TALES OF TURBULENCE: BERT-BASED MULTIMODAL ANALYSIS OF FED COMMUNICATION DYNAMICS AMIDST COVID-19 THROUGH FOMC MINUTES

TÜRBÜLANSIN HİKAYESİ: COVID-19 DÖNEMİNDE FED İLETİŞİM DİNAMİKLERİNİN FOMC TUTANAKLARI ÜZERİNDEN BERT TEMELLİ HİBRİT ANALİZİ

BİLAL TAŞKIN

ASST. PROF. DR. FUAT AKAL
Supervisor

Submitted to
Graduate School of Science and Engineering of Hacettepe University
as a Partial Fulfillment to the Requirements
for the Award of the Degree of Master of Science
in Computer Engineering

September 2023

ABSTRACT

TALES OF TURBULENCE: BERT-BASED MULTIMODAL ANALYSIS OF FED COMMUNICATION DYNAMICS AMIDST COVID-19 THROUGH FOMC MINUTES

Bilal Taşkın

Master of Science , Computer Engineering

Supervisor: Asst. Prof. Dr. Fuat AKAL

September 2023, 82 pages

Although communication of the central banks with the market was still quite limited at the end of the 20th century, and the authorities had agreed that policymakers should hold a more secretive attitude, this situation changed after the 2000s, in light of the Australian example. From there on, the view that decisions should spawn surprise in the market to increase the effectiveness of monetary policy has been replaced by the discourse that claims that the strategy, policy, and short- and long-term goals of the central banks should be precisely understood by the public. Moreover, communication tools have come to the fore as vital support for the monetary policy decision-making processes, especially in countries that have adopted inflation-targeting regimes. Consequently, today, communication is of great importance for the central banks to realize their mission. This study analyzes Federal Open Market Committee (FOMC) minutes using state-of-the-art Natural Language Processing (NLP) techniques. We sought to investigate the effect of the global COVID-19 crisis on the FOMC minutes' pattern and the strength of the Federal Reserve to influence inflation expectations through its primary press releases. To this end, we first quantified minutes leveraging domain-specific pre-trained Bidirectional Encoder

Representations from Transformers models (FinBERTs). Then, we applied Dynamic Time Warping (DTW) to measure temporal sequence proximity over the course of time. To verify our findings, we built multivariable Autoregressive Integrated Moving Average models by injecting an exogenous variable as an indicator function into the time series (ARIMAX). The results suggest that the Federal Reserve has abstained from adjusting its tone and the forward-lookingness setting of its statements for the global epidemic. Therefore, the longstanding association of fed tone and forward-lookingness with consumer inflation expectation has weakened during the crisis.

Keywords: Natural Language Processing, Monetary Policy, Central Banking Communication, COVID-19, Semantic Analysis

ÖZET

TÜRBÜLANSIN HİKAYESİ: COVID-19 DÖNEMİNDE FED İLETİŞİM DİNAMİKLERİNİN FOMC TUTANAKLARI ÜZERİNDEN BERT TEMELLİ HİBRİT ANALİZİ

Bilal Taşkın
Yüksek Lisans, Bilgisayar Mühendisliği
Danışman: Asst. Prof. Dr. Fuat AKAL
Eylül 2023, 82 sayfa

20. yüzyılın sonlarına kadar hakim olan merkez bankacıların ve politika yapıcıların ketum bir davranış sergilemesi gerektiğine dair inanç, 2000'li yılların başından itibaren Avustralya örneğinin ışığında değişime uğramıştır. Böylelikle, alınan kararların piyasalarda sürprizle karşılanması fikri yerini merkez bankalarının stratejilerinin, politikalarının, kısa ve uzun vadeli hedeflerinin kamu yararından tam olarak anlaşılması gerektiği görüşüne bırakmıştır. Bunun da ötesinde, özellikle enflasyon hedeflemesi rejimi uygulayan ülkelerde iletişim araçları politika yapım sürecinin vazgeçilmez bir parçası haline gelmiştir. Sonuç olarak, günümüzde merkez bankalarının misyonlarını gerçekleştirmeleri için iletişim büyük önem taşımaktadır. Bu tez, yeni nesil Doğal Dil İşleme (NLP) tekniklerini kullanarak Federal Açık Piyasa Komitesi (FOMC) tutanaklarını analiz etmektedir. Daha spesifik olmak gerekirse, çalismamizda modern semantik analiz tekniklerinden yararlanarak küresel Covid-19 krizinin FOMC iletişimi üzerindeki etkisini ve Federal Rezerv'in birincil basın bültenleri aracılığıyla enflasyon beklentilerini etkileme gücünü araştırdık. Bu amaçla, önce alana özel önceden eğitilmiş BERT (Bidirectional Encoder Representations from Transformers-Dönüşümleyicilerden Çift Yönlü Kodlayıcı Temsilleri) modellerinden

(FinBERT'ler) yararlanarak FOMC tutanaklarını sayısallaştırdık. Ardından, zamansal dizi yakınlığını zaman boyunca ölçmek için Dinamik Zaman Bükme (DTW) tekniğini uyguladık. Bulgularımızı doğrulamak için, zaman serisilerine indikatör bir fonksiyon enjekte ederek çok değişkenli Otoregresif Entegre Hareketli Ortalama modelleri (ARIMAX) oluşturduk. Sonuçlar, Federal Rezerv'in tonunu ve açıklamalarının ileriye dönüklük derecesini küresel salgın döneminde ayarlamaktan kaçındığını göstermektedir. Bu durum, tüketici enflasyon beklentisi ile uzun dönemden süre gelen FED iletişim tonu ve tutanakların ileriye dönüklük derecesi arasındaki ilişkide zayıflamayı beraberinde getirmiştir.

Keywords: Doğal Dil İşleme, Para Politikasi, Merkez Bankaciligi İletişimi, COVID-19, Semantik Analiz

CONTENTS

Page

TABLES

FIGURES

1. INTRODUCTION

The significance of communicative strategies and the development of apt channels for these strategies has been accentuated since the adoption of inflation-targeting frameworks by numerous central banks from the 1990s onward [2]. The economic scholarly community widely acknowledges that enduring price stability necessitates long-term policy planning. On the other hand, central banks also endeavor to foster consistent national output growth, minimize unemployment, and ensure the smooth operation of financial markets. However, a central bank with the goal of maintaining price stability must adopt a long-term viewpoint, making decisions free from political influences or other groups typically characterized by short-term objectives. The autonomy of central banks in terms of policy formation and institutional operations comes hand in hand with the obligation of being accountable [3]. Thus, it is imperative to establish a robust communication strategy to uphold the transparency of the central bank [4].

Explaining the reasons behind monetary policy decisions and sharing them with the public helps everyone better understand and support these decisions. This also sets clear expectations about the goals of these policies, especially if they're trustworthy. As Muth points out, *"Economic agents make optimum decisions by utilizing all of the information that is available to them"*. This means they use both new and prior information to make decisions [5]. And what is more, Goodfriend suggests that when central banks are more predictable, it makes it easier for markets to anticipate and react to monetary policies, leading to better economic outcomes [6]. In essence, clear communication and being transparent can make monetary policies work better. How well we manage expectations plays a big role in connecting clear communication and the success of monetary policies [7].

Whenever central bankers speak, financial markets are all ears. It has been a lengthy journey for these banks to perfect how they relay their goals and intentions [8]. In today's age, there is an abundance of monetary policy data for financial experts who hang on every word from these banks. Central banks have made their processes transparent: they hold press briefings, release meeting summaries, offer detailed inflation insights, make public announcements, and even address legislative assemblies. Though there is a budding effort to communicate with the broader public via social media platforms like Twitter, the main attention is still on the familiar faces: the financial markets and specialists. Yet, there are moves by certain central banks to enhance the financial understanding of the household members, tapping into the potential of current technology. An initiative that stands out in this effort is ”Economics for All” by the Central Bank of the Republic of Turkey.

Central banks' primary challenge in conveying their messages to the general public is the intricate language they employ, often couched in market-specific terminologies. Recent research suggests that more informal modes of communication tend to be more effective, especially in elucidating monetary policy stances [9]. Yet, central banks continue to prioritize their complex official releases as the main avenue for communication. The intricacy of this situation is humorously encapsulated in a remark made by Alan Greenspan, the former chair of the FED, during an address to a corporate audience as follows, “I guess I should warn you, if I turn out to be particularly clear, you have probably misunderstood what I said.” [10].

Over the past decade, the presence of significant improvements in Natural Language Processing (NLP) has been notably felt in various facets of our daily interactions [11]. With advancements in artificial intelligence technology, natural language processing has steadily improved, delivering more precise and consistent outcomes. Furthermore, Text Mining, a closely related sub-domain, aids in deciphering the nuanced language components present in textual data.

While text mining is widely utilized in fields like politics and marketing, its adoption in economics has historically been more restrained [12]. Yet, text mining could be invaluable for discerning the underlying intentions and perspectives concealed within central bank announcements, given their profound impact on the financial sector [13]. Sentiment analysis, also termed Opinion Mining, employs a text-mining technique designed to detect and classify subjective sentiments within source content, utilizing text analytics, computational linguistics, and natural language processing. Sentiment analysis endeavors to determine an author's stance on a topic or gauge the overall tonal sentiment of a piece of communication [14].

Sentiment analysis is rooted in two primary methodologies: dictionary-based approaches and machine learning-based techniques[1]. The dictionary-based methods rely on a reference list where words are tagged with predefined sentiments and associated scores. This approach's limitation stems from the intricate nuances of language, including metaphors, idioms, and proverbs, making it challenging to encapsulate emotions using rigid parameters. However, a notable merit of dictionary-based systems is that they operate without the necessity for extensive datasets. On the other hand, machine learning techniques, contingent upon the specific algorithm implemented, mandate substantial data for training and often exhibit improved performance as the data input increases.

Supervised and unsupervised learning stand as the two principal approaches within machine learning, each distinguished by its training process and data prerequisites. The choice between supervised and unsupervised methods often depends on the particular characteristics of the issue under consideration. Supervised learning hinges on labeled data for its training process, guiding the model towards specific outcomes. In contrast, unsupervised algorithms operate under the presumption that input data adheres to an unidentified statistical distribution. The goal is to discern this distribution's characteristics [15]. Metrics like intra-cluster and inter-cluster distances, predicated on certain criteria (like cluster size), come into play here. While there are emerging models like semi-supervised learning that operate under relaxed constraints, they still necessitate certain assumptions about the data distribution to optimize results. Such assumptions might relate to the data's inherent distribution patterns, whether in terms of continuity, clustering, or manifold structures.

Occasionally, dictionary-based methods might be categorized as supervised due to the presence of an inherent "ground truth" in lexicons. However, if an analysis employs static weights from a lexicon without any adaptability, it does not fall under either supervised or

[1]Though there's a subtle distinction between a lexicon and a dictionary in linguistic terms, we employ them synonymously in this context.

unsupervised learning. There are no weight adjustments, and therefore, no genuine learning takes place, rendering it distinct from traditional machine learning paradigms.

In this research, we applied edge-cutting text mining and semantic analysis methods to elicit quantitative information about monetary policy tendency and foresight from the FOMC minutes between February 1993 and November 2022. We mainly questioned the potential shifts in the communiqués.

Texts that follow Zipf's Law are typically easier to read and understand because they prioritize the most important words and efficiently use language [16]. Therefore, we initially conducted a Zipfian compliance inspection to see whether FOMC releases comply with Zipf's Law and any significant deviation from its historical outlook. We also glanced at the releases' readability aspect utilizing prevalent readability algorithms. As the last part of the exploratory analysis, we brought out the topic intensity evolution of the documents to discover the general view of the FOMC agenda over time. Thereafter, we analyzed the sentimental and forward-looking inclination of the minutes. Since general-purpose dictionaries are one-size-fits-all solutions, and generally, central bank communiqués have a high-formality level and incorporate technical expressions, we embraced a machine-learning approach. Specifically, we exploited a financial domain-specific pre-trained NLP model, FinBERT, to unearth the semantic orientation of the minutes [17]. We further ran another investigation on the forward-lookingness of the meeting releases utilizing another pre-trained and financial domain-specific BERT model, FinBERT FLS (Forward-lookingness), which looks for soothsaying ingredients through the texts and returns both types of forward-lookingness and its intensity [18].

To observe the possible impacts of global health crisis on the relationship between the communication tone of the FED and market movements, we implemented a well-known signal processing technique, Dynamic time warping (DTW). We measured the pair-wise similarity of the time series as temporal sequences, i.e., Consumer Inflation Expectations (CIE) for one-year and three-year ahead horizon and time series derived from the minutes exclusively. Finally, we modeled each time series using an autoregressive integrated moving average method with a Covid-19 indicator as an exogenous variable.

We use advanced Natural Language Processing (NLP) methods, specifically using ”domain-specific pre-trained Bidirectional Encoder Representations from Transformers” models (FinBERTs), to analyze the Federal Open Market Committee (FOMC) minutes. Through the application of Dynamic Time Warping (DTW) for evaluating the proximity of temporal sequences and the construction of multivariable Autoregressive Integrated Moving Average models (ARIMAX), we depart from previous research by investigating the influence of the global Covid-19 crisis on the structure of FOMC minutes and the Federal Reserve's capacity to impact inflation expectations through its primary press releases. The results unveil that the Federal Reserve has maintained a consistent tone and forward-looking perspective in its statements throughout the crisis, challenging the previously established connection between the Fed's tone, forward-looking stance, and consumer inflation expectations. In conclusion, this thesis offers valuable insights into the evolving dynamics of central bank communication and its implications for monetary policy within the context of a global crisis.

The thesis is structured in the following manner: In Chapter 1, we outline our motivation, contributions, and the thesis's scope. Chapter 2 offers a comprehensive background, encompassing the methods and techniques that form the basis of our methodology. Chapter 3 reviews the literature of opinion mining and polarity exploration research on financial and economics-related texts, particularly central bank releases. In Chapter 4, we present the research methodology, provide a concise overview of the data, delve into the identification of abstract topics, and present the results of exploratory analyses. Chapter 5 presents the empirical path for semantic and forward-lookingness modeling. Chapter 6 reports the results. Chapter 7 concludes the results, draws some policy implications, and suggests future routes for further research.

-

2. BACKGROUND OVERVIEW

Natural Language Processing (NLP) is an expanding research realm situated at the crossroads of artificial intelligence and linguistics. Its core aim is to develop computational models and algorithms that enable machines to grasp, interpret, and fabricate human language in a meaningful and coherent fashion. NLP seeks to narrow the divide between human communication and machine comprehension, ushering in numerous applications across domains like information retrieval or sentiment analysis. By arming computers with the capability to process and scrutinize extensive volumes of text data, NLP holds the potential to fundamentally transform our interactions with technology and our access to information.

The development of NLP has been propelled by advancements in machine learning, particularly deep learning, and the availability of large-scale datasets. With the advent of transformer models like BERT and GPT-4, which leverage pre-training and fine-tuning techniques, NLP has witnessed significant breakthroughs in various language understanding and generation tasks (see Figure 2.1). These models have demonstrated state-of-the-art performance in tasks such as text classification, named entity recognition, sentiment analysis, and machine translation. Consequently, NLP has garnered substantial interest from academia, industry, and research communities, giving rise to many applications reliant on language processing and comprehension. The continuous advancements in NLP techniques and their practical applications hold immense potential for transforming various sectors, improving user experiences, and enabling new avenues for innovation and discovery.

While NLP encompasses a wide array of techniques and methodologies, one of the critical components underpinning language understanding is semantic analysis. Semantics delves into the meaning behind words and how context can influence this meaning, a crucial consideration when training models to comprehend language nuances. As the subsequent sections will explore, semantic analysis is not a monolithic approach; instead, it can be achieved through various methods. Two of the most prevalent approaches are 'machine learning methods' and 'lexicon-based methods.' Machine learning methods leverage

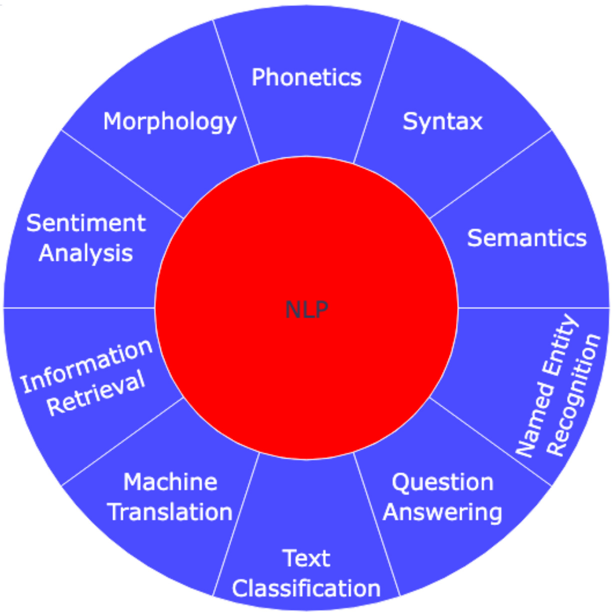


Figure 2.1 Various subfields within the domain of Natural Language Processing (NLP)

data-driven techniques, especially those amplified by the recent advances in deep learning. On the other hand, lexicon-based methods, which have their roots in linguistic studies, rely on predefined lists or dictionaries of words paired with their semantic orientations. Each approach has its strengths and weaknesses, but together they form the bedrock of our current understanding and implementation of sentiment analysis in NLP.

In the ensuing parts of this section of the thesis, we will embark on a comprehensive journey exploring the intricacies of sentiment analysis. Delving deep into its primary techniques and approaches, we will demystify the algorithms and methodologies that power this crucial aspect of textual data analysis. Following our exploration of sentiment, we will pivot to the realm of temporal data, providing insights into temporal sequence similarity measurements and the vast domain of time series analysis. These techniques are pivotal in understanding patterns, trends, and sequences in temporal data, offering a nuanced perspective on information that evolves over time. Lastly, but certainly not least, we will turn our attention to the foundational stage of any analytical process: data preprocessing, and BERT (Bidirectional Encoder Representations from Transformer) models in greater detail as well. This segment will shine a light on the vital techniques employed to cleanse, transform, and prepare data, ensuring its optimal utility for subsequent analysis.

2.1. Sentiment Analysis

Sentiment analysis, also known as sentiment understanding or sentiment processing, is a branch of NLP that focuses on extracting meaning and understanding from text or speech. Together with the smart algorithms adjusted to the semi-structured nature of text data it goes beyond the surface-level analysis of words and sentences and delves into the deeper context and interpretation of language. In other words, the objective of sentiment analysis is to enable machines to comprehend and interpret human language in a way that is similar to how humans understand it.

Various techniques and algorithms have been employed to analyze the relationships between words, phrases, and sentences and extract the underlying meaning for opinion mining. By applying advanced algorithms and machine learning models, the sentiment analysis aims to derive the intended meaning, context, and implications behind the text, facilitating a deeper level of understanding and enabling applications such as question-answering systems, sentiment analysis in social media, document classification, and information retrieval. In a nutshell, sentiment analysis plays a crucial role in bridging the gap between human language and machine understanding, enabling more sophisticated and intelligent interactions between humans and machines. In the following part, we explore the main methods used for sentiment analysis in NLP, focusing on machine learning and lexicon-based approaches.

2.1.1. Machine Learning Methods

Machine learning techniques have played a transformative role in advancing sentiment analysis tasks in NLP. These methods allow computers to automatically learn patterns and extract meaningful representations from large-scale datasets. It involves training models on labeled data to learn patterns and relationships in language, enabling them to automatically understand and interpret the meaning of text or speech.

In machine learning-based sentiment analysis, the process typically involves several steps. First, a dataset with labeled examples is collected, where each example is associated with the desired sentiment analysis task (e.g., sentiment classification, entity recognition). Then, features are extracted from the text, which can include word embeddings, syntactic information, or other relevant linguistic features. These features are used as input to machine learning algorithms such as support vector machines (SVM), random forests, or deep learning models like recurrent neural networks (RNNs) or transformers.

During the training phase, the model learns the relationships between the input features and the target sentiment analysis task. The model is optimized by adjusting its internal parameters to minimize the difference between the predicted outputs and the ground truth labels in the training data. Once the model is trained, it can be used to predict the semantic properties or interpret the meaning of unseen text or speech data. Several prominent machine-learning methods have been employed in sentiment analysis:

2.1.1.1. Supervised Learning Supervised learning is a commonly employed methodology in sentiment analysis, which involves training a model using annotated data. Each instance in the data is associated with a specific sentiment analysis task, such as sentiment analysis where each instance is labeled as positive, negative, or neutral. Through the training process, the model captures the underlying patterns and relationships between the input text and the corresponding labels. Various algorithms including Support Vector Machines (SVM), Naive Bayes, Decision Trees, and Random Forests, as well as deep learning models like Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) can be utilized. Once trained, the model can then be utilized to predict the semantic properties of new, unseen text.

2.1.1.2. Word Embeddings Word embeddings are condensed vector portrayals of words crafted to encapsulate intricate semantic connections among them. These representations are forged by subjecting models to extensive textual data sets with the aim of forecasting the contextual utilization of words. Prominent algorithms, including Word2Vec, GloVe, and FastText, have found widespread use in creating these word embeddings. The resulting embeddings encode semantic similarities and interconnections, thus empowering various sentiment analysis tasks, including word sense disambiguation, sentiment similarity assessment, and entity recognition. By leveraging word embeddings, models can substantially enhance their understanding of word semantics and contextual nuances within a given text. In Figure 2.2, we present the core concepts of word embeddings. This 2D illustrative space visually demonstrates the fundamental principle behind embeddings: words that share semantic meanings are closer in the vector space. While the actual embeddings often reside in higher-dimensional spaces, this simplified representation provides a clear intuition of the underlying idea.

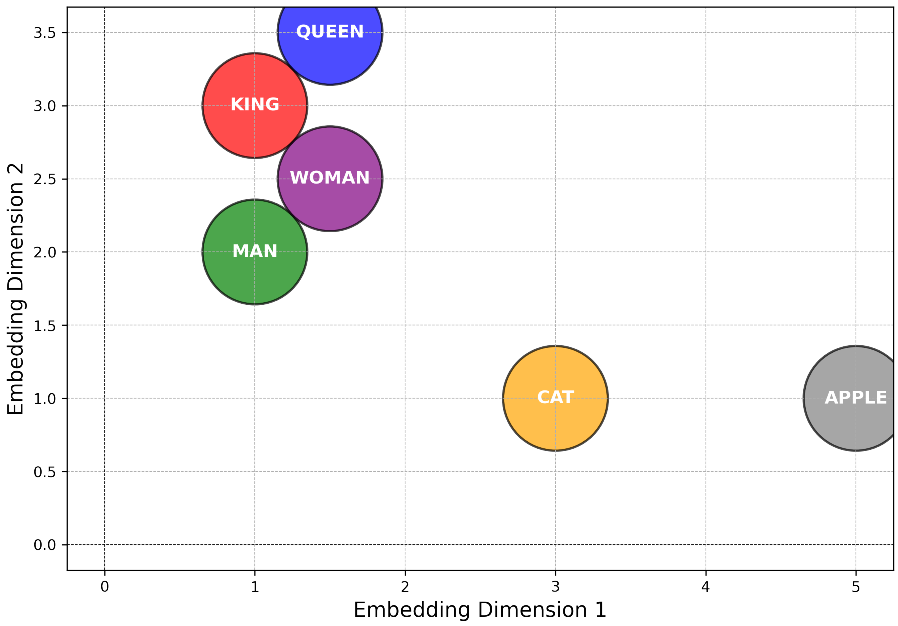


Figure 2.2 Conceptual visualization of word embeddings in a 2D space. Words with perceived semantic similarities are positioned closer together. This diagram provides an illustrative representation and does not depict actual embeddings derived from trained models.

2.1.1.3. Recurrent Neural Networks (RNNs) RNNs, or Recurrent Neural Networks, belong to a class of neural networks that demonstrate strong effectiveness in modeling sequential data. They excel at capturing intricate dependencies and temporal relationships among words within a given text. The nature of RNNs involves sequential processing, where each word is sequentially presented as an input, along with the hidden state inherited from the preceding word. Notably, variants of RNNs such as Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRUs) have gained popularity due to their ability to mitigate the vanishing gradient problem and effectively model long-term dependencies. In various tasks where the contextual information and word order significantly impact the results, RNNs find extensive application.

2.1.1.4. Transformers Transformers have sparked a revolution in the field of natural language processing, profoundly impacting sentiment analysis tasks. The Transformer architecture, epitomized by cutting-edge models like BERT (Bidirectional Encoder Representations from Transformers), has exhibited remarkable prowess across diverse applications. By harnessing self-attention mechanisms, Transformers possess the unique ability to capture long-range dependencies by simultaneously attending to all positions within an input sequence. This attention-based approach enables Transformers to effectively gauge the importance of individual words or phrases in their contextual surroundings. Consequently, Transformers have demonstrated exceptional proficiency in tasks such as question answering, text classification, and language generation. Leveraging the power of pre-trained models like BERT, practitioners can fine-tune these models to achieve state-of-the-art performance in specific sentiment analysis tasks.

2.1.1.5. Transfer Learning Transfer learning has emerged as a cornerstone technique in modern machine learning and, more specifically, in Natural Language Processing (NLP). At its heart, transfer learning is the practice of leveraging knowledge gleaned from pre-trained models on vast datasets, subsequently applying this acquired knowledge to specific tasks, such as sentiment analysis, which might have less available data. This method is particularly powerful due to the inherent capability of pre-trained models to encapsulate general linguistic patterns, intricate syntactic structures, and deep semantic relationships drawn from extensive training corpora.

Popular models like BERT or GPT stand as testament to the potency of transfer learning. Initially trained on colossal corpora spanning billions of words, these models embed a comprehensive understanding of language. Yet, their true versatility shines when they are fine-tuned on more specific, smaller labeled datasets for downstream tasks, such as sentiment detection in product reviews or emotion classification in texts. By doing so, the models do not start their learning journey from scratch. Instead, they build upon the linguistic foundation laid during their pre-training phase, aligning their vast knowledge to the nuances and specificities of the target task.

The benefits of this approach are manifold. Not only does transfer learning lead to improved performance in tasks like sentiment analysis, but it also substantially reduces the training time, as the models are not starting from a point of complete ignorance. This efficiency is especially valuable when labeled data for the target task is scarce—a frequent challenge in NLP endeavors.

2.1.1.6. Ensemble Methods In the ever-evolving landscape of sentiment analysis, ensemble methods have risen to prominence, offering sophisticated strategies to enhance model predictions. These techniques hinge on the principle of harnessing the collective power of multiple models to offset individual weaknesses and amplify strengths. By integrating a variety of models, ensemble methods aspire to achieve higher robustness, increased accuracy, and a more nuanced understanding of sentiment across diverse datasets.

Among the pantheon of ensemble strategies, Bagging, Boosting, and Stacking stand distinguished, each providing a unique approach to the assembly of models. Bagging—or Bootstrap Aggregating—diversifies its approach by training multiple instances of a model on different subsets of the training data. These instances then collaboratively decide on the final prediction, often through an averaging mechanism or a majority vote. Such an approach is instrumental in reducing variance, offering a more stable prediction landscape.

Boosting, contrastingly, is dynamic in its approach. Models are trained sequentially, with each iteration paying particular attention to the mistakes made by the previous one. This iterative correction ensures that challenging data points, which might be misclassified initially, receive amplified attention in subsequent models, leading to a progressive refinement of the prediction accuracy.

Stacking, perhaps the most intricate of the trio, involves layering models. Initial models—often from diverse algorithmic families—make predictions, and a subsequent meta-learner then learns from these predictions to make the final decision. This hierarchical approach leverages the expertise of multiple models, ensuring that the diverse strengths of individual models contribute to a more robust and refined prediction.

The value proposition of ensemble methods in sentiment analysis is undeniable. Beyond the obvious benefits of enhanced accuracy, they play a pivotal role in countering overfitting, a notorious challenge in machine learning. Moreover, by leveraging multiple models, ensemble methods inherently imbue the prediction process with a capacity to handle uncertainty, ensuring more consistent and reliable sentiment interpretations across varied textual landscapes.

2.1.1.7. Neural Attention Mechanisms Neural attention mechanisms have rapidly become an indispensable facet of advanced machine learning architectures, pioneering a paradigm shift in how models process and interpret vast swathes of data. Rooted in the human cognitive process of selective focus, attention mechanisms infuse neural networks with an ability to discern and prioritize pertinent information, much like how our brains zoom in on critical details amidst a deluge of stimuli.

In the context of sentiment analysis, attention mechanisms serve as adept curators of information. Traditional neural models, while powerful, often suffer from a myopic understanding of lengthy texts, losing critical nuances as they progress through sentences. Enter attention mechanisms, which empower these models to continuously recalibrate their focus, dynamically weighting words or phrases based on their contextual relevance. This selective magnification ensures that pivotal sentiments—whether subtly embedded or glaringly apparent—are duly recognized and incorporated into the model's interpretation.

Beyond sentiment analysis, the versatility of attention mechanisms has been showcased in a plethora of NLP tasks. Machine translation, for instance, has been revolutionized by the introduction of attention. Instead of processing source and target languages in isolation, attention-equipped models can now draw intricate relationships between individual words or phrases across languages, resulting in translations that are not just linguistically accurate but also contextually enriched. Similarly, in text summarization, these mechanisms enable models to distill voluminous content into concise summaries, cherry-picking the most salient details without sacrificing coherence or context.

The beauty of attention mechanisms lies in their adaptability. Whether incorporated into the loops of Recurrent Neural Networks (RNNs) or the multi-headed architectures of Transformers, they augment these structures with a refined sense of discernment. Particularly in the latter, as seen in models like BERT and GPT, attention mechanisms have been pivotal in capturing long-range dependencies, knitting together disparate pieces of information to form a cohesive understanding of textual narratives.

2.1.1.8. Graph-based Methods Natural Language Processing (NLP) has long grappled with the challenge of effectively capturing and representing the intricate web of relationships that underpin textual content. One of the most compelling responses to this challenge has been the advent of graph-based models. These models stand at the intersection of linguistic semantics and structured data representation, aiming to provide a holistic view of language's interconnectedness.

At the core of a graph-based model lies the graph itself, a structured representation that seeks to encapsulate the multifaceted associations among words or entities. Within this graphical framework, individual nodes serve as anchors that represent specific words, phrases, or more broadly, entities (see Figure 2.3). These nodes are not isolated islands; instead, they are intricately connected via edges. These edges, far more than mere connectors, signify semantic relationships. Whether it is a syntactic relationship, like that of a subject to its verb, or a deeper semantic connection, like synonymy or antonymy, these edges capture the essence of how language components relate to one another.

In practical applications, such as sentiment analysis, graph-based models offer unique advantages. Consider a product review that praises a product's design but critiques its functionality. A graph-based model could effectively discern the contrasting sentiments by mapping positive associations to the design aspect and negative ones to functionality, all while understanding the intricate relationship between these components. Such granularity is pivotal in ensuring sentiment analysis results that are not only accurate but also contextually rich.

Beyond sentiment analysis, the potential applications of graph-based models in NLP are vast. Knowledge graphs, for instance, use a similar principle to relate entities and facts, providing a structured representation of general knowledge. Another application is in text summarization, where graphs can help in identifying the most salient pieces of information by mapping dependencies and importance levels among textual elements.

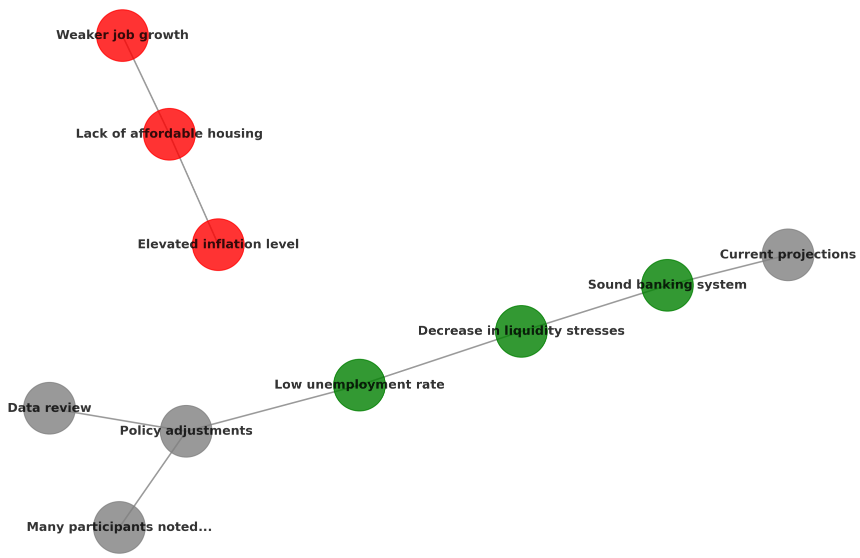


Figure 2.3 A representation of sentiment propagation in graph-based analysis. Each node represents a statement or opinion, with colors indicating the sentiment: green (positive), red (negative), and gray (neutral). Edges illustrate the interconnectedness or influence between statements. The graph exemplifies how sentiments can be influenced by and spread across related nodes in complex conversational structures.

Diving deeper into the methodologies, graph-based approaches exhibit a spectrum of strategies tailored for different NLP tasks:

2.1.1.9. Adjacency Matrices At its core, an adjacency matrix provides a binary representation indicating the presence or absence of a connection (edge) between pairs of nodes. In sentiment analysis, for instance, this could map syntactic relationships between words in a sentence.

2.1.1.10. Spectral Clustering This technique leverages the eigenvalues of the Laplacian matrix derived from a graph to segment it into clusters. In the context of NLP, this can aid in grouping semantically similar words or documents.

2.1.1.11. Random Walks Simulating a path that moves from one node to another randomly, random walks can aid in tasks like word sense disambiguation, where the model determines the most probable meaning of a word based on its neighboring context.

2.1.1.12. Graph Neural Networks (GNNs) Infusing traditional neural network concepts with graph-based structures, GNNs have the ability to propagate information across nodes, making them apt for tasks where context from distant words or entities needs to be considered.

2.1.1.13. PageRank Originally designed for ranking web pages, this algorithm evaluates the importance of nodes within a graph. In NLP, it can help prioritize the significance of words or entities in larger texts.

While the aforementioned are just a few approaches under the vast canopy of graph-based methods, they collectively underscore the versatility and depth of these models. Whether it's extracting sentiment, clustering documents, or disambiguating word meanings, the structured, relational nature of graphs endows them with a unique capability to harness the multifaceted dimensions of language.

2.1.2. Lexicon-based Methods

Lexicon-based approaches in NLP involve utilizing pre-existing lexical resources, such as dictionaries and thesauri, to analyze and understand the semantic characteristics of textual data. These methods are especially beneficial in scenarios where there is a scarcity of labeled data or when addressing domain-specific tasks. Prominent lexicon-based techniques employed in NLP include:

2.1.2.1. Sentiment Lexicons Sentiment analysis commonly relies on the utilization of sentiment lexicons, which encompass words annotated with their respective sentiment polarities, namely positive, negative, or neutral. Through the process of matching words from the text with entries in the lexicon, sentiment polarity can be assigned to the text. Prominent lexicons such as SentiWordNet and VADER (Valence Aware Dictionary and sEntiment Reasoner) are extensively employed in sentiment analysis tasks. These lexicons serve as a foundational resource for sentiment classification and aid in discerning the overall sentiment conveyed within a given text.

2.1.2.2. WordNet WordNet, a pervasive lexical database, serves as a fundamental resource in the organization of words into synsets, which encompass groups of synonymous words. This comprehensive framework not only embraces semantic associations like hypernymy (is-a), hyponymy (part-of), and meronymy (member-of) but also plays a pivotal role in a myriad of natural language processing (NLP) endeavors. These encompass, but are not limited to, word sense disambiguation, semantic similarity assessment, and information retrieval. By harnessing the power of WordNet, NLP systems gain an enhanced ability to comprehend the intricate nuances and contextual subtleties of words, thereby empowering more precise and refined sentiment analyses.

2.1.2.3. FrameNet FrameNet is a lexical resource that is centered around capturing the semantic nuances of words by examining their functions within conceptual frames. These frames encompass a collection of interconnected words and the specific semantic roles they fulfill in varying contexts. FrameNet proves particularly valuable in tasks such as semantic role labeling and information extraction. By establishing associations between words and frames, along with their corresponding roles, FrameNet enhances the comprehension of semantic connections within textual data, thereby facilitating a more profound analysis of textual meaning.

In essence, sentiment analysis plays a critical role in natural language processing (NLP) by enabling machines to extract meaning and comprehend the contextual nuances of human language. Within this domain, both lexicon-based methods and machine-learning approaches hold significant importance. Lexicon-based methods, such as the utilization of sentiment lexicons, WordNet, and FrameNet, offer valuable insights into semantic relationships and contribute to a deeper understanding of textual meaning. However, the landscape of sentiment analysis has been substantially transformed by the advent of machine learning techniques.

Machine learning approaches, encompassing methodologies like word embeddings, neural networks, and deep learning models, have emerged as powerful tools for advancing sentiment analysis tasks in NLP. These techniques have exhibited remarkable performance gains and have become integral in diverse NLP applications. By harnessing the potential of machine learning, NLP systems can effectively capture and analyze the intricate semantics inherent in human language, resulting in enhanced performance and more precise interpretation of textual data.

2.2. Temporal Sequence Proximity

Temporal sequence proximity refers to the assessment of the nearness or resemblance between events or elements arranged in a time-oriented sequence. This fundamental concept plays a pivotal role across diverse domains, enabling the identification of patterns, dependencies, or associations that are specific to the temporal dimension. Nevertheless, comparing temporal sequences with disparate frequencies poses a significant challenge due to disparities in time scales and sampling rates.

The Euclidean distance serves as a straightforward metric for proximity, computing the direct distance between corresponding points in the sequences. It quantifies the dissimilarity between two sequences by considering the disparities in their values at each time instance. Smaller Euclidean distances indicate higher similarity. However, the comparison of temporal sequences with varying frequencies holds utmost importance in numerous scenarios. For instance, in the domain of finance, it aids in the detection of market trends and correlations between assets traded at distinct frequencies. In healthcare, it enables the analysis of patient data collected with irregular frequency. In the context of language modeling, it assists in comprehending the temporal dynamics of text, such as the occurrence of words or the evolution of topics over time. Hence, the development of methodologies for measuring the proximity of sequences with different frequencies becomes imperative for accurate analysis and interpretation of temporal data. To tackle the challenge of comparing sequences with disparate frequencies, several techniques have been devised:

2.2.1. Down-sampling

Down-sampling refers to the process of decreasing the frequency of a higher-frequency sequence to align it with a lower-frequency sequence. This adjustment is typically achieved by either averaging or decimating the higher-frequency data points. On the other hand, up-sampling involves increasing the frequency of the lower-frequency sequence by employing interpolation methods or advanced techniques like Fourier analysis. By aligning the sequences at a comparable time scale, down-sampling and up-sampling techniques facilitate the measurement of proximity between the sequences, allowing for meaningful analysis of their relationships.

2.2.2. Dynamic Time Warping (DTW)

Dynamic Time Warping (DTW) is a widely utilized technique in temporal sequence analysis to quantify the similarity or proximity between sequences of different lengths or time scales. It enables flexible alignment by mapping corresponding elements of the sequences, accommodating variations in timing or duration. The underlying concept behind DTW is that sequences may exhibit similar patterns or shapes while experiencing differences in speed or temporal progression. DTW addresses this challenge by warping or stretching the sequences in the time dimension to achieve the best alignment that captures the underlying similarities. This adaptive alignment capability of DTW is particularly valuable in scenarios where precise temporal correspondence is essential for accurate analysis and interpretation of temporal data. Through the application of DTW, researchers and practitioners gain a powerful tool to delve into and comprehend the intricate dynamics present in diverse temporal sequences.

DTW operates by aligning two temporal sequences in a dynamic programming framework. The basic idea is to find an optimal warping path that aligns the sequences in a way that minimizes the overall cost of warping. The cost of warping is typically defined as the dissimilarity or distance between corresponding data points in the sequences. The algorithm finds the path with the minimum cumulative cost, taking into account possible shifts in time and scaling of the sequences.

DTW can handle time series data with different frequencies, as it allows for local time shifts, meaning that it can align data points that occur at different time points in the sequences. This makes it particularly useful for comparing sequences with irregular or asynchronous time intervals. Additionally, DTW can handle sequences with different lengths, as it allows for partial alignments and does not require sequences to have the same number of data points.

One of the advantages of DTW is its flexibility in capturing complex temporal patterns in time series data. It can capture both linear and nonlinear relationships between data points, making it suitable for a wide range of applications, such as speech recognition, gesture recognition, and financial time series analysis. Intuitively, Figure 2.4 represents the logic behind DTW by comparing it with its Euclidian counterpart.

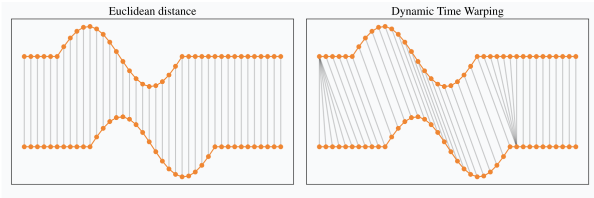


Figure 2.4 Graphical representation of the comparison between Dynamic Time Warping and Euclidean distance [1]

Given two time series, denoted as $X = x_1, x_2, ..., x_n$ and $Y = y_1, y_2, ..., y_n$, where n and m are the lengths of X and Y, and x_i and y_i represent the values of the time series at the time i and j, respectively. DTW aims to find a mapping between the two time series X and Y, which minimizes the cumulative distance between corresponding data points, subject to certain constraints [19]. Let y_j as $d(x_i, y_j)$ be the distance between two data points x_i, which could be any valid distance metric, such as Euclidean distance, Manhattan distance, or any other distance measure based on the problem domain. The DTW algorithm computes a similarity matrix, denoted as $D = d(i, j)$, where $d(i, j)$ represents the distance between x_i and y_j. This similarity matrix has the exact dimensions as the time series data X and Y.

The DTW algorithm uses a dynamic programming approach to compute the similarity matrix D in an iterative manner. The key idea is to compute the minimum cumulative distance from the starting point $(0, 0)$ to each point (i, j) in the similarity matrix D, by considering the minimum of the three possible previous points: $(i-1, j)$, $(i, j-1)$, and $(i-1, j-1)$.

The recurrence relation for computing the similarity matrix D is given by:

$$D(i,j) \ = \ d(x_i, y_j) \ + \ min(D(i-1,j), \ D(i,j-1), \ D(i-1,j-1)) \tag{1}$$

where $D(i, j)$ represents the minimum cumulative distance from the starting point $(0, 0)$ to the point (i, j) in the similarity matrix, and $d(x_i, y_j)$ represents the distance between x_i and y_j. Once the similarity matrix D is computed, the DTW distance between two time series X

and Y is given by the value at the bottom-right corner of the similarity matrix, i.e., $D(n, m)$, where n and m are the lengths of X and Y, respectively.

2.2.3. Wavelet Transform

The motivation behind the Wavelet Transform arises from the need to analyze signals using functions that exhibit both temporal and frequency localization. In contrast to traditional Fourier analysis, which employs global sinusoidal functions that offer frequency information but lack time localization, the Wavelet Transform capitalizes on functions with localized properties.

The fundamental concept underlying the Wavelet Transform is the recognition that signals often encompass localized features or events that possess significance for analysis. By employing wavelet functions that exhibit temporal localization, the Wavelet Transform adeptly captures these localized features.

From a mathematical standpoint, the Wavelet Transform entails a sequence of convolution and downsampling operations. The wavelet function, which is small and localized, undergoes shifting and scaling operations across the signal to extract valuable frequency information over time.

The selection of a specific wavelet function, the determination of decomposition levels, and the adjustment of other parameters can be tailored to the signal's characteristics and the analysis objectives. Different wavelet functions offer distinct properties and are well-suited for specific types of signals or applications. Additionally, the Wavelet Transform is frequently employed in conjunction with Downsampling techniques to reduce data size and focus on the most pertinent frequency components.

By providing a representation that encompasses both time and frequency domains, the Wavelet Transform has found extensive use in diverse domains, including signal processing, image analysis, time-series analysis, and others, where localized analysis and the capture of transient features are essential considerations.

2.3. Time-Series Modeling

Time-series modeling is a fundamental methodology employed to analyze and predict data that exhibits temporal dynamics. It encompasses a diverse set of statistical and machine-learning techniques designed to capture the inherent dependencies and patterns inherent in sequential data. These models play a vital role in numerous domains, such as finance, economics, weather forecasting, and stock market analysis, to name a few.

A time series can be described as a sequence of data points that are collected at regular intervals of time. It represents the progressive evolution of a variable or system as time progresses. Time-series data typically comprises two fundamental components:

- *Temporal Order:* The observations in a time series are ordered chronologically, with a clear time index associated with each data point.
- *Temporal Dependencies:* Time series often exhibit dependencies, meaning that the value at a given time point is influenced by the values at previous time points. These dependencies can manifest as trends, seasonality, cyclic patterns, or other forms of correlation.

There are several types of time-series models, each suited for different data characteristics and modeling objectives:

2.3.1. Autoregressive Models (AR)

These models consider that the value at a given time point depends linearly on past observations. AR models capture the trend and serial correlation in the data.

An autoregressive model of order $p(AR(p))$ represents the value of a variable at a given time point as a linear combination of its past values up to lag p. Mathematically, an $AR(p)$ model can be expressed as:

$$Y(t) = c + \phi(1)Y_{(t-1)} + \phi(2)Y_{(t-2)} + ... + \phi(p)Y_{(t-p)} + \epsilon(t) \tag{2}$$

Here, $Y_{(t)}$ denotes the value of the variable at time t, c represents a constant term, $\phi(1), \phi(2), ..., \phi(p)$ are the autoregressive coefficients, $Y_{(t-1)}, Y_{(t-2)}, ..., Y_{(t-p)}$ correspond to the lagged values of the variable, and $\epsilon_{(t)}$ denotes the error term at time t.

The autoregressive coefficients quantify the impact of the respective lagged values on the current value. Positive values indicate a positive correlation, while negative values indicate a negative correlation. The magnitude of the coefficients determines the strength of the relationship between the variable and its past values, with larger magnitudes indicating a stronger influence.

2.3.2. Moving Average Models (MA)

MA models assume that the value at a given time point depends on the weighted average of past error terms. They capture short-term changes and random fluctuations. A Moving Average model of order $q(MA(q))$ expresses the value of a variable at a given time point as a weighted sum of the most recent q error terms. Mathematically, an $MA(q)$ model can be represented as:

$$Y_{(t)} = \mu + \epsilon_{(t)} + \theta(1)\epsilon_{(t-1)} + \theta(2)\epsilon_{(t-2)} + ... + \theta(q)\epsilon_{(t-q)} \tag{3}$$

Here, $Y_{(t)}$ denotes the value of the variable at time t, μ represents the mean of the series, $\epsilon(t)$ represents the error term at time t, and $\theta(1), \theta(2), ..., \theta(q)$ are the MA coefficients. The MA coefficients determine the influence of the corresponding lagged error terms on the current value.

The MA coefficients quantify the impact of the respective lagged error terms on the current value. Positive values indicate a positive correlation, while negative values indicate a negative correlation. The magnitude of the coefficients determines the strength of the influence, with larger magnitudes implying a stronger impact.

2.3.3. Autoregressive Moving Average (ARIMA) Models

ARIMA models find extensive application in the field of time series analysis, effectively capturing the autoregressive and moving average components inherent in the data. Additionally, they incorporate an integration component to handle non-stationary data, where the statistical properties change over time. By considering these essential components, time series models enable the accurate representation and analysis of temporal patterns and dependencies in various domains.

An ARIMA model is characterized by three components: autoregressive (AR), moving average (MA), and integration (I). The AR component captures the linear relationship between a variable and its past values, the MA component models the influence of the past error terms, and the I component handles non-stationarity by differencing the data.

ARIMA model can be expressed as:

$$\begin{aligned}\Delta^d Y_{(t)} = c + \phi(1)\Delta^d Y_{(t-1)} + \phi(2)\Delta^d Y_{(t-2)} + ... + \phi(p)\Delta^d Y_{(t-p)} \\ +\theta(1)\epsilon_{(t-1)} + \theta(2)\epsilon_{(t-2)} + ... + \theta(q)\epsilon_{(t-q)}.\end{aligned} \tag{4}$$

Here, $\Delta^d Y_{(t)}$ represents the differenced variable at time t, c is a constant term, $\phi(1), \phi(2), ..., \phi(p)$ are the autoregressive coefficients, $Delta^d Y_{(t-1)}, \Delta^d Y_{(t-2)}, ..., \Delta^d Y_{(t-p)}$ correspond to the differenced lagged values, $\theta(1), \theta(2), ..., \theta(q)$ are the moving average coefficients, and $\epsilon_{(t-1)}, \epsilon_{(t-2)}, ..., \epsilon_{(t-q)}$ denote the error terms.

The integration component of ARIMA models, denoted by the differencing operator Δ^d, is used to transform the non-stationary time series into a stationary one. The parameter d represents the order of differencing required to achieve stationarity. Differencing involves computing the difference between consecutive observations to remove trends and seasonality from the data.

Parameter estimation techniques, such as maximum likelihood estimation (MLE) or least squares, are commonly used to find the values that optimize the likelihood of the observed data or minimize the difference between the observed values and the values predicted by the ARIMA model. On the other hand, selecting the appropriate orders for the AR, MA, and differencing components of an ARIMA model is crucial for accurate modeling. Statistical measures such as the Akaike Information Criterion (AIC) or the Bayesian Information Criterion (BIC) can be used to assess the goodness-of-fit of different models while considering their complexity. Additionally, graphical tools such as the autocorrelation function (ACF) and partial autocorrelation function (PACF) can provide insights into the presence of autocorrelation and guide the selection of optimal orders.

2.4. Data Preprocessing

Text preprocessing plays a vital role in NLP and significantly impacts the efficacy of algorithms. Prior to conducting text data preparation, several important considerations must be addressed. These include decisions on removing punctuation marks and stop-words, segmenting the text into sentences, words, or even individual letters, determining whether to use words in their original form or convert them to their root, and establishing strategies for handling misspelled words.

Accurate tokenization, for instance, enhances the precision of part-of-speech (POS) tagging, while preserving multi-word expressions can improve reasoning and machine translation. Preprocessing also plays a pivotal role in reducing dimensionality when working with extensive corpora. However, it is crucial to exercise caution during preprocessing to avoid excessive loss of information.

2.4.1. Stop-word Removal

Stop words are available in abundance in any human language. These words are often considered a single collection of words, which may indicate diverse matters to various applications. However, considering all candidate words from determiners (e.g., the, a) to prepositions (e.g., above, across) to some adjectives (e.g., good, excellent) can be an appropriate stop word list in some cases. Removing stop words discards low-level information from the text to give more focus to the relevant elements. Besides that, it reduces the data set size and thus cuts down the training time due to the fewer tokens involved in the process. Nevertheless, one should be careful while creating a stop word list and assuredly consider domain-specific wording.

Moreover, the removal of stop words may not be a good decision at all in some specific scenarios. That is, eliminating some words can be troublesome since these words might be decisive in contextualizing the intention. For instance, ignoring adjectives like ”upside” or ”ongoing” as well as negations like ”not” may cause algorithms to malfunction in polarity research as it changes the valence of the passage and omits the context. Thus, stop words are retained when using a contextual model like BERT.

2.4.2. Normalization and Tokenization

To reduce inflectional forms and occasionally derivatively related forms of a word to a single base form is the aim of the normalization phase of preprocessing. Lemmatization and stemming are the two most-common normalization techniques applied in NLP to reduce data irregularity and bring it closer to a predefined standard. Both stemming and lemmatization algorithms share the objective of reducing words, including inflectional forms and sometimes derivationally related forms, to a common base form. However, these two techniques differ in their approach and characteristics. The straightforwardness and lesser computational complexity of stemming prompted us to use it in this study. Therefore, following cleaning redundant words from the corpora, we put the remaining part of the documents into the stemming process for normalization. We preferred Porter stemmer so as not to prune too tightly and, thus, avoid any possible costly data loss[2].

2.4.3. Word Embeddings

To adequately manipulate and analyze text data, it is crucial to represent the corpus in a mathematically manipulable format. This entails converting words into numerical representations that can be understood by computers, as computers primarily operate using numerical data. However, the task of capturing semantic and syntactic relationships extends beyond a simple mapping of words to numbers within a sentence or document. A comprehensive representation that encompasses the semantic and syntactic properties of the corpus requires sizable numerical encoding.

Various word embedding methods, such as One Hot Encoding, Bag-of-Words, TF-IDF, Word2Vec, and FastText, have been developed to address this need. These techniques aim to capture the context of paragraphs or preceding sentences, along with extracting their semantic and syntactic properties and similarities. Among these methods, we opted for the Bag-of-Words technique in this study due to its simplicity, flexibility, and scalability in representing the text data within the documents.

It is important to note that text preprocessing procedures were exclusively conducted for the exploratory analysis segment of this study. This happens because FinBERT models involve contextual word representations, signifying that the input depiction of a particular token is formed by aggregating the corresponding token, segment, and position features. Consequently, the input data for FinBERT models does not necessitate the same preprocessing effort as other methods.

[2]Lancaster algorithm, another well-known stemmer, is one of the most aggressive stem-ming methods. Approximately, it has two times more stemming rules than the Porter method and tends to over stem a lot of words.

2.4.4. Transformations for Stationarity

Stationarity pertains to the constancy of statistical properties in a time series over time. Specifically, a stationary time series maintains consistent mean, variance, and autocorrelation characteristics. The presence of trends, seasonality, or other patterns can hinder accurate modeling and forecasting of the data.

Non-stationarity poses a significant challenge in time series analysis, as many forecasting methods assume stationarity. When a time series is non-stationary, its statistical properties change over time, making it difficult to apply traditional time series models and obtain reliable forecasts. Moreover, non-stationary time series can lead to spurious correlations, where variables appear to be correlated due to common trends or non-stationary patterns.

Polarity or semantic analysis, commonly known as sentiment analysis, is a natural language processing (NLP) technique that involves determining the sentiment or emotional tone expressed in a piece of text, such as a sentence, paragraph, or document. It involves automatically identifying and categorizing the sentiment conveyed in the text as positive, negative, or neutral. Polarity analysis aims to understand the subjective opinion, emotion, or sentiment expressed in text data, which can be valuable for various applications such as social media monitoring, customer feedback analysis, brand perception analysis, and market research. For the machine-learning approach, polarity analysis is typically performed using algorithms that are trained on labeled data, where the sentiment of the text data is manually annotated, allowing the model to learn patterns and associations between words, phrases, and sentiment labels.

On the other hand, forward-lookingness, also known as future or forward-looking information, refers to the aspect of information or data that pertains to the future or predicts future events, trends, or outcomes. It mostly engages in analyzing and interpreting data that provide insights or forecasts about what may happen in the future. Forward-lookingness is often used in various fields, such as economics, finance, business, and technology, to make informed decisions, develop strategies, and anticipate potential risks and opportunities.

Forward-lookingness can encompass a wide range of data types, including economic indicators, market trends, predictive analytics, scenario modeling, and expert opinions. It allows stakeholders to proactively plan and adapt to changing circumstances, make strategic decisions, and stay ahead in a dynamic and uncertain environment.

The following subsections introduce the tools we exploited and the route pursued throughout the analysis.

2.5. Bidirectional Encoder Representations from Transformers (BERT)

The development of deep learning models for sentiment analysis has undergone a significant transition from the early use of LSTM-based models to the more recent emergence of BERT models. LSTM models were initially used for sentiment analysis and were successful in capturing temporal dependencies in sequential data. However, the emergence of transformer-based models, such as BERT, enabled more effective handling of contextual information by exploiting self-attention mechanisms. BERT's success in capturing rich contextual information from pre-trained language models led to the development of BERT-based sentiment analysis models, which have consistently outperformed LSTM-based models in recent years.

Bidirectional Encoder Representations from Transformers (BERT) is a state-of-the-art natural language processing (NLP) model that has revolutionized the field of language understanding and representation [20]. BERT is a pre-trained deep learning model that is capable of capturing bidirectional contextualized representations of words and sentences, leading to significant advancements in various NLP tasks such as text classification, named entity recognition, question answering, and sentiment analysis.

At the heart of BERT lies the Transformer architecture, a neural network that enables efficient and parallel processing of input sequences. Unlike traditional NLP models that process words or sentences sequentially, BERT is designed to process input text in parallel, allowing for faster and more effective representations. During pre-training, BERT employs a masked language modeling (MLM) objective, where it randomly masks out words in a sentence. The model is trained to predict the masked words based on the contextual information from the surrounding words. This aspect of BERT enables it to learn bidirectional contextual representations, as it must rely on both the left and right context to predict the masked words accurately. In contrast to traditional word embeddings that are static and do not change based on the context, bidirectional nature allows BERT to capture the meaning of a word within the context of the entire sentence, enabling it to grasp the intricate nuances of language, such as word sense disambiguation, syntactic structure, and semantic relationships. By leveraging a masked language model (MLM) training approach, BERT can predict missing words in a sentence, forcing it to understand the dependencies between words and the context in which they appear.

BERT's input representation is a combination of multiple embeddings to capture different types of information from the input text. This combination ensures that BERT not only understands the words themselves but also their context and position in a sentence, as well as which sentence they belong to when dealing with pair-input tasks. Figure 2.5 describes the function of each of the embedding layers in BERT.

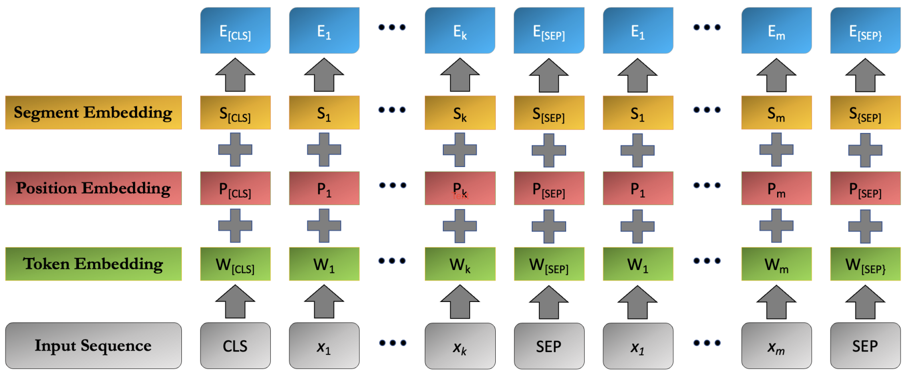


Figure 2.5 BERT input representation. The input embeddings is the sum of the token embeddings, segmentation embeddings, and the position embeddings.

BERT has been widely adopted in the NLP community and has achieved state-of-the-art results on a wide range of benchmark data sets. It has also been used as a base model for fine-tuning specific NLP tasks. Researchers and practitioners can fine-tune BERT on task-specific data with minimal modifications to the model architecture. This transfer learning capability of BERT has significantly reduced the need for large amounts of task-specific labeled data, making it a powerful tool for many real-world NLP applications with limited data availability.

BERT has also inspired further research and advancements in the field of NLP, leading to the development of variants and extensions of the original BERT model, such as ALBERT, ELECTRA, and RoBERTa. These variants have further pushed the boundaries of NLP performance and opened up new research avenues in areas such as multilingual NLP, low-resource NLP, and domain adaptation.

There are BERT adaptations for the finance and economics domains as well. Two notable examples are FinBERT and FinBERT FLS [17][18]. These models are designed to specifically cater to the unique characteristics and requirements of financial text data, making them particularly well-suited for tasks related to financial sentiment analysis, financial event prediction, and financial risk assessment.

2.5.1. FinBERT

FinBERT is a domain-specific adaptation of BERT that has been trained on a large corpus of financial documents, such as financial reports, earnings calls, and news articles related to finance and economics. By pre-training on such domain-specific data, FinBERT can capture the nuances and domain-specific language used in financial texts, which can differ from general-purpose language. Hence, it can perceive financial jargon, industry-specific terminology, and domain-specific contextual information, which is critical for accurately analyzing financial text data.

FinBERT FLS (forward-looking sentence), on the other hand, is a variant of FinBERT that is fine-tuned to detect the forward-looking structure of the sentences in financial documents. These models can identify phrases or sentences that indicate future events, predictions, or expectations, which can be crucial for decision-making in finance. For instance, FinBERT can recognize phrases that may contain a projection into the future and identify forward-looking information in financial text data. This enables analysts to understand better the prospects, risks, and opportunities associated with financial entities or events and make more informed decisions.

FinBERT and FinBERT FLS have proven to be highly effective in various financial text analysis tasks. They have been used in applications such as sentiment analysis of financial news, predicting stock price movements based on financial sentiment, and identifying financial risks and opportunities. These domain-specific BERT mod-els have demonstrated superior performance compared to general-purpose BERT mod-els in financial text analysis tasks, showcasing the importance of domain-specific adaptations for specialized domains in economics and finance.

3. RELATED WORK

A large part of NLP research that has been done in recent years focused on analyzing the direction and severity of users' thoughts and emotions on a particular topic on social media data. Though several superimposed mechanisms have been proposed to extend the functionality of general-purpose dictionaries for a given domain, these efforts still need the ability to correctly interpret the true meanings of the words in highly technical documents. One of the main reasons for this is that the standard gold dictionaries used in lexicon-based methods cannot produce successful results in do-main-specific texts.

Lucca et al. proposed a method to measure the polarity of the content of central bank disclosure documents. They exclusively employed Google's and Factiva's semantic orientation scoring systems to extract quantified meaning at both sentence and document levels. They conducted an empirical analysis of the statements released by FOMC after its policy meetings [21]. Using vector autoregression (VAR) models, they found that it takes more than a year for a significant change in the policy documents to be accompanied by a policy interest rate change. Furthermore, according to the results of their analysis, they claimed that short-term nominal treasury yields are more sensitive to changes in policy rates around policy announcements, while long-term treasuries spy on the changes in policy communication.

Szyszko et al. scrutinized six European economies and investigated the drifts of central banks' tone on their consumer inflation expectations between 2010 and 2019 [22]. After deriving the relevant content from the releases leveraging Latent Dirichlet Allocation (LDA) technique, they took advantage of a dictionary launched by Loughran and McDonald (LM) to quantify consumer review surveys since they have the opinion that consumers' knowledge is relatively low. Thus, any lexicon well-tailored for the policy texts would be too specific, and the LM dictionary was a good option for that purpose [23].

Tumala et al. employed an opinion-mining technique to analyze the efficiency of the Central Bank of Nigeria's communication hand-outs by building a specific lexicon using their previous meeting communiqués for the last fifteen years [24]. Besides readability and word occurrence frequency check, they use topic modeling for a detailed assessment of the subject density evolution of the documents throughout the years.

Kahveci et al. ran an investigation on a set of aspects as certainty, and opti-mism, of press releases of three specific central banks, namely FED, European Central Bank (ECB), and Central Bank of the Republic of Turkey (CBRT). They exploited Diction 7[3] to appraise the changes in communication transparency [25]. They claimed that while there is no significant shift in the tone of ECB and CBRT, an upward trend in certainty and a decline in the optimism of FED declarations were remarkable.

Park et al. quantified the Bank of Korea's meeting minutes to capture partial consequences of the central bank's manner for the national macroeconomic variables [26]. To this end, they applied contiguous sequences of the word (n-gram) approach to a field-specific Korean dictionary (eKoNLPy) since some words might mislead the results if they leave alone, e.g., recovery conveys an applauding meaning while sluggish recovery does not. They also built a

[3] A computer-assisted program for determining the tone of a verbal message using pre-defined dictionaries to process a passage and then it compares the results to built-in norms.

machine learning-based model leveraging Naïve Bayes Classifier and replicated the analysis for several well-known lexicons for comparison. Their study strongly suggests that the machine learning approach outperforms other dictionary-based models to explain current policy changes and provide insight into future movements.

Moniz et al. presented another framework that gauges the influence of central bank communications on investors' interest rate expectations in the United Kingdom [27]. Using an automated summarization algorithm, TextRank, to detect word-level communities, they adopted the LDA (Latent Dirichlet allocation) technique with a naïve classifier to infer topic clusters from the documents for verification purposes. Finally, they mined the sentiment score with the help of the General Inquirer dictionary, a set of procedures for identifying recurrent patterns within text documents built by the Massachusetts Institute of Technology [28]. Giuseppe Bruno from the Bank of Italy constructed a setup similar to Moniz's [29]. He took a close look at Bank of Italy's previous twenty Governor's Concluding Remarks documents from 1996 to 2015 and analyzed the evolution of the documents for different aspects, such as formality, memorability, polarity, etc., over time. He concluded that the sentiment of these documents stayed relatively neutral during the period, whereas with different short-term economic conditions, it occasionally showed slight volatility.

Sohangir et al. assessed the performance of three famed lexicons in the opinion extraction field using StockTwits, a financial-social network data set [30]. They included a bunch of machine learning algorithms in the scope of their study. Their experiment showed that TextBlob[4] intends to classify too many words as neutral compared to other lexicons and machine learning models, seemingly making it inappropriate for texts that contain technical jargon. They also found that only the VADER dictionary[5] outperforms baseline machine learning algorithms in terms of both accuracy and computational complexity.

[4]TextBlob is a popular open-source Python library for NLP that provides a simple interface for common NLP tasks such as part-of-speech tagging, noun phrase extraction, sentiment analysis, and translation, among others.

[5]VADER (Valence Aware Dictionary and sEntiment Reasoner) is a pre-trained rule-based sentiment analysis tool developed by Hutto and Gilbert in 2014 [31].

Devlin et al. introduced a new language representation model called BERT, which stands for Bidirectional Encoder Representations from Transformers [20]. They designed BERT to pre-train deep bidirectional representations from an unlabeled text by joint conditioning on all layers' left and right contexts. Then, they inserted one additional output layer to create a model that struggles with a wide range of tasks, such as question answering and language inference, without substantial task-specific architecture modifications. They demonstrated the importance of bidirectional pre-training for language representations and its capability that allows the same pre-trained model to tackle a broad set of NLP tasks successfully.

Howard et al. claimed another pre-training of a language model on a target do-main corpus enhances classification performance [32]. They conducted further pre-training processes in pursuit of discovering whether such adaptation would be advantageous for the realm of finance and economics. To do this, they severally conducted two experiments exploiting TRC2-financial corpus, a subset of Reuters' TRC2 corpora, and Financial PhraseBank from the study of Malo et al. [33]. In the last step, 16 professionals with backgrounds in business and finance annotated these sentences. Although Financial Phrasebank is relatively small, they concluded that using data from the direct target provided better target domain adaptation.

Araci et al. employed a transfer learning path and introduced a downstream language model based on BERT, called FinBERT, to deal with financial domain-specific texts for NLP tasks [34].

Yang et al. built another BERT model addressing domain-specific precision issues in NLP [35]. They gathered vast financial domain corpora comprising the three most representatives in finance and business communications, which amounts to over 4.9 billion tokens. Based on their initial pre-training, they also developed a model that classifies a sentence according to its forward-looking tendency. Although it varies according to the models' fine-tuning parameters and data set, their experiment results show a substantial improvement (between 4.3% and 29.2%) of FinBERTs over the generic BERT models.

This study will contribute to the literature since it is the first study that questions the effects of the recent pandemic on central banking communication by exploiting the state-of-the-art BERT models, off-domain tools, and further statistical analysis.

4. METHODOLOGY

This thesis aims to explore the potential effects of the Covid-19 crisis on the communication tone employed by the Federal Reserve (FED), while also examining the dynamics of the relationship between FED statements and consumer expectations before and after the crisis. The primary objective of this research is to conduct a thorough analysis of how the Covid-19 crisis might have impacted the way that the FED communicates and its efficacy in shaping consumer expectations across different time periods. By doing so, this research endeavors to address the following key inquiries:

- How have FED statements evolved regarding readability level, Zipfian compatibility, and topic distribution over time? Is there any significant change in the trend during the pandemic?
- What has occurred in the last decade years in the landscape of the co-movement between the semantic orientation and forward-lookingness intensity of FED statements and consumer inflation expectations? Has there been a change in the appearance of this association with the Covid-19 period?
- Has the pandemic significantly impacted the polarity orientation and forward-looking strength of the communiqués?

To answer these questions, we followed the workflow given in Figure 4.1.

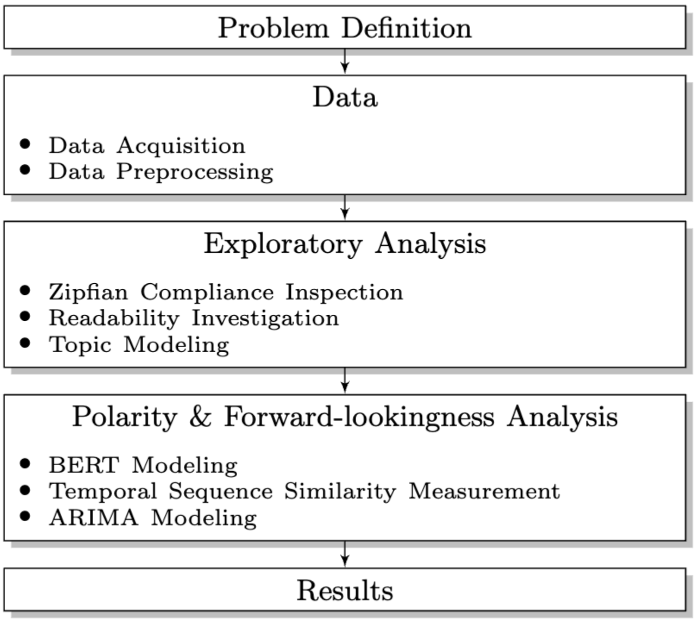


Figure 4.1 Research Workflow

4.1. Data

4.1.1. Data Acquisition

We acquired the sampling data from Federal Reserve Board's official page using web-scrapping techniques and then placed each meeting release in a separate text document. That way, we obtained a corpus comprising 214 FOMC meeting minutes re-leases from February 1992 to November 2022[6]. Additionally, we extracted monthly consumer inflation expectation data for the one- and three-year ahead[7] time horizon from the Federal Reserve Bank of New York. Below is the snapshot from the Committee Policy Action section of the FOMC meeting on June 14-15, 2022:

"In their discussion of monetary policy for this meeting, members agreed that overall economic activity appeared to have picked up after edging down in the first quarter. Job gains had been robust in recent months, and the unemployment rate had remained low. Members

[6] https://www.federalreserve.gov/monetarypolicy/materials/
[7] https://www.newyorkfed.org/microeconomics/sce#/

also agreed that inflation remained elevated, reflecting supply and demand imbalances related to the pandemic, higher energy prices, and broader price pressures."

"Members concurred that the invasion of Ukraine by Russia was causing tremendous human and economic hardship. Members agreed that the invasion and related events were creating additional upward pressure on inflation and were weighing on global economic activity. With the effects of the invasion of Ukraine by Russia already materializing, members considered it appropriate to omit from the June statement the sentence conveying the high uncertainty associated with the implications of the invasion for the U.S. economy. Members also agreed that COVID-related lockdowns in China were likely to exacerbate supply chain disruptions. In light of these developments, members remarked that they remain highly attentive to the upside risks to inflation and would be nimble in responding to incoming data and the evolving outlook."

Federal Open Market Committee – Minutes of the Meeting of June 14–15, 2022

In the course of our investigation, we embarked on several preprocessing measures. These encompassed the elimination of stop-words, standardization of data, segmenting the text into meaningful units (tokenization), and converting these tokens into a format suitable for computation (vector representation). Furthermore, to suit our ARIMA model, we introduced a stationary transformation to the data. These preparatory measures were crucial in refining our data and preparing it for the in-depth analysis that followed.

For the purpose of examining the data's stationarity, we employed two widely-recognized tests: the Augmented Dickey-Fuller (ADF) and the Kwiatkowski–Phillips–Schmidt–Shin (KPSS) tests, evaluating them at different levels. Nonetheless, it is imperative to recognize that these tests are not without their challenges. Each has its constraints in pinpointing different forms of non-stationarity and operates within specific assumptions. To elucidate, the ADF test operates under the presumption of residuals being normally distributed and the accurate specification of the autoregressive (AR) process. On the other hand, the KPSS test functions on the basis of having finite variance and errors that are not serially correlated. Given these intricacies, our approach involved a holistic review of the test outcomes, taking into account the inherent advantages and limitations of each. Detailed outcomes from these tests can be referred to in Table 4.1.

A careful scrutiny of the results revealed that for the Cov-19 indicator, there was not substantial evidence in the ADF test to dismiss the null hypothesis (H_0). This insinuates the potential stationarity of the data. Conversely, for the Cov-19 series, the hypothesis of stationarity was dismissed across all significance thresholds. These collective findings point towards the existence of a unit root, indicating a propensity for the series to follow a random walk pattern without reverting to the mean. As a logical deduction, it becomes evident that the Cov-19 series exhibits non-stationary characteristics.

Subsequently, we applied the first difference to all the series in consideration. The outcome of this process led to the rejection of the null hypothesis (H_0) in the ADF test. Yet, in the KPSS test at a significance level of 1%, the null hypothesis was upheld. These insights were instrumental for the modeling steps that followed.

Table 4.1 Unit-root and Stationarity Test Results

	ADF		KPSS	
Variable	test stat.	*p*-value	test stat.	*p*-value
Specific FLS	-3.622	0.032**	0.126	0.087*
non-Specific FLS	-4.118	<0.01***	0.119	0.099*
Positivity	-4.235	<0.01***	0.091	0.100
Negativity	-3.235	0.083*	0.300	<0.01***
Neutrality	-3.643	0.030**	0.154	0.043**
Cov-19	-1.349	0.849	0.651	<0.01***

Note: The null hypothesis of the ADF test (H_0) argues the presence of unit roots, while the alternative hypothesis (H_A) asserts the absence of unit roots. Contrarily, a claim on the stationarity of the series frames the (H_0) hypothesis of the KPSS test and vice versa. A *p*-value less than the significance level provides evidence to reject the null hypothesis at that significance level. Conversely, values larger than the significance level imply insufficient evidence to reject the null.
($***$: Significant at 99 percent confidence interval, $**$: Significant at 95 percent confidence interval, $*$: Significant at 90 percent confidence interval.)

4.2. Exploratory Analysis

This part of the thesis delves into notable observations drawn from the press releases of FOMC minutes spanning the past thirty years of committee gatherings. Initially, our focus was directed at evaluating the quantity of paragraphs and words in these documents. The visual depiction of these patterns over time can be observed in Figure 4.2. Interestingly, there was a consistent volume in the committee meetings' documentation up until 2007. However, post-2007, there was a noticeable expansion in the length of these releases. The surge in document size became particularly pronounced after 2010.

Subsequent to this, we embarked on a comprehensive statistical review of these documents. This involved assessing them based on standard computational linguistic attributes, particularly word frequency. This was then complemented by an assessment of the legibility of the minutes, employing various readability evaluation metrics.

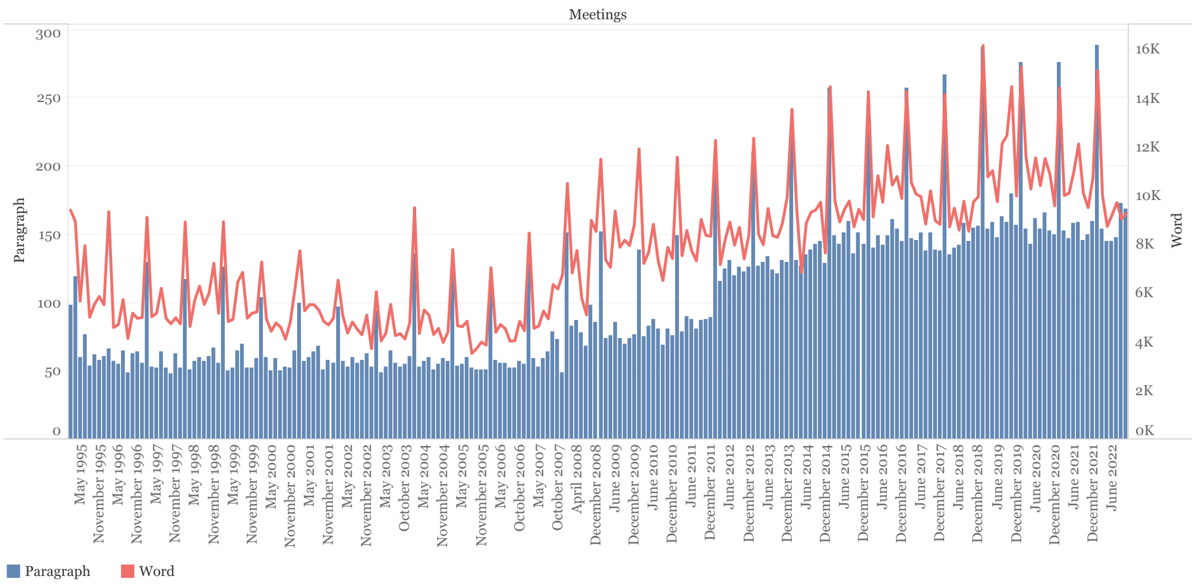


Figure 4.2 Word and Paragraph Volume

4.2.1. Zipfian Compliance

Zipf's Law theorizes that within a body of natural language texts, the occurrence of individual words is inversely linked to their standings in the frequency hierarchy. For example, if the most commonly occurring word appears 1,000 times, the second most frequent word is anticipated to appear about half as frequently, and the third word approximately a third as often, and so on. Furthermore, an associated principle from information theory proposes that a text, to optimize its information conveyance using a restricted vocabulary, should adhere to Zipf's Law [36].

Upon generating Zipf's tables for the consolidated corpus annually, we utilized the Kolmogorov-Smirnov test[8] to evaluate if the data originates from a population exhibiting Zipfian distribution characteristics. The test results did not reject the null hypothesis, indicating that the corpus exhibits a Zipfian nature, with a consistent significance threshold of 5% maintained throughout all years. There is no compelling evidence to assert that the sample data diverges from a Zipfian distribution. Consequently, our prevailing interpretation is that the minutes align with Zipfian principles, and the emergence of the Covid-19 pandemic has not introduced deviations negating the Zipfian essence of FOMC statements (refer to Table 4.2).

Table 4.2 Kolmogorov-Smirnov Test Results

	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
KS stat.	0.705	0.683	0.692	0.674	0.702	0.735	0.763	0.684	0.712	0.733
p-value	0.174	0.201	0.189	0.213	0.177	0.140	0.112	0.200	0.165	0.143

Note: (H_0): The data follow the Zipfian distribution. (H_A): The data do not follow the Zipfian distribution. A p-value less than the significance level provides evidence to reject the null hypothesis at that significance level. Conversely, values larger than the significance level imply insufficient evidence to reject the null.
($***$: Significant at 99 percent confidence interval, $**$: Significant at 95 percent confidence interval, $*$: Significant at 90 percent confidence interval.)

[8] This is a non-parametric method used to determine the congruence of continuous or discontinuous one-dimensional probability distributions, either by comparing a sample with a benchmark probability distribution or by juxtaposing two samples.

4.2.2. Readability Investigation

The concept of readability pertains to the ease with which a reader can comprehend a given text. Various scoring systems have been developed to quantify the potential difficulty of a text by considering specific attributes known to influence complexity, such as the average length of sentences or the prevalence of intricate words. It is important to differentiate between the readability and the intelligibility of a document. While readability assesses the structural and linguistic aspects of the text, intelligibility might vary based on the reader's existing knowledge on the subject and the intrinsic qualitative and quantitative features of the material. For our study, we utilized three prominent readability indices: the Flesch-Kincaid grade level, the Gunning-Fog Index, and the Automated Readability Index (ARI). A visual representation detailing the temporal variations in the readability of FOMC statements can be observed in Figure 4.3.

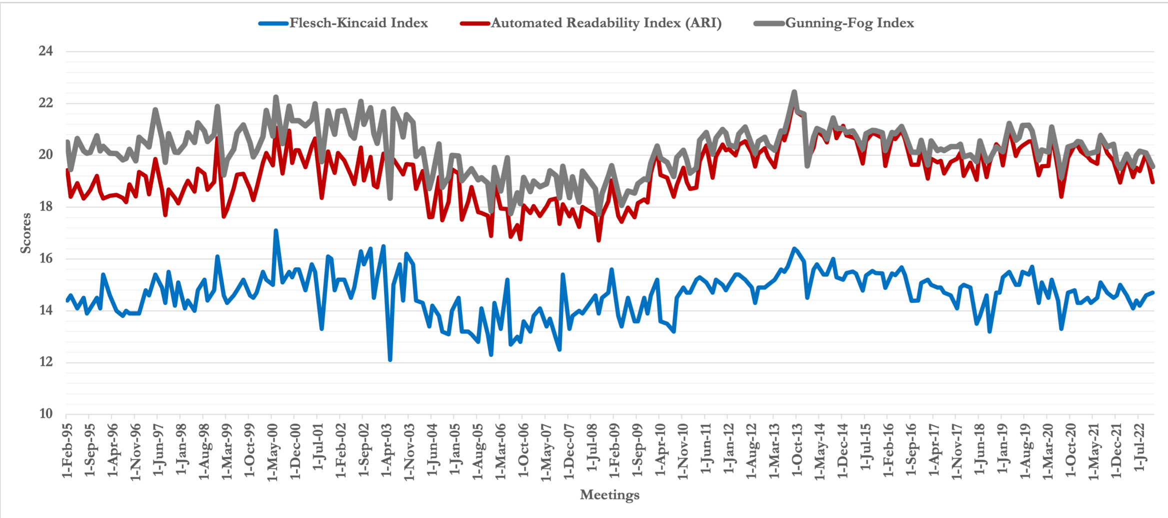


Figure 4.3 FOMC Minutes Releases' Readability Indices

The Flesch-Kincaid Grade Level is modeled on the US educational framework, implying the educational level necessary to grasp a particular text. For instance, a score of 13 suggests that the material is comprehensible to a college-level student. On the Gunning-Fog scale, scores between 17-20 and above 20 represent post-graduate and advanced post-graduate levels of understanding, respectively. Meanwhile, the Automated Readability Index (ARI) provides a figure that approximates the reader's age needed for clear comprehension. To contextualize, the 12th grade, which is the concluding year of US high school before college, aligns with a 17-year-old's reading capability.

It is noteworthy that post-2010, ARI and Gunning-Fog measurements displayed considerable congruence. However, during earlier times, they exhibited discrepancies, with Gunning-Fog consistently registering higher scores than the ARI. An intriguing observation is the diminishing readability levels of FED communications post-1990s, with a pronounced surge presumably linked to the financial turmoil of 2008. In recent years, all metrics have plateaued around the collegiate standard. Yet, both the Gunning-Fog and ARI metrics exhibit a subtle decline in their trends.

4.2.3. Topic Modeling

Topic modeling, often referred to as topic extraction, is a statistical technique primarily used to identify the primary themes within a set of documents. Typically falling under the domain of unsupervised machine learning, topic modeling aims to uncover hidden patterns and structures in textual data. That is, algorithms pinpoint themes based on discernible patterns, like word clusters and their recurrence. Conversely, topic classification falls under supervised learning. It leverages rule-driven systems, or learning algorithms, which are meticulously trained using hand-labeled data and predetermined categories. Only upon familiarizing with the training data can these machine learning algorithms competently categorize new, unseen texts in accordance with these predefined categories.

For the purposes of our research, we adopted the Latent Dirichlet Allocation (LDA) technique. This generative probabilistic model, proposed by Blei and colleagues [37], allowed us to explore implicit themes and ascertain their prominence within documents. Subsequently, we scrutinized the variations in these weights during the FOMC sessions over the years.

After subjecting the words to the previously discussed preprocessing phase, we then grouped them. Post clustering, we assigned labels to these groupings by considering the terms that appeared most frequently within each cluster. Figure 4.4 visually depicts the progression in topic intensity across various periods of the FED's FOMC minutes. Furthermore, Table 4.3 showcases the ten most recurrent words for each identified theme.

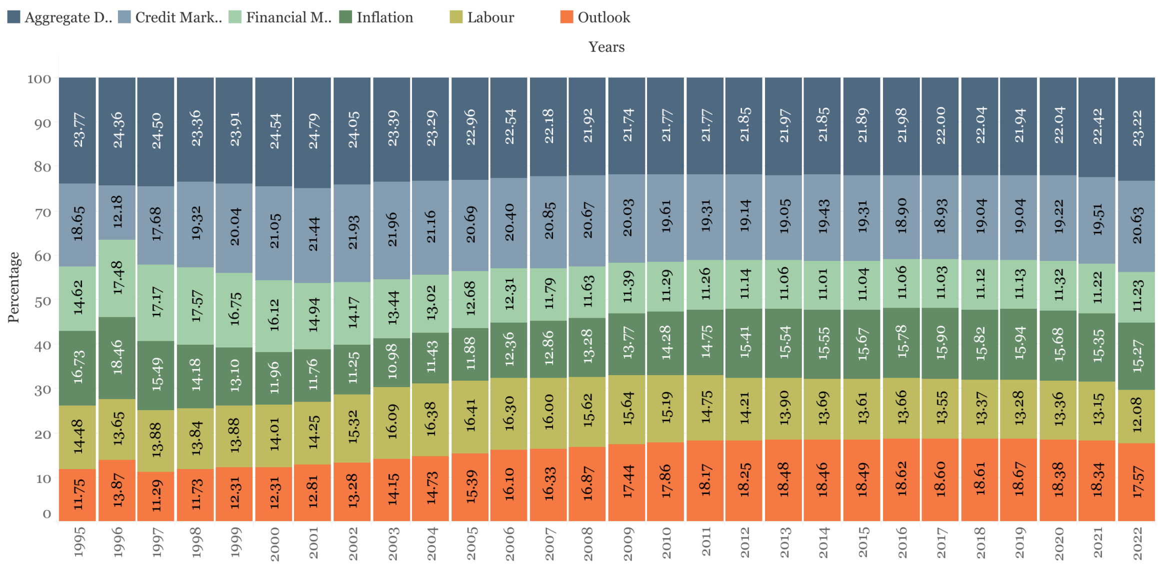


Figure 4.4 Relative Topic Intensity of FOMC minutes

The topic of "Aggregate Demand" has consistently dominated FOMC communiqués from 1995 to 2022. It stands out, claiming precedence over other themes by consistently maintaining a weight of more than 20 percent throughout the years under review. Following closely, the "Credit Market", which represents the arena where agents negotiate and trade debts, has often occupied the second most significant segment of the meeting discussions. Although it retained this position for most years, there was a notable exception in 1996. In the post-2000 era, this topic has consistently constituted approximately a fifth of the meeting's discourse.

Meanwhile, the theme of "Inflation" underwent some fluctuations. Starting with a share of 16.73 percent in 1995, its prominence dwindled to nearly 11 percent in the early 2000s. However, a resurgence was observed post-2005, stabilizing at about 15 percent in the latter years. It is noteworthy that the proportional representation of these subjects in Fed meetings hasn't seen any marked alterations, especially post-2010. The discussions found a specific trajectory post the global financial meltdown.

Lastly, discussions related to ”Outlook” have gradually increased since the onset of the 2000s. By 2010, it consistently accounted for an estimated 18 percent of the meeting’s agenda, a trend that has since been sustained.

Table 4.3 Word Ranking by Topics

Aggregate Demand	Credit Market	Financial Market	Inflation	Labour	Outlook
quarter	loan	market	inflat	particip	particip
increas	yield	fund	price	labor	polici
spend	remain	financi	expect	rate	inflat
product	credit	secur	year	busi	econom
real	period	rate	consum	continu	committe
busi	market	treasuri	measur	sector	would
decline	bank	period	staff	employ	rate
good	declin	asset	percent	market	risk
manufacture	bond	term	energi	recent	expect
sale	intermeet	feder	month	hous	feder

5. POLARITY AND FORWARD-LOOKINGNESS ANALYSIS

Sentiment analysis, also known as polarity or semantic analysis, is a branch of natural language processing (NLP) dedicated to deciphering the emotional tone or sentiment expressed within text, whether it’s a single sentence, a paragraph, or an entire document. Its core function is to automatically identify and classify this sentiment into categories like positive, negative, or neutral. This analytical approach aims to uncover the subjective opinions, emotional nuances, or overall sentiment encapsulated in text data, rendering it valuable for a wide range of applications, including social media tracking, customer feedback evaluation, brand perception assessment, and market research. For the machine-learning approach, polarity analysis is typically performed using algorithms that are trained on labeled data, where the sentiment of the text data is manually annotated, allowing the model to learn patterns and associations between words, phrases, and sentiment labels.

On the other hand, forward-lookingness, also known as future or forward-looking information, refers to the aspect of information or data that pertains to the future or predicts future events, trends, or outcomes. It mostly engages in analyzing and interpreting data that provide insights or forecasts about what may happen in the future. Forward-lookingness is often used in various fields, such as economics, finance, business, and technology, to make informed decisions, develop strategies, and anticipate potential risks and opportunities. Forward-lookingness can encompass a wide range of data types, including economic indicators, market trends, predictive analytics, scenario modeling, and expert opinions. It allows stakeholders to proactively plan and adapt to changing circumstances, make strategic decisions, and stay ahead in a dynamic and uncertain environment.

The following subsections introduce the specific BERT models and techniques we exploited and the route pursued throughout the analysis.

5.1. BERT Modeling and FinBERT

We extracted quantitative information from the documents using FinBERT and FinBERT FLS. FinBERT assigns labels to each sentence in the documents, classifying them as positive, negative, or neutral. At the same time, FinBERT FLS produces an outcome, categorizing sentences as either specific forward-looking-sentence, non-specific forward-looking-sentence, or not-forward-looking-sentence. Labeling a sentence as Specific FLS implies that it is a definitive judgment about the future of a specific entity. In contrast, non-specific FLS spells its future-oriented structure without referring to matters pointedly. As the name suggests, the not-FLS tags indicate that the sentence has no forward-looking nature. We calculated the average scores for each class in a meeting-wise manner as follows:

where total scores (i.e., Total Positive Score) are simply the sum of all scores produced at the sentence level for each class (i.e., Positive). Figures 5.1 and 5.2 illustrate the time series for each defined variable with a joined bar graph representing the word counts on the opposite axis.

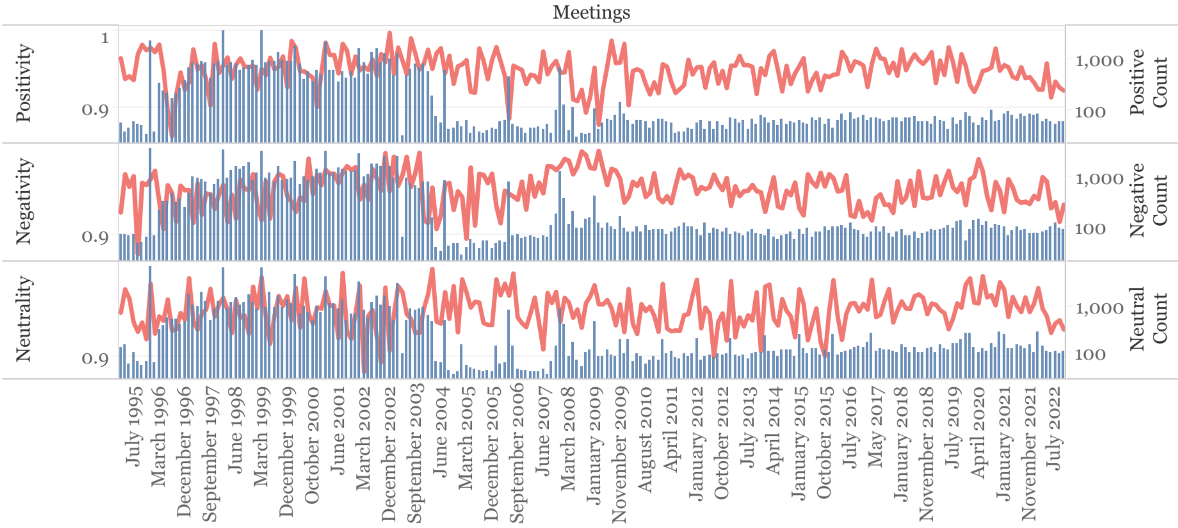


Figure 5.1 Scores (line graph on the left-axis) and Total Sentence Counts (bar graph on the right-axis) -for each class- Results for Polarity Search (e.g. Positive, Negative, and Neutral)

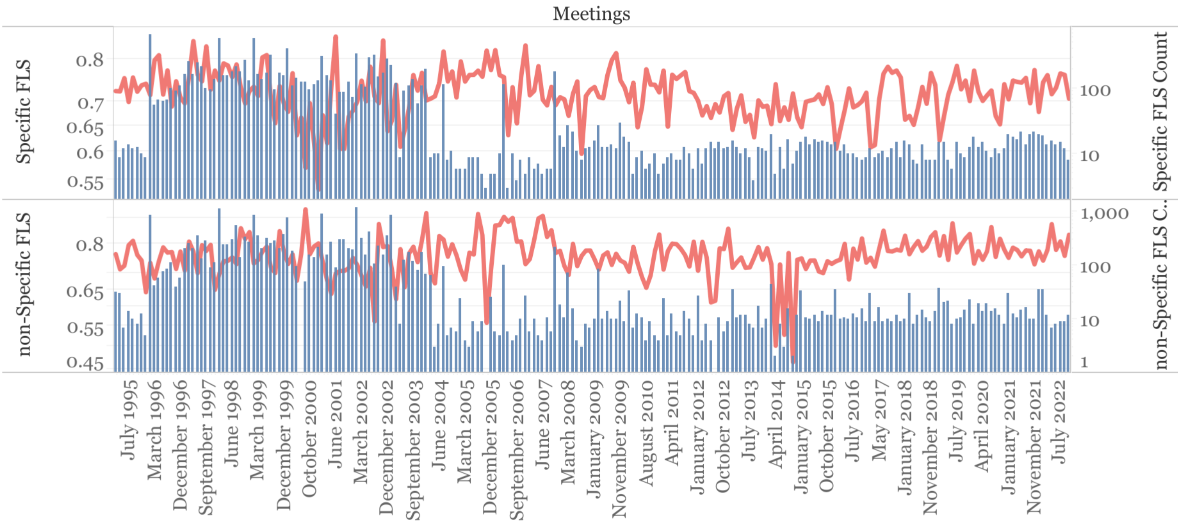


Figure 5.2 Scores (line graph on the left-axis) and Total Sentence Counts (bar graph on the right-axis) -for each class- Results for Polarity Search (e.g. specific FLS and non Specific FLS)

As Figures 5.1 and 5.2 show, there is a substantial decrease in the magnitude of the volume of the minutes after 2004. Remarkably, the decline in the count of neutral sentences appears to be more pronounced than positive and negative ones. This trend suggests that the Federal Reserve (FED) has displayed a discernible inclination towards es-chewing statements that lack a definitive direction in following the attitude change. However, towards the culmination of a period spanning approximately three years, there emerges a resurgence in the volume of neutral expressions. Subsequently, in the aftermath of the 2007-2008 financial crisis, it is discerned that the volume of sentences within minutes exhibits a state of relative stability.

Naturally, the shrinkage in the volume of the minutes goes along with the specific and non-specific forward-looking volume. Yet, the forward-looking strength of the FOMC releases fluctuates more than the polarity direction. Moreover, the harsh fluctuations experienced occasionally in forward-looking scoring are interesting. We see such fluctuations between '99 and '03 for specific forward-lookingness and also in 2005 and between '13 and '15 for non-specific forward-lookingness. However, after the adjustment period, as mentioned earlier, the forward-lookingness volume has kept its disheveled appearance than the polarity.

5.2. Temporal Sequence Similarity and Dynamic Time Warping (DTW)

A temporal sequence is a series of data points or observations ordered chronologically. It illustrates the progression of a variable or phenomenon across time, where every data point is linked to a particular time moment or time frame.

Measuring the similarity between two temporal sequences with different frequencies can be challenging due to differences in the granularity or resolution. A well-known approach is to resample or aggregate the data to a standard frequency before calculating similarity measures. Several other methods commonly employed for measuring the similarity between temporal sequences with different frequencies include Dynamic Time Warping (DTW), cross-correlation, Fourier-based methods, Pearson correlation, and wavelet-based methods.

Resampling or aggregating data to a common frequency for measuring similarity between temporal sequences with different frequencies may have limitations. It can lead to loss of information, potential misalignment or distortion of temporal patterns, and introduce biases.

Careful consideration should be given to the potential implications and limitations of this approach, such as the impact on the accuracy and validity of the similarity measures and the potential loss of fine-grained temporal dynamics. Alternative methods that account for the inherent differences in frequency between the sequences, such as dynamic time warping or wavelet-based methods, may be considered to mitigate these limitations. Therefore, we employed Dynamic Time Warping method, a popular technique for measuring similarity between two temporal sequences with different frequencies.

Using the DTW method, we attained the proximity between consumer inflation expectations and each series. Since the consumer inflation expectation behavior can vary depending on the time horizon, we exclusively considered the short- and medium-term expectations for our analysis. Figures 5.3 and 5.4 show the distances between consumer inflation expectations and each score series over the course of time.

Figures 5.3 and 5.4 show that the distance between the sequences shows extreme similarity for the one-year and three-year consumer inflation expectations for the pre-crisis period. We follow that the distance between the sequences has increased to 16 units for one-year expectations and 10 units for three-year expectations with the global epidemic. Indeed, it is not much striking since, as the literature says, consumer inflation expectations tend to be more stable over longer time horizons. Short-term expectations can be more volatile and subject to sudden changes in economic conditions or perceptions of economic policy. On the other hand, while the distance between three-year inflation expectations and the FOMC scoring series in the post-crisis period shows the normalization trend more clearly, we need to observe more data for one-year expectations.

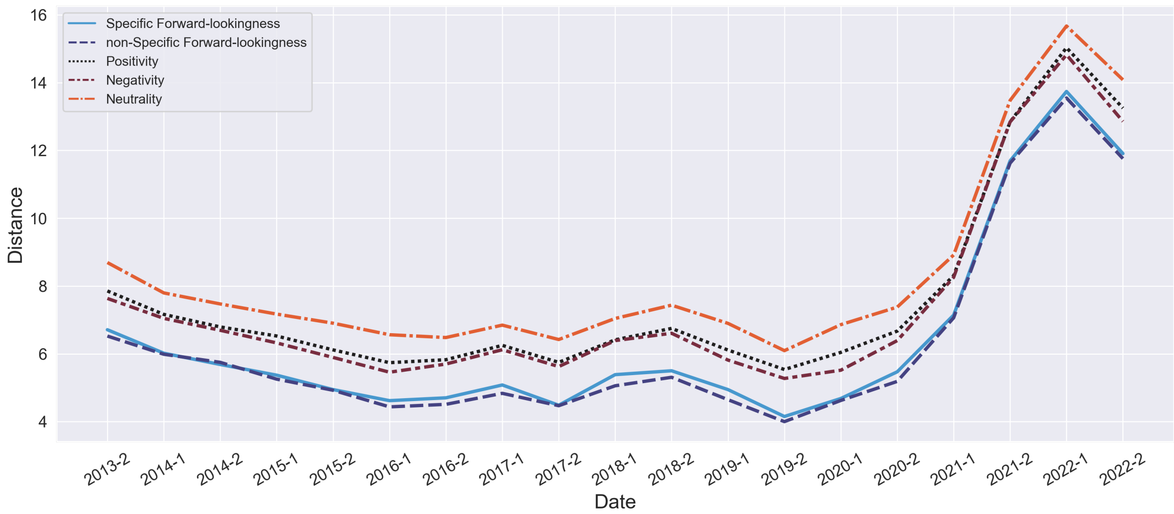


Figure 5.3 Distances for one-year ahead Consumer Inflation Expectations

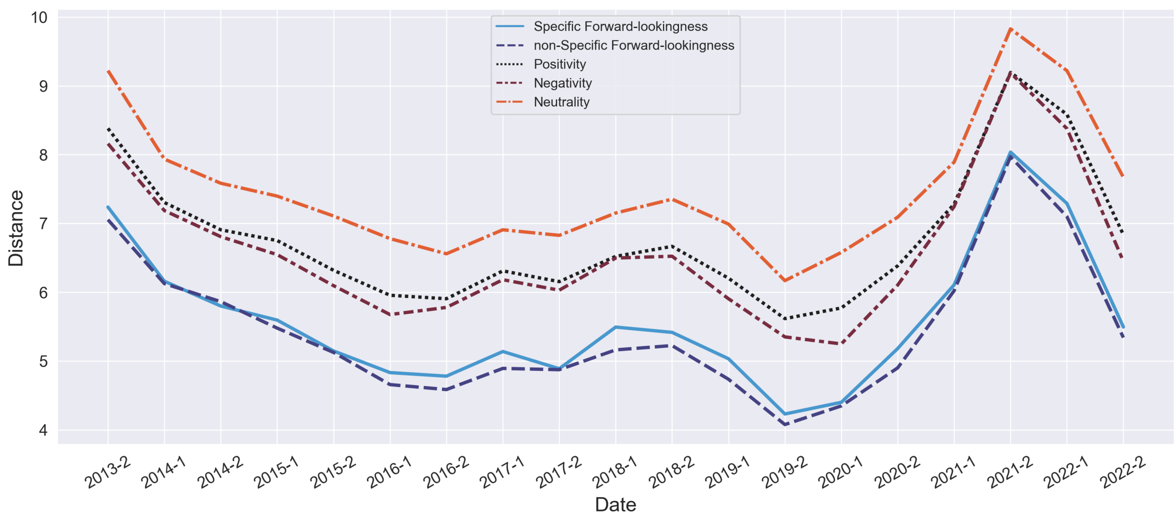
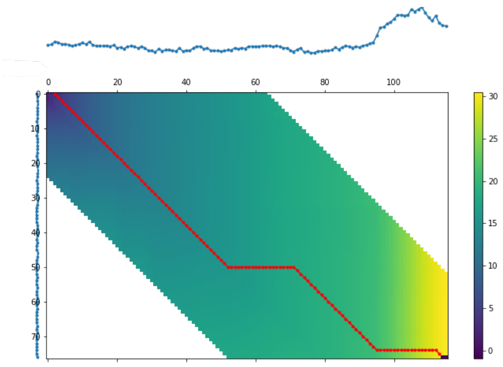
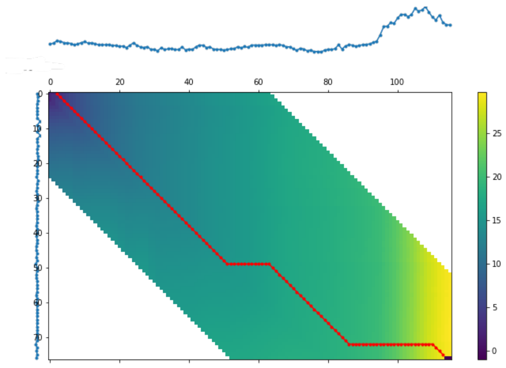


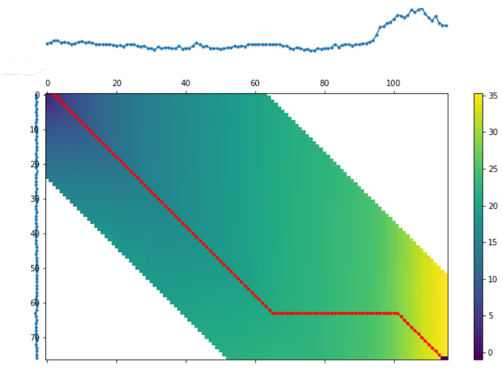
Figure 5.4 Distances for three-year ahead Consumer Inflation Expectations



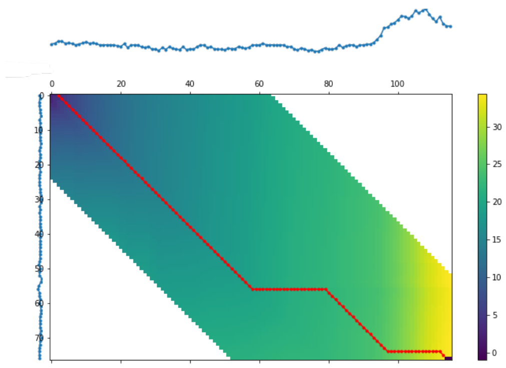
Specific FLS



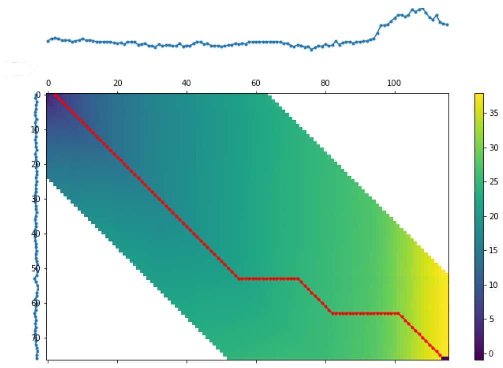
non-Specific FLS



Positivity

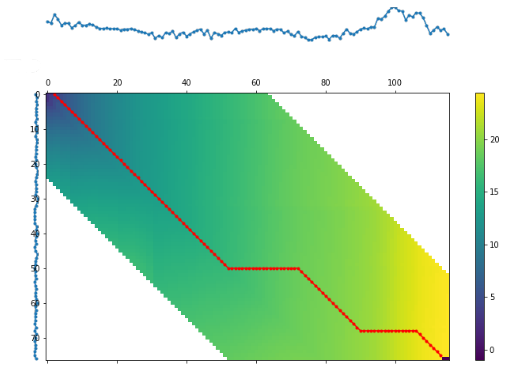


Negativity

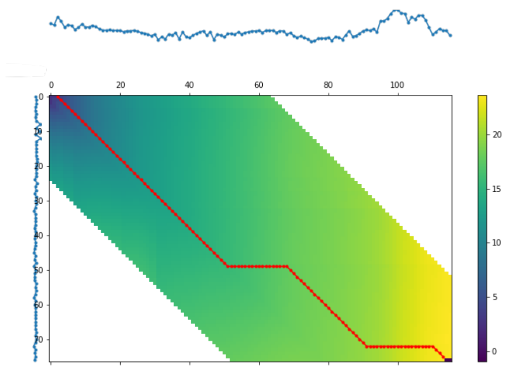


Neutrality

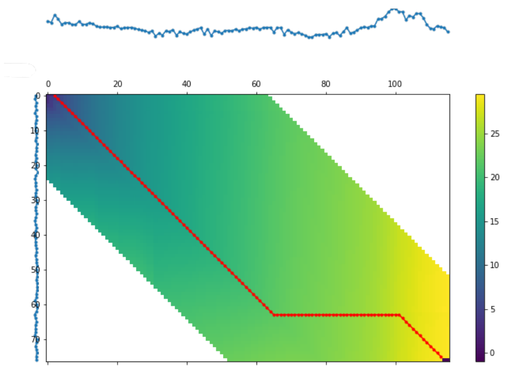
Figure 5.5 Accumulated Cost Matrices and Warping Paths for one-year ahead Consumer Inflation Expectations



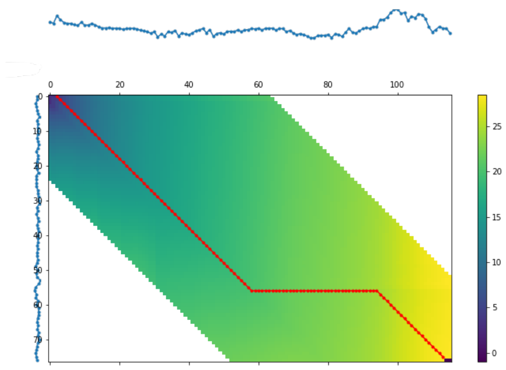
Specific FLS



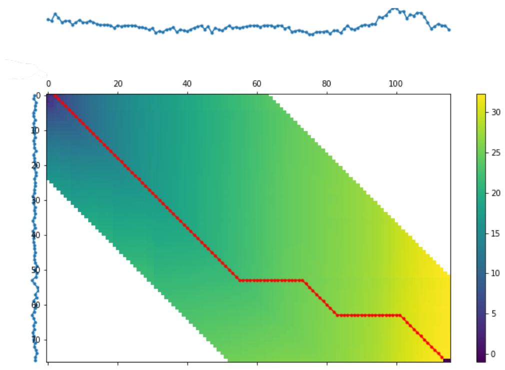
non-Specific FLS



Positivity



Negativity



Neutrality

Figure 5.6 Accumulated Cost Matrices and Warping Paths for three-year ahead Consumer Inflation Expectations

The accumulated cost matrix represents the cumulative cost of aligning the pairwise series of sequences at each point in time. The cost at each position is computed based on a distance metric between the elements at that position and a set of neighboring positions in the accumulated cost matrix. The objective is to identify the most favorable warping path within the matrix, characterized by the lowest cumulative cost. This path signifies the optimal alignment between the two sequences.

Additionally, the warping path signifies the ideal mapping or correspondence between the sequences, signifying that the data points in both sequences exhibit the highest degree of similarity. A shorter path or a lower accumulated cost indicates a higher similarity between the sequences, whereas a longer path or a higher accumulated cost indicates a lower similarity. The warping path can be interpreted as a sequence of pairs of indices from the two sequences that indicate which elements are matched or aligned. The warping paths can also provide information on the time lag or shift between the two sequences, indicating any temporal misalignment or time lag in the similarity patterns. Figures 5.7 and **??** visualize all possible warping paths for one- and three-year ahead produced by the DTW algorithm, respectively.

Note that a diagonal move is a match between the two sequences. In contrast, off-diagonal moves imply either duplication of one point of one sequence (expansion) or elimination of one of the points (contradiction). Observe that those pairwise warping paths for one- and three-year-ahead are almost identical except for the negativity series, whose three-year-ahead path flourishes a salient deviation from its one-year-ahead pattern. Thus, the patterns between the series obtained from the Fed minutes and the one- and three-year ahead consumer inflation expectations are alike. Additionally, all pair-wise series showed a phase shift once or twice during the last decade. The accumulated similarity matrix and paths produced by the DTW algorithm showed a non-monotonic pattern during these periods, suggesting that the scoring series (e.g., positivity) experienced a temporal distortion or phase shift compared to the other time series (e.g., consumer inflation expectation).

5.3. ARIMAX Modeling

ARIMAX (AutoRegressive Integrated Moving Average with eXogenous variables) is an extension of ARIMA that allows for the inclusion of exogenous variables, which are external factors that may impact the time series being forecasted. Exogenous variables are included as additional predictors in the ARIMA model, along with the autoregressive, moving average, and integrated components. It is particularly useful when there are external factors that are expected to influence the time series data and need to be accounted for in the forecasting process [[38]]. Below equation constructs ARIMA(p, I, q):

$$Y_t = c + (\phi_1 * Y_{t-1}) + ... + (\phi_p * Y_{t-p}) + \varepsilon_t - (\theta_1 * \varepsilon_{t-1}) - ... - (\theta_q * \varepsilon_{t-q}) \tag{5}$$

where:
y_t: The value of the time series variable at time t,
$\phi_1, ..., \phi_p$: The parameters of the AR component,
$\theta_1, ..., \theta_q$: The parameters of the MA component.

Let $\beta_1, ..., \beta_m$ be the parameters of the exogenous variables X_1, X_2,..., X_m. Then the general form of an ARIMAX(p, I, q) model can be expressed correspondingly [39]:

$$\begin{aligned} Y_t = c + (\phi_1 * Y_{t-1}) + ... + (\phi_p * Y_{t-p}) + (\beta_1 * X_{1,t}) + ... + (\beta_m * X_{m,t}) \\ + \varepsilon_t - (\theta_1 * \varepsilon_{t-1}) - ... - (\theta_q * \varepsilon_{t-q}) \end{aligned} \tag{6}$$

To determine the appropriate model order, we went through the autocorrelation function (ACF) and partial autocorrelation function (PACF) plots and identified the autoregressive (AR) and moving average (MA) components. We traveled through the possible model neighborhoods as well and evaluated the prospective models by their AIC (Akaike Information Criterion) values to find the best fit. And finally, we embedded a dummy variable into the equation as an indicator function. Tables 5.1 and 5.2 present the coefficients, standard errors, and p values. Moreover, Figure **??** presents the residuals which are the differences between the actual observed values and the predicted values of the time series after fitting the models. That is, unexplained variation between the observed values of a dependent variable and the values predicted by the model.

Based on the results in Tables 5.1 and 5.2, coefficients associated with the COVID-19 indicator variable are not statistically significant for any time series, as indicated by p-values that are larger than any predefined significance level (e.g., $\alpha = 0.05$). This implies that including the indicator function variable did not significantly contribute to the models' predictive performance for the respective time series. On the other hand, the residuals of the polarity models displayed more desirable characteristics than the forward-lookingness models, as they showed less prominent patterns and deviations from the expected behavior, suggesting a better fit to the data.

Table 5.1 Coefficient - Standard Error and p-value Tables for Polarity Modeling

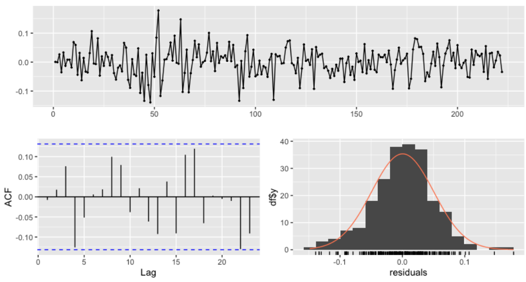
	Positivity			Negativity			Neutrality		
	coeff	S.E.	p-val	coeff	S.E.	p-val	coeff	S.E.	p-val
ar(1)	0.229	0.068	0.001***	0.862	0.071	$\approx 0***$	0.841	0.060	$\approx 0***$
ar(2)	0.238	0.069	0.001***	-	-	-	-	-	-
ma(1)	-0.988	0.025	$\approx 0***$	-1.524	0.0126	$\approx 0***$	-1.443	0.098	$\approx 0***$
ma(2)	-	-	-	0.524	0.125	$\approx 0***$	0.443	0.097	$\approx 0***$
Xreg	-0.017	0.023	0.472	-0.016	0.039	0.687	0.008	0.051	0.883

Note: Xreg: Covid-19 Indicator Function
($***$: Significant at 99 percent confidence interval, $**$: Significant at 95 percent confidence interval, $*$: Significant at 90 percent confidence interval.)

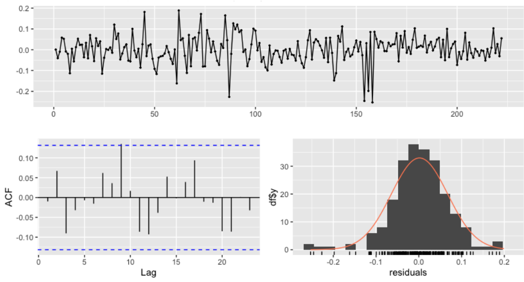
Table 5.2 Coefficient - Standard Error and p-value Tables for Forward-lookingness Modeling

	Specific			non-Specific		
	coeff	S.E.	p-val	coeff	S.E.	p-val
ar(1)	0.150	0.084	0.073	0.0238	0.066	$\approx 0***$
ma(1)	-0.870	0.049	$\approx 0***$	-1.000	0.018	$\approx 0***$
Xreg	-0.011	0.030	0.721	0.027	0.015	0.075*

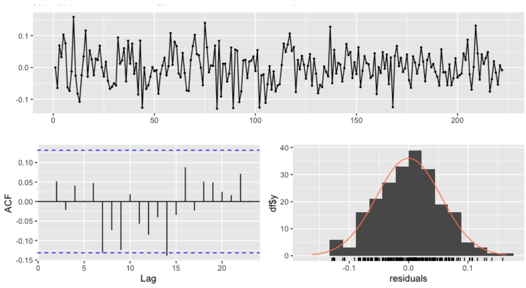
Note: Xreg: Covid-19 Indicator Function
($***$: Significant at 99 percent confidence interval, $**$: Significant at 95 percent confidence interval, $*$: Significant at 90 percent confidence interval.)



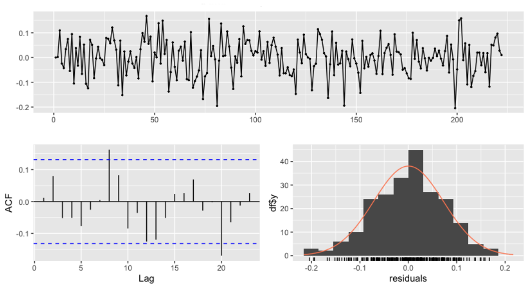
ARIMAX(1,1,1) on Specific FLS



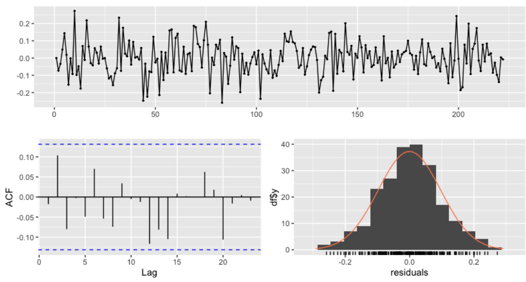
ARIMAX(1,1,1) on non-Specific FLS



ARIMAX(0,1,2) on Positivity



ARIMAX(0,1,2) on Negativity



ARIMAX(1,1,1) on Neutrality

Figure 5.7 ARIMAX Residuals

6. RESULTS

This section presents the results from our comprehensive examination of the Federal Reserve (FED) communication statements, both prior to and in the aftermath of the pandemic. The observations shed light on how FED communication evolved during these significant periods.

Firstly, there is an unmistakable rise in the volume of words and paragraphs, which remained fairly constant up to 2007. From mid-2007 onwards, an upward trajectory became evident, with a notable peak in 2011. An interesting observation is the recurrent trend of lengthier texts during the first meetings each year. This repeated pattern could stem from the FED's annual practice of conducting a thorough assessment of financial and economic landscapes, leading to the setting of monetary policy goals for the upcoming year. To articulate these assessments and policy directions, more extensive communications become necessary.

Post the 1990s, the clarity of FED releases saw a dip but witnessed a sharp upswing, possibly due to the financial turmoil of 2008. Subsequently, most readability indices settled at an undergraduate academic level. Yet, a minor declining trend persisted in the Gunning-Fog and ARI indices. Moreover, the consistency with Zipf's Law suggests a logical progression in the text, emphasizing pivotal concepts. Our evaluations confirm that this structural alignment with Zipf's Law remained unchanged across the timeline.

Subjects like "Aggregate Demand" and "Credit Market" predominantly shaped the discussions during FOMC sessions. Notably, inflation-related issues, which saw diminished focus in the early 2000s, regained prominence post-2005, consistently representing about 15% of the discourse in the last ten years.

Post-2004, there was a discernible reduction in the Federal Reserve's minutes, especially in neutral statements relative to positive or negative ones. This hints at the FED's increased inclination towards making more decisive statements. Nonetheless, after a triennial period, neutral sentiments made a comeback. The 2007-2008 financial meltdown saw stabilization in the volume of statements. This reduction was paired with variations in both specific and vague forward-looking statements. Forward-looking language saw intriguing shifts between

1999-2003 for specificity and 2005 and 2013-2015 for vagueness. Post an adjustment phase, the volume of forward-looking phrases displayed sporadic changes, especially when juxtaposed with polarity.

Our analysis indicated that the inclusion of the COVID-19 indicator did not significantly amplify the predictive capability of our models for the associated time series, as the p-values surpassed the set significance threshold. In other words, adding the COVID-19 indicator did not notably boost the forecasting power of our models. Furthermore, the polarity model residuals were more congruent with expectations than those of the forward-looking models, indicating a better data fit.

While any residual discrepancies might pose concerns regarding the models' precision and robustness, our research unravels crucial nuances in FED's communication style and focus. For future endeavors, alternative modeling techniques could be considered, or the current models could be enriched by incorporating new variables or refining existing ones. Such enhancements will pave the way for a more profound understanding of central bank communications and their ramifications for both monetary policies and the larger economic framework.

On the other hand, even though there are now numerous domain-specific, fine-tuned versions of BERT-based language models available, these models still face limitations when it comes to extracting significant and contextually meaningful variations. This limitation arises from the extensive presence of rigid structures in high-formality technical documents, such as central bank and government announcements, which can impede their ability to capture nuanced deviations. Further research areas may encompass exploring potential solutions to address these weighting issues.

7. CONCLUSION

The COVID-19 pandemic has had profound and far-reaching effects on countries' economies worldwide, resulting in widespread disruptions in supply chains, business closures, job losses, and fiscal challenges. The economic impacts of the pandemic have been diverse and complex and continue to evolve, posing significant challenges for policymakers and necessitating unprecedented measures to mitigate the economic fallout.

On the other hand, central banks' communication strategies play a pivotal role in shaping consumer and market expectations. It helps to guide decision-making, influence behavior, and manage perceptions about monetary policy and economic outlook. As the literature tells, transparent, timely, and effective communication is crucial in enhancing credibility and building trust, fostering stability in financial markets and the broader economy. A valid communication strategy and the invention of tools adequately aligned with these strategies are essential catalysts for shaping consumer perceptions and guiding economic behavior. Therefore, central banks must adapt their communication accordingly in the face of unexpected circumstances as they play a pivotal role for their countries in navigating through crises.

Unprecedented challenges demand unconventional remedies. Traditional approaches to monetary policy may not be enough to address the current economic challenges, and financial authorities need to be open and adaptable in their communication and decision-making to promote economic recovery. Specifically, central banks ought to embrace communication strategies that are both transparent and forward-looking. These strategies should inform financial markets and the public regarding their policy positions while remaining adaptable and responsive to abrupt shifts in economic conditions.

During the Covid-19 pandemic, central banks worldwide have demonstrated their commitment to effective communication and adaptability. They have implemented a range of measures, including interest rate cuts, liquidity injections, asset purchases, and loan support programs. Moreover, they have adopted proactive communication strategies to ensure that their policy actions are well-understood and aligned with the changing economic landscape. Central banks have engaged in frequent press conferences, issued timely policy statements, and provided regular updates on the evolving situation. These efforts have helped to manage market expectations, maintain financial stability, and foster confidence in the economic recovery process.

Looking ahead, central banks face ongoing challenges as they navigate the path to economic recovery in a post-pandemic world. The lessons learned from the COVID-19 crisis underscore the need for central banks to continuously evaluate and refine their communication strategies. Adapting to the evolving needs and expectations of market participants and the public will be crucial. Central banks should prioritize transparency, clarity, and accessibility in their communication channels, ensuring that their messages reach a wide range of stakeholders. Embracing technological advancements and leveraging digital platforms can enhance the effectiveness of communication efforts.

Furthermore, central banks should also recognize the importance of collaboration and coordination with other policy authorities, such as fiscal authorities and regulatory agencies. Integrated and synchronized policy communication can amplify the impact of policy measures, instill confidence, and facilitate a harmonized response to economic challenges.

To conclude, the COVID-19 pandemic has underscored the pivotal importance of central bank communication in navigating economic challenges. Central banks must adapt their communication strategies to provide transparency, foster stability, and guide economic behavior. The unconventional nature of the crisis necessitates flexible and responsive approaches, with long-term foresight to ensure economic recovery. Central banks have shown their resilience and ability to navigate through uncertain times by implementing comprehensive communication strategies and proactive policy measures. Our study contributes to the existing literature on central bank communication and paves the way for further research on the effects of the pandemic and the effectiveness of different communication strategies in navigating economic crises.

Moving forward, central banks must remain vigilant, continuously evaluate the effectiveness of their communication efforts, and adapt to the changing needs and expectations of the stakeholders they serve. By doing so, central banks can play a crucial role in promoting stability, guiding economic behavior, and fostering a resilient and inclusive post-pandemic recovery.

Notwithstanding the constraints mentioned earlier, our research adds to the body of knowledge on central bank communication and paves the way for future inquiries into the potential consequences of the global pandemic. Subsequent studies could delve into the enduring impacts of central bank communication in times of crises and evaluate the efficacy of various communication strategies in alleviating economic and financial upheavals.

Despite the availability of specialized BERT-based models, their ability to extract nuanced variations remains limited in highly formal technical documents like central bank announcements. Future research could explore ways to address these challenges and enhance their effectiveness.

Last but not least, it is essential to recognize that significant economic crises, like the Global Financial Crisis of 2007-2008, could potentially be analyzed using a similar methodology, allowing for comparative studies. However, delving into such analyses is beyond the scope of this study.

REFERENCES

[1] Romain Tavenard. An introduction to dynamic time warping. `https://rtavenar.github.io/blog/dtw.html`, **2021**.

[2] Klaus Schmidt-Hebbel and Martin Carrasco. *The Past and Future of Inflation Targeting: Implications for Emerging-Market and Developing Economies*, pages 583–622. **2016**. ISBN 978-81-322-2838-7. doi:10.1007/978-81-322-2840-0_18.

[3] Jakob Haan, Cristina Bodea, Raymond Hicks, and Sylvester Eijffinger. Central bank independence before and after the crisis. *Comparative Economic Studies*, 60, **2018**. doi:10.1057/s41294-017-0050-4.

[4] Charles Freedman and Douglas Laxton. Inflation targeting pillars: Transparency and accountability. *International Monetary Fund, IMF Working Papers*, 09, **2009**. doi:10.5089/9781451874075.001.

[5] John F. Muth. Rational expectations and the theory of price movements. *Econometrica*, 29(3):315–335, **1961**. ISSN 00129682, 14680262.

[6] Marvin Goodfriend. Interest rates and the conduct of monetary policy. Working Paper 90-06, Federal Reserve Bank of Richmond, **1990**.

[7] Alan Blinder, Jakob Haan, Michael Ehrmann, Marcel Fratzscher, and David-Jan Jansen. Central bank communication and monetary policy: A survey of theory and evidence. *Journal of Economic Literature*, 46:910–45, **2008**. doi:10.1257/jel.46.4.910.

[8] O. Issing. *The Long Journey of Central Bank Communication*. Karl Brunner distinguished lecture series. MIT Press, **2019**. ISBN 9780262355995.

[9] Shiwei Su, Ahmad Hassan Ahmad, and Justine Wood. How effective is central bank communication in emerging economies? an empirical analysis of the chinese money markets responses to the people's bank of china's policy

communications. *Review of Quantitative Finance and Accounting*, 54, **2020**. doi:10.1007/s11156-019-00822-7.

[10] Rupert Cornwell. If i've made myself too clear, you must have misunderstood me, **1996**.

[11] W. Rayward. Michael d gordin, scientific babel : How science was done before and after global english.isis review. *Isis*, 107:665–666, **2016**.

[12] David Bholat, Stephen Hans, Pedro Santos, and Cheryl Schonhardt-Bailey. *Text mining for central banks*. Centre for Central Banking Studies, Bank of England, **2015**.

[13] Stephen Hansen and Michael McMahon. Shocking language: Understanding the macroeconomic effects of central bank communication. *Journal of International Economics*, 99, **2015**. doi:10.1016/j.jinteco.2015.12.008.

[14] Tiejian Luo, Su Chen, Guandong Xu, and Jia Zhou. *Sentiment Analysis*, pages 53–68. **2013**. ISBN 978-1-4614-7201-8. doi:10.1007/978-1-4614-7202-5_4.

[15] Maite Taboada. Sentiment analysis: An overview from linguistics. *Annual Review of Linguistics*, 2, **2016**. doi:10.1146/annurev-linguistics-011415-040518.

[16] Ramon Ferrer-i Cancho and Ricard Sole. Two regimes in the frequency of words and the origins of complex lexicons: Zipf?s law revisited. *Journal of Quantitative Linguistics*, 8:165–173, **2001**. doi:10.1076/jqul.8.3.165.4101.

[17] Allen H. Huang, Hui Wang, and Yi Yang. Finbert: A large language model for extracting information from financial text*. *Contemporary Accounting Research*, **2022**.

[18] Yi Yang. Yiyanghkust/finbert-fls · hugging face, **2022**.

[19] Lei Ge and Shun Chen. Exact dynamic time warping calculation for weak sparse time series. *Applied Soft Computing*, 96:106631, **2020**. ISSN 1568-4946. doi:https://doi.org/10.1016/j.asoc.2020.106631.

[20] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. **2018**.

[21] David Lucca and Francesco Trebbi. Measuring central bank communication: An automated approach with application to fomc statements. NBER Working Papers 15367, National Bureau of Economic Research, Inc, **2009**.

[22] Magdalena Szyszko, Aleksandra Rutkowska, and Agata Kliber. Do words affect expectations? the effect of central banks communication on consumer inflation expectations. *The Quarterly Review of Economics and Finance*, 86, **2022**. doi:10.1016/j.qref.2022.07.009.

[23] Tim Loughran and Bill Mcdonald. When is a liability not a liability? textual analysis, dictionaries, and 10-ks. *Behavioral & Experimental Accounting eJournal*, **2010**.

[24] Mohammed Tumala and Babatunde Omotosho. A text mining analysis of central bank monetary policy communication in nigeria. *Central Bank of Nigeria Journal of Applied Statistics*, 10:73–107, **2019**. doi:10.33429/Cjas.10219.3/6.

[25] Eyup Kahveci and Aysun Odabaş. Central banks' communication strategy and content analysis of monetary policy statements: The case of fed, ecb and cbrt. *Procedia - Social and Behavioral Sciences*, 235:618–629, **2016**. doi:10.1016/j.sbspro.2016.11.039.

[26] Ki Young Park, Young Joon Lee, and Soohyon Kim. Deciphering monetary policy board minutes through text mining approach: The case of korea. *SSRN Electronic Journal*, **2019**. doi:10.2139/ssrn.3312561.

[27] Andy Moniz and Franciska de Jong. Predicting the impact of central bank communications on financial market investors' interest rate expectations. In *The Semantic Web: ESWC 2014 Satellite Events*, page 144–155. Springer-Verlag, Berlin, Heidelberg, **2022**. ISBN 978-3-319-11954-0. doi:10.1007/978-3-319-11955-7_12.

[28] Philip Stone, Dexter Dunphy, Marshall Smith, and Daniel Ogilvie. *The General Inquirer: A Computer Approach to Content Analysis*, volume 4. **1966**. doi:10.2307/1161774.

[29] Giuseppe Bruno. Text mining and sentiment extraction in central bank documents. *2016 IEEE International Conference on Big Data (Big Data)*, pages 1700–1708, **2016**.

[30] Sahar Sohangir, Nicholas Petty, and Dingding Wang. Financial sentiment lexicon analysis. pages 286–289. **2018**. doi:10.1109/ICSC.2018.00052.

[31] C.J. Hutto and Eric Gilbert. Vader: A parsimonious rule-based model for sentiment analysis of social media text. **2015**.

[32] Jeremy Howard and Sebastian Ruder. Universal language model fine-tuning for text classification. pages 328–339. **2018**. doi:10.18653/v1/P18-1031.

[33] Pekka Malo, Ankur Sinha, Pekka Korhonen, Jyrki Wallenius, and Pyry Takala. Good debt or bad debt: Detecting semantic orientations in economic texts. *Journal of the Association for Information Science amp; Technology*, 65(4):782–796, **2014**.

[34] Dogu Araci. Finbert: Financial sentiment analysis with pre-trained language models, **2019**. doi:10.48550/ARXIV.1908.10063.

[35] Yi Yang, Mark UY, and Allen Huang. Finbert: A pretrained language model for financial communications. **2020**.

[36] Yueying Zhu, Benwei Zhang, Qiuping Wang, Wei Li, and Xu Cai. The principle of least effort and zipf distribution. *Journal of Physics: Conference Series*, 1113:012007, **2018**. doi:10.1088/1742-6596/1113/1/012007.

[37] David Blei, Andrew Ng, and Michael Jordan. Latent dirichlet allocation. In T. Dietterich, S. Becker, and Z. Ghahramani, editors, *Advances in Neural Information Processing Systems*, volume 14. MIT Press, **2001**.

[38] Wiwik Anggraini, Retno Vinarti, and Yuni Kurniawati. Performance comparisons between arima and arimax method in moslem kids clothes demand forecasting: Case study. *Procedia Computer Science*, 72:630–637, **2015**. doi:10.1016/j.procs.2015.12.172.

[39] G. Respondek, Joachim Gröger, Jens Floeter, and Axel Temming. Variability of fishing effort for the german brown shrimp (crangon crangon) fishing fleet: Influencing factors, and seasonal and spatial patterns. *ICES Journal of Marine Science*, 71, **2014**. doi:10.1093/icesjms/fsu016.