustments. Monetary policy surprises are thus provided to the first stage of the model as a single dimension input, as represented by the right-hand side input node on Figure A.1.

The other input for the first stage network are statements themselves, in the form of textual data as discussed in Section 4.4. Since we are working with a transformer model, little preprocessing is required, leaving text as close as possible to original with just minor changes — such as converting text to lower case — in order for the network to learn relations and what to attend to. Therefore, the most important preprocessing step is text tokenization, converting terms to tokens known to the pre-trained model. As a matter of fact, the tokenizer functionality used in this step is made available alongside with pre-trained model, as both should be compatible. This functionality takes care of formatting the input as BERT expects[1]. Finally, statements are provided to the model as whole documents, with just some previous screening to remove disclosures unrelated to policy communication.

As previously noted, the second stage network, shown in Figure A.2, has a very similar design when compared to the first stage. The difference lies in the right-hand side input, with second stage network receiving communication state as a feature extracted from the first stage. The higher dimension of this input represents BERT's output embedding dimension, as detailed ahead. In practical terms, to calculate communication state we detach the first

[1] Technically, BERT's inputs are actually two vectors, which the tokenizer provides: the first contains the identification number of tokens that result from the tokenization step and the second takes care of masking irrelevant inputs, such as [PAD] tokens used to complete the expected sentence length.

stage fine-tuned BERT and score statements again. We also build the second stage network using the fine-tuned version of BERT from the first stage. This procedure leverages our two stage approach to further adapt textual embeddings to the monetary policy communication domain.

Moreover, passing communication state as an exogenous feature to second stage model results in added practical value to our framework. Despite being originally estimated based on near-term monetary decision statements, the fact that this feature assumes the form of BERT's contextual embeddings makes it possible to apply our second stage neural network to other sources of monetary policy communication. Therefore, as discussed in the main text, besides meeting minutes, the usual committee members' speeches or press manifestations can be used to extract monetary policy tone as well. This highlights that our framework stands as a flexible method to calculate the multi-dimension sentiment index from novel textual data, despite its source or frequency. The feature estimated from novel communication would then become the new communication state, providing input for next iterations. This way, our second stage neural network is designed to make it possible to be detached from the framework for continuous assessment of monetary policy sentiment.

B.2 Hidden and Output Layers

Among the hidden layers of our framework's neural networks in Figures A.1 and A.2, we should first turn our attention back to BERT's outputs. As BERT provides contextual embeddings and given that our goal is to extract sentiment from the entire document, we use the special [CLS] token placed in the beginning of documents to capture their contextualized representations, as recommended by Devlin *et al.* (2019)[2]. The [CLS] token, introduced in Apendix A, is the one that carries the embeddings for the entire document, which we have mentioned by the end of Section 4.5. These embeddings are vectors with the same dimension of the BERT model, which in our case is 768. These dimensions capture different aspects of text, some of which are expected to contain semantic clues to the tone in monetary policy communication. Therefore, still as recommended by Devlin *et al.* (2019), BERT's output vectors are passed to a fully connected layer for the model to learn the patterns associated with the sentiment contained in training data. However, before that, theses vectors are merged with other relevant features in order for the model to use a comprehensive set of information that may result in yield curve shifts. These additional features come from the other inputs of our deep learning architecture, which differ for first and second stage designs.

For the first stage, additional information assume the form of monetary policy decision

[2] Technically, this is done by using huggingface's BERT implementation with the so called pooler output, which consists of the special [CLS] token further passed through a fully connected layer with tanh activation function.

surprises, as the yield curve adjustments used to train the model are measured for the day that market participants incorporate this information in prices. This input passes through a fully connected layer with ReLU activation. The layer transforms the scalar measuring policy surprise into an 8-dimensional vector that, associated with the activation function, is expected to help the model incorporate different non-linear relations coming from policy surprises[3]. For instance, market participants may react to positive surprises in different ways that they do to negative surprises. Or they can react non-linearly to the magnitude of surprises, with small surprises being overlooked while larger surprises driving a complete economic outlook revision. The increased dimension of this hidden layer is intended to let the network learn such patterns present in data. Finally, ReLU activation, which is usually adopted for fully connected hidden layers, transforms each of the dimensions individually, leaving positive values unaltered and zeroing negative values:

$$ReLU(x) = \max(0, x). \tag{B.1}$$

As for the second stage, it has a somewhat similar treatment to communication state, the information received in addition to the document we want to extract sentiment from. As mentioned, this feature assumes the form of BERT's vector representation for previous communication, which is first stage's decision statement when we are working with monetary policy meeting minutes. Therefore, the main difference between this input and policy surprise is its dimension: now we are dealing with a 768-dimensional vector. This way, the fully connected hidden layer must have its dimension adjusted as well. Accounting for that, the hidden layer through which the communication state in the second stage network is passed reduces its dimension to one fourth of the original size, as depicted in Figure A.2, resulting in a 192-dimensional vector[4]. Again, the layer includes ReLU activation, with the same rationale as discussed before. It is also important to note that, even though we do not explicitly input the innovations in communication as a feature itself, the design of our framework allows the deep learning model to learn non-linear patterns of such innovations from the communication state provided. The hidden layer just discussed is crucial for that design.

With all the information properly processed by each network, the merged vectors are passed to the classification layers. At this point, we have decision statement embeddings merged with decision surprise vector for the first stage; and meeting minute embeddings merged with communication state for the second. Hereon, both stages act the same, using a fully connected layer to combine all the information received to produce three outputs, one

[3] The dimension of this hidden layer is actually one of the hyperparameters set during network design, with the 8-dimensional vector producing the best results.

[4] Similarly to the first stage network, the dimension of this hidden layer is a hyperparameter with value chosen according to the best achieved results.

for each of the term structure of interest rates factors calculated from principal components. The result is thus a vector in $\mathbb{R}^3$, which represents the expected shifts along the yield curve's level, slope and curvature, and is the final output when we use the framework for regression tasks. Finally, when the model is configured for binary classification problems, this output is further passed through a sigmoid activation function, which transforms each dimension individually according to Eq. B.2. The new output assumes the form of probabilities p_k of yield curve factor increase in dimension k.

$$S(x) = \frac{1}{1 + e^{-x}}. \tag{B.2}$$

Appendix C – Data Labeling

The first step to understand the models' training strategy is to detail our data labeling approach, as this is how we will inform the model what are the expected outcomes for parameter optimization. As we have discussed in the main text, the financial literature usually adopts one of two alternative approaches to label data for supervised learning tasks. The first involves doing it manually, usually by one or more experts annotating each observation. The second resorts to market data to infer sentiment from price reactions to studied events. In this case, labels can be a continuous variable representing price changes, a binary variable indicating prices reacting by moving up or down (usually 1 for price increase and 0 for price reduction) or a categorical variable, indicating levels of price changes (as in positive, neutral or negative changes, for example). In our case, we follow the market data approach, considering communication release as the event driving price movements. Furthermore, since we are interested in a multidimensional sentiment index, the labels derive from changes in each of yield curve's factor scores.

Data labeling is one of the key aspects by which regression and binary classification approaches differ. When we set our model for binary classification problems, training data labels should be binary variables informing the direction of factor movement. Therefore, we set the labels to 1 when factor scores increase and to 0 when they reduce in days that markets react to communication events. It is interesting to recall that, from the nature of yield curve and how its factors capture its movements, factor score increase (reduction) represent negative (positive) sentiment. This is not adjusted for in the labeling process, making our sentiment index's interpretation coherent with yield movements. Formally, Eq. C.1 is the expression used to calculate binary classification labels, $y_{k,t}^{class}$ for factor dimension k in time t, using factor shifts in Eq. 4.5:

$$y_{k,t}^{class} = \begin{cases} 1, & \Delta z_{k,t} > 0 \\ 0, & \Delta z_{k,t} \leq 0. \end{cases} \tag{C.1}$$

As for when we configure our framework for regression problems, labels should be in the form of continuous variables. We thus use standardized score shifts, which produces improved regression results over raw shifts. This process is done in the same train-score fashion as described in Section 4.3 for yield curve standardization, so we do not incorporate future market information in any given time of our dataset[1]. Standardized scores are also

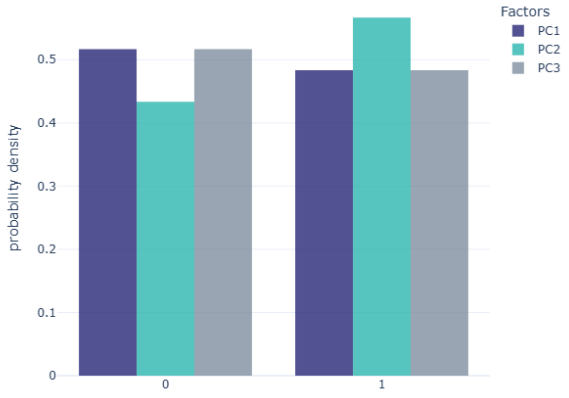
[1] Recall that we use a period prior to the beginning of our study to calculate standardization parameters, namely mean and standard deviation, and use them to standardize time series during the study period. This makes sure that we do not use future information at any moment for yield curve modeling, as discussed in Section 4.3.2, and sentiment modeling as discussed here.

not adjusted for sentiment direction, assuring consistency of interpretation with binary classification and coherence with yield curve movements. Eq. C.2 shows how we calculate regression labels $y_{k,t}^{regr}$ again from factor shifts in Eq. 4.5:

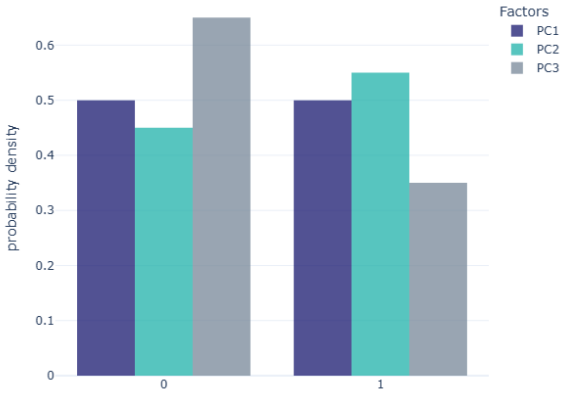
$$y_{k,t}^{regr} = \mathcal{Z}_{\Delta z}^{train}(\Delta z_{k,t}), \tag{C.2}$$

where $\mathcal{Z}_{\Delta z}^{train}(\cdot)$ is the standardization function based on mean and standard deviation calculated for Δz during training period.

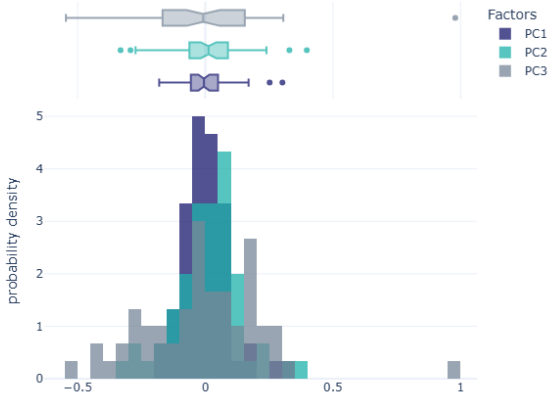
Figure C.1 shows the distributions of labels across all three factor dimensions for first stage statements and for second stage minutes and for the binary classification and regression setups. We can see that this data is rather balanced across every dimension and thus do not require any further processing. These three dimension variables computed from yield curve factor movements are then provided to the framework as labels associated with each textual data document for our supervised training step.



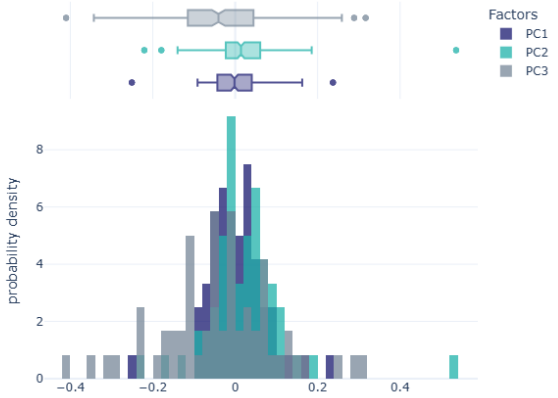
(a) Classification - Statements



(b) Classification - Minutes



(c) Regression - Statements



(d) Regression - Minutes

Figure C.1 – Label Distribution

Appendix D – Training

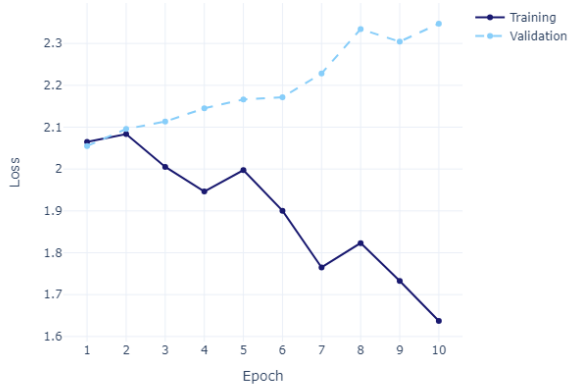
D.1 Binary Classification

As for the framework training strategy, we will first discuss the general aspects along with the binary classification setup. Binary classification is the most usual setup for sentiment analysis, given that manually labeled data usually assumes the form of binary or, at most, categorical data. The main challenge here was to work with a relatively small dataset when compared to the usual deep learning application in natural language processing. To tackle that issue, we began by following the main guidelines provided by Devlin *et al.* (2019) in setting hyperparameters. However, BERT fine-tuning on small datasets is known for being highly dependent on network's hyperparameters (Zhang *et al.*, 2020), the reason why we ended up experimenting on a broad range of alternatives, guided by the recommendations of both Devlin *et al.* (2019) and Zhang *et al.* (2020). All in all, our framework's final hyperparameters are set as follows:

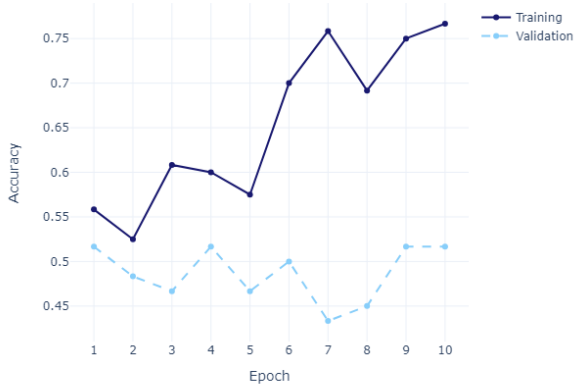
- Network architectures of both stages, along with each layer's dimensions, are the ones depicted in Figures A.1 and A.2;
- BERT implementation is $\text{BERT}_{\text{BASE}}$ (Devlin *et al.*, 2019);
- Number of yield curve factors, thus the output dimension, is set to $K = 3$;
- For both stages, data is randomly split into train and test sets, with proportion of 0.67 to 0.33. Important to note that, despite having data up to December 2024, we train our model with data only up to the end of 2023, in order to leave new data unseen by the model in training phase;
- Number of epochs is set to 10 for both stages, larger than Devlin *et al.* (2019) usual approach and in line with evidence of Zhang *et al.* (2020), however much smaller than their results for small sample BERT fine-tuning;
- Batch size is set to 8, trying to further improve generalization from reduced size;
- Dropout rate is set to 0.5 in every dropout node, except from BERT's output in first stage, which is set to 0.25. These high numbers are fundamental to assure that the model will not overfit from overtraining, since we have increased number of epochs;
- Optimizer follows BERT's recommended AdamW (Kingma; Ba, 2017; Loshchilov; Hutter, 2019);
- Learning rate is set to $2e^{-5}$ in both stages (Devlin *et al.*, 2019);
- Loss function is binary cross entropy, widely adopted for multilabel classification problems in machine learning. It is described by Eq. D.1:

$$\frac{1}{N}\sum_{n=1}^{N} y_n \cdot \log p_n + (1 - y_n) \cdot \log(1 - p_n), \tag{D.1}$$

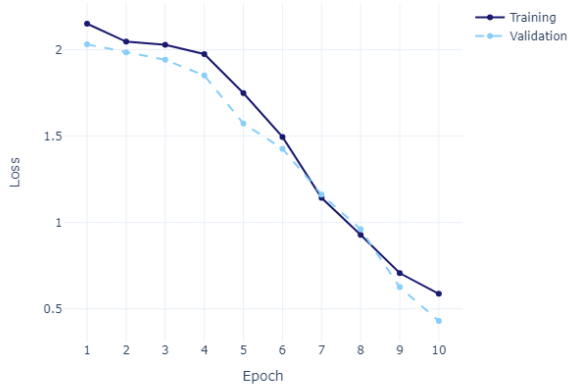
where N is the total number of classifications, given by number of observations times number of factors; p_n is the model's predicted probability and y_n is actual output label;



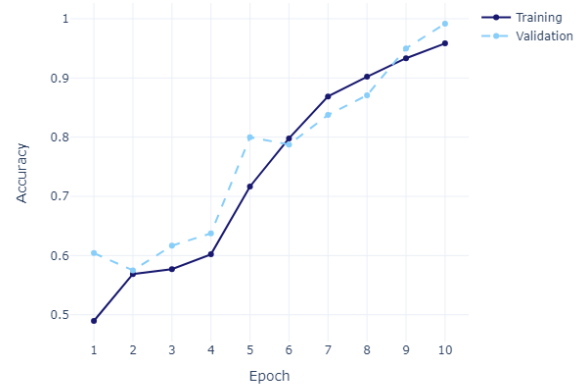
(a) Loss 1st Stage



(b) Accuracy 1st Stage



(c) Loss 2nd Stage



(d) Accuracy 2nd Stage

Figure D.1 – Binary Classification Model Training-Validation Metrics

Figure D.1 shows fine tuning phase loss and accuracy for training and validation data splits for both stages. From it, we can see that the main part of our model, the second stage network — which delivers the final output — presents very good performance both in and out-of-sample. We can see from Panel D.1c that convergence of the loss function is stable and from Panel D.1d that accuracy by the end of the 10-epoch training phase achieved over 95% in training and in validation splits.

D.2 Regression

We should recall that the regression approach differs from binary classification in the fact that the predicted output becomes a continuous variable for each factor, instead of a binary indicator of increase or decrease. This approach is possible given our market-driven data labeling strategy. The regression setup follows the same overall training strategy described and hyperparameters listed in Appendix D.1. The exceptions, besides the labeling process detailed in Appendix C and the network design explained in Appendix B, is in loss and evaluation metrics. In addition to that, we also doubled the number of epochs from the binary classification approach and set the dropout rate for BERT's output in first stage to 0.5, matching the rate used to train other layers.

For evaluation metric, we use the R^2 score, a standard in the machine learning literature for regression evaluation. For loss, we use the also widely adopted mean squared error (MSE), given by Eq. D.2. Noteworthy, loss metric adjustment is specially important since it is used for network parameters' optimization.

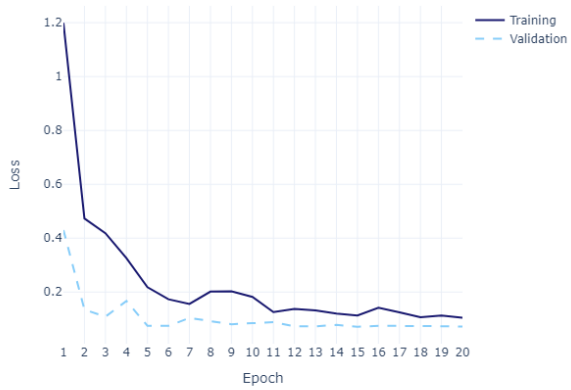
$$MSE = \frac{1}{N}\sum_{n=1}^{N}(y_n - \hat{y}_n)^2, \tag{D.2}$$

where N is the total number of outputs, given by number observations times number of factors, and y_n and $\hat{y}_n$ are actual and predicted factor shifts, respectively.

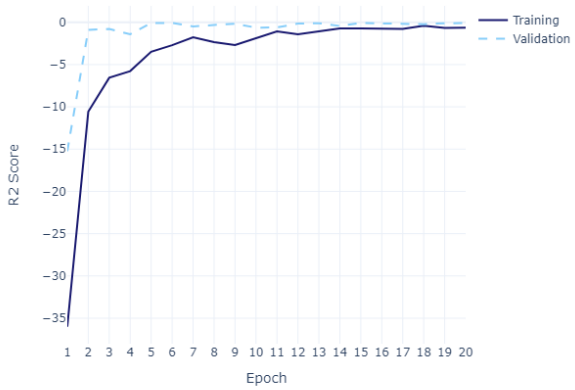
Regression training statistics are presented in Figure D.2. First, we can see from second stage metrics in Panels D.2c and D.2d that the increased number of epochs is necessary to assure converge. In terms of model evaluation, we see from Panel D.2d that out-of-sample R^2 score converged to over 80% by the end of the experiment, corroborating the good performance from the previous classification exercise. We can also see that validation metrics presented better results than training metric as shown in every panel of Figure D.2. Despite seemingly counterintuitive, this pattern is in fact a direct results of how large dropout rates we have to adopt to prevent the model from overfitting from the combination of small dataset and increased number of epochs. One must recall that dropout is active during training but turned out for validation, resulting in a more comprehensive model for the latter. As consequence, we can see that our training strategy successfully delivered a set of parameters of consistent performance.

D.3 Single Stage Regression

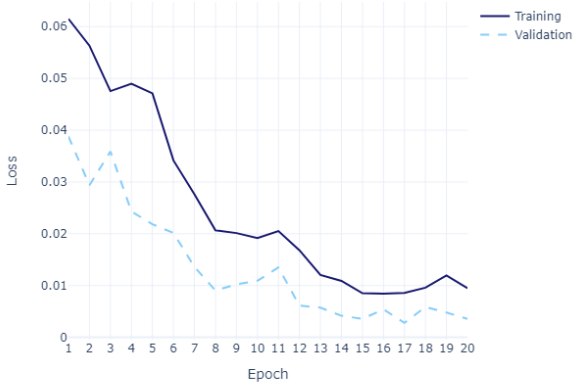
In order to evaluate the contribution of our two stage approach to model's performance, we ran an alternative regression experiment, including only the second and final stage of the baseline framework. In this benchmark experiment, the neural network has access only



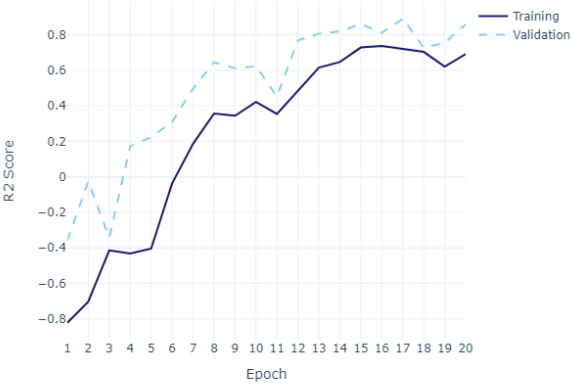
(a) Regression Loss 1st Stage



(b) Regression R2 1st Stage



(c) Regression Loss 2nd Stage



(d) Regression R2 2nd Stage

Figure D.2 – Regression Model Training-Validation Metrics

to the current communication released by the central bank in the form of monetary policy meeting minutes, without information of communication state know by economic agents. The deep learning model is thus adapted to expect only the textual input, as shows the additional neural network representation in Figure D.3.

The benchmark regression experiment has its results shown in Figure D.4. First, we can see a similar pattern in terms of dropout effects over train and validation sets. Moreover, final results do not differ much, with our baseline two stage framework outperforming benchmark setup by a close margin in both loss and R^2 score. However, we can see from the comparison of both panels in Figure D.4 to Panels D.2c and D.2d that convergence is dramatically affected by our two stage framework. As a matter of fact, even the scales of the plots differ in order of magnitude.

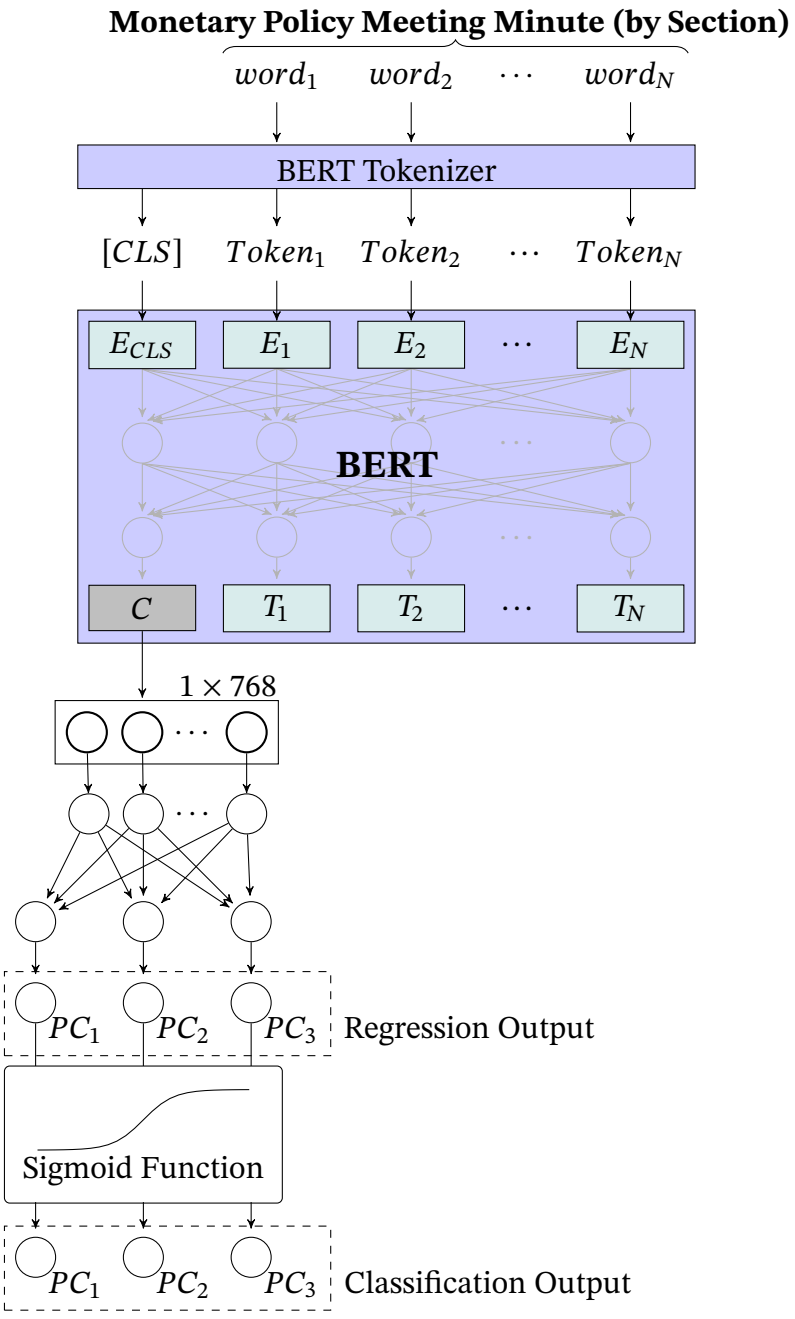
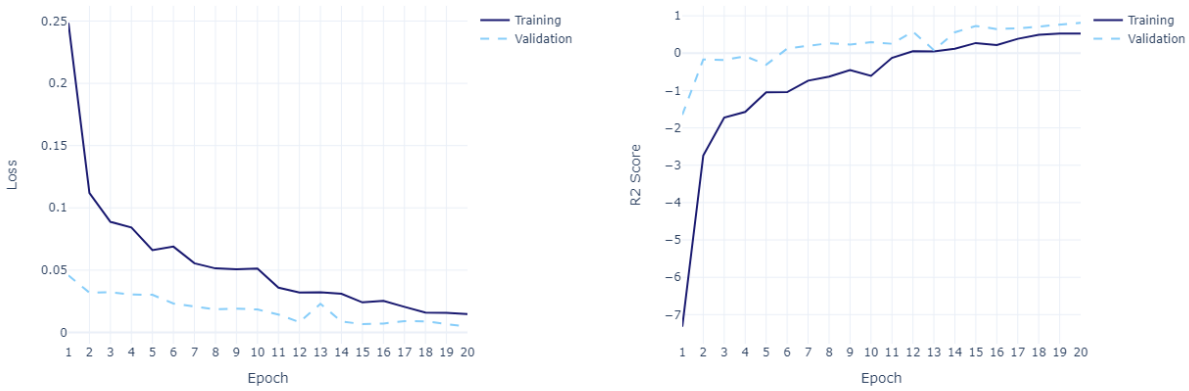


Figure D.3 – Single Stage Model Neural Network



(a) Regression Benchmark Loss

(b) Regression Benchmark R2

Figure D.4 – Regression Benchmark Training-Validation Metrics


UnB