Essays on Natural Language Processing and Central Banking.

Inaugural - Dissertation

zur

Erlangung der wirtschaftswissenschaftlichen Doktorwürde

des Fachbereichs Wirtschaftswissenschaften

der Philipps-Universität Marburg

eingereicht von:

Johannes Zahner, M.Sc.

aus Waiblingen

Erstgutacher/in:	Prof. Dr. Elisabeth Schulte
Zweitgutachter/in:	Prof. Dr. Bernd Hayo
Einreichungstermin:	03.08.2021
Prüfungstermin:	08.09.2021
Erscheinungsort:	Marburg
Hochschulkennziffer:	1180

Acknowledgments

Failures and the ability to learn from them are the foundations of scientific achievement. While this thesis contains four essays, they only represent a small fraction of the ideas pursued and effort invested. My gratitude – and these acknowledgements – are particularly directed towards those who have kept me going when failure seemed certain and success improbable. The thesis you are currently reading would not be possible without a great number of people listening to my ideas, correcting my blunders, and critiquing my work. While it is impossible to name everyone, there are a some who stood out.

First and foremost, I would like to thank my advisors, who have been invaluable sources of wisdom and assistance. Elisabeth Schulte provided me with the opportunity and means to pursue my PhD in her research group. Her assistance, particularly during difficult times, allowed me to continue exploring research topics and ideas in which I was naturally interested, even if they often did not appear to have scientific potential at the time. Bernd Hayo, whose interest for my work on the one hand and repeated destruction of my use of empirical methods on the other, was immensely valuable to me. And Marc Steffen Rapp, whose exploratory curiosity led me to machine learning and neural networks, a topic (and hobby) that I have been following closely ever since.

I owe a huge debt of gratitude to my wife, Ayca Cevener Zahner; I simply couldn't have done it without her sacrifices, optimism, and unique sense of humor.

Ideas generally emerge as a result of collaboration with peers who share common interests in similar subjects. I was fortunate to have met and worked with Jonas Gross and Martin Baumgärtner, for whom – after countless hours – I have developed the utmost respect and admiration. Having met few people as diligent as Jonas and as dedicated as Martin, I will miss our time together.

Finally, I would like to thank Dominik Alfke & Alexandra Bünger, Nadine Blöbaum, Maximilian Burger, Sebastian Friedrich & Alba Palazzolo, Patrick Hoffmann, Matthias & Hatice Mayer, Tilo Mohaupt Christoph Rössler, Leander Stecker, Madeleine Schröter, and Simone Zahner for the countless hours of joy and laughter.

Dokumentation der Ko-Autorenschaft

Kapitel 1 wurde von Jonas Gross und mir verfasst. Die Idee für das Projekt stammt von Jonas Gross, die Datenanalyse und technische Umsetzung von mir. Die Analyse haben wir gemeinsam durchgeführt.
Kapitel 2 habe ich in Alleinautorenschaft verfasst.
Kapitel 3 wurde von Kai Henseler, Bernd Hayo, Marc Steffen Rapp und mir verfasst. Die Idee für das Papier, sowie die erste Version stammen von Kai Henseler, Bernd Hayo und Marc Steffen Rapp. Die Text-Analyse in der neusten Fassung stammt von mir und die Überarbeitungen wurden gemeinsam durchgeführt.
Kapitel 4 wurde von Martin Baumgärtner und mir verfasst. Die Idee und das Konzept des Projekts stammen von mir, die Zusammenführung des Textkorpus, sowie die technische Umsetzung der Sprachmodelle von Martin Baumgärtner. Die Evaluation & Applikationen haben wir gemeinsam durchgeführt.

Summary

Humans generally interact, communicate, and form social structures using natural language. Due to the high dimensionality of language, much of the wealth of information from these interactions has been barred from the economic profession. However, recent technological advancements lead to increasing use of text as an underlying datasource in economic and financial applications. This trend has been further accelerated by Nobel laureate Robert J. Shiller's presidential address to the American Economic Association *"Narrative Economics"*, in which he argues for more elaboration on narratives – stories that affect individual decisions and collective actions – by the economic scientific community. Addressing this gap in the literature, research has been published utilizing textual information to quantify latent variables such as uncertainty, forecasting macroeconomic variables in real time, and asset price predictions.

In conjunction with the rise of natural language processing applications, there has been a shift in perspective on monetary policy with regards to central bank transparency and communication. Transitioning from the presumption that monetary policy is limited to interest rate actions, communication has advanced to become a key tool in the central banker's toolbox. Ever since, words are used to anchor expectations and self-enforce the central banks' desired equilibrium path. As a result, research on monetary policy has been relentless in the pursuit of adopting novel techniques as well as incorporating new unstructured data sources such as news-articles, press conference statements, and speeches. This string of literature is regularly complemented by an extension of the traditional empirical toolbox, borrowing novel techniques from the field of machine learning.

The here presented cumulative dissertation consists of four essays that touch on all these fields, namely text as data, monetary policy, and machine learning. My primary focus is on the European Central Bank (ECB), but the methodology and ideas can be extended to other central banks as well.

Throughout this thesis, textual information is incorporated from different data sources, analyzed using different techniques in order to approximate different latent variables. As a result, text is employed as a dependent variable at times and

as an independent variable at other times. Specifically, the first essay leverages the relative frequency of terms used in ECB press statements as anecdotal evidence for the diversity of the central banks' communication with regard to their topics, whereas the second essay counts positive and negative terms in speeches to approximate the latent variable of central bank loss. The third essay examines the impact of linguistic complexity on financial market participants by conducting a readability test on the ECB's introduction statements, and the final essay dives into computational linguistics to develop a novel central bank-specific language model for better quantifying monetary policy communication. The following is a brief summary of the four essays included in this thesis.

My first essay analyzes rule-based monetary policy in the euro area before and after the financial crisis. Jonas Gross and I argue that the environment in which policymakers operate is far more complex than traditional model-based analysis of policy rules permits. We complement this view with evidence from ECB press conferences, demonstrating that the central bank discusses a wide range of topics beyond the traditional Taylor-rule variables. Since each variable has the potential to be relevant in understanding the central bank's reaction function, we combine a literature review with natural language processing to identify a set of potential determinants. The traditional approach of selecting a single interest rate response function is then contrasted by applying a Bayesian model averaging approach to these determinants. We account for model uncertainty by including a large number of determinants and estimating a total of 33.000 different model combinations.
Our results suggest that in contrast to the ongoing criticism, the ECB primarily reacts to inflation in its interest rate decision. In fact, our analysis finds that inflation is a significant variable in almost all of the examined model combinations. Furthermore, we find that the ECB reacts to changes in economic activity determinants such as unemployment and production as well. These economic activity indicators were a priority for the ECB prior to the financial crisis but have since declined in relevance, suggesting that inflation is the sole driver of monetary policy decisions in the post-crisis period. Finally, we assess our findings with textual

evidence from the ECB press conferences, where, in accordance with the previous results, we find the same shift.

My second essay focuses on the ECB's objective itself, quantifying the central bank's satisfaction with current economic conditions through textual analysis. By maximizing an implied objective function, the ECB is assumed to pursue inflation targeting with a subordinate focus on supporting the general economic policy of the European Union. I compute the central bank's sentiment using the ECB's public communication by counting the number of positive and negative words in speeches, allowing me to quantify the objective. Assuming a typical functional form for the objective allows me to estimate the optimal levels with respect to inflation and economic activity, i.e. the bliss points in which the central banks communication is the most positive.
Using a dictionary approach to estimate the sentiment index yields several interesting results. The most surprising is, unquestionable, a concave inflation objective with an implied inflation target beyond the banks' mandate and best described as 'above, but close to 2%'. Deviations from this bliss point appear to lower the satisfaction, and hence the optimistic language in speeches. With respect to the subordinate objective, I find a convex objective towards output growth and a linear objective towards the unemployment rate. Furthermore, my results suggest that deviations from the primary objective, the inflation rate, appear to have no greater effect on the speeches' language than deviations from either of the subordinate objectives. In fact, in contrast to inflation, both output and unemployment are consistently significant variables. Finally, contrary to findings in the United States, financial market conditions have no significant influence on the ECB's sentiment.

In the third essay, Bernd Hayo, Kai Henseler, Marc Steffen Rapp, and I investigate the impact of central bank communication on financial markets. We are particularly interested in the communication's complexity and how it affects financial market trading. To examine this relationship empirically, we employ high-frequency data from European stock index futures during the introductory statement of the ECB's press conferences. A readability test on the introductory

statement during the press conference determines the statements' linguistic complexity. In conjunction with the central banks' unique communication design, we are able to separate the effect of verbal complexity on trading during the introductory statements and the subsequent Q&A session. Our sample contains announcements of novel Unconventional Monetary Policy Measures (UMPM), enabling us to investigate whether the content of the introductory statements interacts with the reaction of traders to its linguistic complexity.
We find that the Q&A sessions are – in terms of linguistic complexity – less complex and thus more comprehensible. When UMPM are announced, contemporaneous trading volumes are negatively correlated with complexity, resulting in a temporal shift of trading towards the less complex Q&A session. This shift is first indication that financial markets respond to linguistic complexity in a context-specific manner. This line of reasoning is strengthened further by the observation that events containing UMPM are less similar in terms of wording to previous statements. As a result, we believe that financial market traders are underreacting to novel complex information in introductory statements regarding UMPM. The subsequent discussion and clarification of the cognitively costly content during the Q&A session mitigates this effect, shifting trading from the introductory statement phase to the Q&A phase of the ECB's press conference.

The final essay concerns the quantification of central bank communication, i.e. it explores how text in monetary policy can be effectively summarised and analysed. Martin Baumgärtner and I propose a novel language model, build on machine learning, as a tool to quantify central bankers qualitative information.
The necessity and feasibility of measuring central bank communication in this manner stems from two major developments in the fields of monetary policy and machine learning over the last two decades. On the one hand, central bankers' communication, as well as its analysis, has increased substantially. This progress necessitates some form of quantification of the qualitative components, a research topic dominated by dictionary approaches. On the other hand, advances at the intersection of linguistics and computer science enabled the use of machine learning to train language models capable of adequately capturing the languages multidi-

mensionality and context-dependence. The resulting models are regularly open source. However, the technical jargon of central bankers renders them generally unsuitable for use in the field. This essay aims to apply computational linguistics research to monetary policy by developing a language model exclusively trained on central bank communication.

To accomplish this, we gather a large and diverse text corpus, which we use to compare a number of state-of-the-art machine learning algorithms. Choosing the most promising, we develop a central bank specific language model.

Several applications are presented to showcase the broad applicability of our language model. First, we propose a novel technique for comparing central banks, affirming that similarity is driven by mutual objectives. Next, we construct a time-series index that reflects the ECB's willingness to act as a lender of last resort. The index suggests that communication similar to Mario Draghi's 'whatever it takes' speech can calm financial markets during times of high uncertainty. The third application emphasizes the presence of prejudices even in central bankers' technical language. We demonstrate how social patterns, such as occupational gender distribution, are reflected in their communication. The final application is a forecasting exercise that suggests that speeches may be more accurate predictors than previous research suggests.

Zusammenfassung

Menschliche Interaktion, Kommunikation, sowie die Bildung sozialer Strukturen ereignen sich in der Regel mittels natürlicher Sprache. Durch die hohe Dimensionalität der Sprache war ein Großteil der reichhaltigen Informationen, die sich aus diesen Interaktionen ergeben, den Wirtschaftswissenschaften bisher vorenthalten. Jüngste technologische Fortschritte führen jedoch zu einer zunehmenden Verwendung von Text als Datenbasis in ökonomischen Anwendungen. Diese Entwicklung wurde durch die Rede des Nobelpreisträgers Robert J. Shiller vor der American Economic Association *"Narrative Economics"* weiter beschleunigt, in der er sich für eine stärkere Berücksichtigung von Narrativen – Geschichten, die individuelle Entscheidungen und kollektive Handlungen beeinflussen – durch die wirtschaftswissenschaftliche Gemeinschaft einsetzt. In den letzten Jahren wurden Forschungsarbeiten veröffentlicht, die textbasierte Informationen nutzen, um latente Variablen wie Unsicherheit zu quantifizieren, makroökonomische Variablen in Echtzeit zu prognostizieren und Vermögenspreise vorherzusagen.

Zur selben Zeit hat sich die Sichtweise der Geldpolitik im Hinblick auf die Transparenz und Kommunikation der Zentralbank gewandelt. Statt Geldpolitik auf Zinsmaßnahmen zu beschränken, ist die Kommunikation zu einem wichtigen Instrument im Werkzeugkasten der Zentralbanker geworden. Heute werden Worte verwendet, um Erwartungen zu verankern und den von den Zentralbanken gewünschten Gleichgewichtspfad zu unterstützen. Infolgedessen hat die Erforschung der Geldpolitik unermüdlich neue Methoden zum Messen von Text und neue unstrukturierte Datenquellen wie Nachrichten, Pressekonferenzerklärungen und Reden in ihre Forschung einbezogen. Die Literatur wird dabei laufend durch eine Erweiterung der traditionellen empirischen Instrumente aus dem Bereich des maschinellen Lernens ergänzt.

Die hier vorgelegte kumulative Dissertation besteht aus vier Aufsätzen, die sich auf alle diese Bereiche beziehen, nämlich Text als Daten, Geldpolitik und maschinelles Lernen. Mein Hauptaugenmerk liegt auf der Europäischen Zentralbank (EZB), aber die Methoden und Ideen lassen sich auch auf andere Zentralbanken übertragen.

In dieser Dissertation wird Text aus unterschiedlichen Quellen einbezogen und mit unterschiedlichen Techniken analysiert, um unterschiedliche latente Variablen zu ermitteln. Infolgedessen wird der Text manchmal als abhängige Variable und manchmal als unabhängige Variable verwendet. Der erste Aufsatz zählt die relative Häufigkeit von Begriffen in EZB Presseerklärungen, um die Vielfalt der Kommunikation der Zentralbank in Bezug auf ihre Themen anekdotisch zu belegen, während der zweite Aufsatz sich auf positive und negative Begriffe in Reden konzentriert, um die latente Variable des Zentralbankverlustes zu approximieren. Im dritten Aufsatz werden die Auswirkungen sprachlicher Komplexität auf Finanzmarktteilnehmer untersucht, und der letzte Aufsatz wendet sich an die Computer-Linguistik zur Entwicklung eines neuartigen zentralbankspezifischen Sprachmodells zur besseren Quantifizierung der geldpolitischen Kommunikation. Im Folgenden finden Sie eine kurze Zusammenfassung der vier Aufsätze dieser Dissertation.

Mein erster Aufsatz analysiert die regelbasierte Geldpolitik im Euroraum vor und nach der Finanzkrise. Jonas Gross und ich argumentieren, dass das Umfeld, in dem geldpolitische Entscheidungsträger agieren, weitaus komplexer ist, als es traditionelle modellbasierte Analysen der geldpolitischen Regeln erlauben. Wir ergänzen diese Sichtweise durch Erkenntnisse aus EZB-Pressekonferenzen, die zeigen, dass die Zentralbank ein breites Spektrum an Themen jenseits der traditionellen Taylor-Regel-Variablen diskutiert. Da jede Variable das Potenzial hat, für das Verständnis der Reaktionsfunktion der Zentralbank relevant zu sein, kombinieren wir eine Literaturanalyse mit der Auswertung von Pressekonferenzen, um eine Reihe potenzieller Einflussfaktoren zu ermitteln. Der traditionelle Ansatz einer einzigen Zinsreaktionsfunktion wird dann durch die Anwendung von Bayesian Model Averaging auf diese Determinanten gegenübergestellt. Da wir eine große Anzahl von Determinanten mit einbeziehen, schätzen wir insgesamt 33.000 verschiedene Modellkombinationen.
Unsere Ergebnisse deuten darauf hin, dass die EZB entgegen der anhaltenden Kritik bei ihren Zinsentscheidungen in erster Linie auf Änderungen in der Inflation reagiert. Die Inflationsrate stellt in fast allen untersuchten Modellkom-

binationen eine signifikante Variable dar. Darüber hinaus stellen wir fest, dass die EZB auch auf Veränderungen bei Konjunkturindikatoren wie Arbeitslosigkeit und Produktion reagiert. Wir stellen jedoch eine Verschiebung der Gewichtung im Laufe der Zeit fest. Konjunkturindikatoren hatten vor der Finanzkrise für die EZB Priorität, haben aber seitdem an Bedeutung verloren, was darauf hindeutet, dass die Inflationsrate die einzige treibende Kraft für geldpolitische Entscheidungen seit der Krise ist. Abschließend bewerten wir unsere Ergebnisse anhand von Textdaten aus den EZB-Pressekonferenzen, wo wir in Übereinstimmung mit den vorherigen Ergebnissen die gleiche Verschiebung feststellen.

Mein zweiter Aufsatz konzentriert sich auf die Zielsetzung der EZB und quantifiziert die Zufriedenheit der Zentralbank mit den aktuellen wirtschaftlichen Bedingungen anhand einer Textanalyse. Es wird angenommen, dass die EZB durch die Maximierung einer impliziten Zielfunktion ein Inflationsziel verfolgt, dem die Unterstützung der allgemeinen Wirtschaftspolitik der Europäischen Union untergeordnet ist. Ich berechne die Zufriedenheit der Zentralbank anhand der öffentlichen Kommunikation der Zentralbank, indem ich die Anzahl der positiven und negativen Wörter in den Reden zähle. Dies ermöglicht es mir die Zielfunktion zu quantifizieren. Unter Annahme einer typischen funktionalen Form kann ich dann die optimalen Niveaus in Bezug auf Inflation und Wirtschaftstätigkeit ermitteln, d.h. die Werte, bei denen die Kommunikation der Zentralbank am positivsten ist.
Die Verwendung eines Wörterbuchansatzes zur Schätzung des Zielfunktion führt zu mehreren interessanten Ergebnissen. Das überraschendste Resultat dürfte die konkave Inflationszielfunktion mit einem impliziten Inflationsziel von 'über, aber nahe 2%' sein. Bei Abweichungen von dieser Inflationsrate wird die Sprache in den Reden pessimistischer. In Bezug auf das untergeordnete Ziel stelle ich eine konvexe Zielfunktion im Hinblick auf das Produktionswachstum und eine lineare Zielfunktion im Hinblick auf die Arbeitslosenquote fest. Darüber hinaus deuten meine Ergebnisse darauf hin, dass Abweichungen vom primären Ziel, der Inflationsrate, keinen größeren Einfluss auf die Sprache der Reden haben als Abweichungen von einem der Sekundärziele. Im Gegensatz zur Inflation sind

vielmehr sowohl die Produktion als auch die Arbeitslosigkeit durchweg signifikante Variablen. Abschließend lässt sich feststellen, dass die Finanzmarktbedingungen im Gegensatz zu den Ergebnissen in den Vereinigten Staaten keinen signifikanten Einfluss auf die Reden der EZB haben.

Im dritten Aufsatz untersuchen Bernd Hayo, Kai Henseler, Marc Steffen Rapp und ich die Auswirkungen von Zentralbank-Kommunikation auf die Finanzmärkte. Wir interessieren uns insbesondere wie die Komplexität dieser Kommunikation den Finanzhandel beeinflusst. Um ebendiese Beziehung empirisch zu analysieren, verwenden wir Hochfrequenzdaten von europäischen Aktienindex-Futures während der Einführungsstatements der EZB-Pressekonferenzen. Ein Lesbarkeits-Test bestimmt die sprachliche Komplexität dieser Statements. In Verbindung mit dem einzigartigen Pressekonferenz-Konzept der EZB sind wir in der Lage, die Auswirkungen von erhöhter sprachlicher Komplexität auf den Handel zu messen, sowie die Effekte zwischen der einleitenden Erklärung und der anschließenden Q&A-Runde zu trennen. Da unsere Stichprobe Ankündigungen unkonventioneller geldpolitischer Maßnahmen (UMPM) enthält, sind wir in der Lage zu analysieren, ob der Inhalt der Erklärungen mit der Reaktion der Händler auf ihre sprachliche Komplexität zusammenhängt.

Wir stellen fest, dass die Q&A-Runden – in Bezug auf die sprachliche Komplexität – weniger komplex und damit verständlicher sind. Nur wenn UMPM angekündigt werden, korrelieren die Handelsvolumina negativ mit Komplexität, was zu einer zeitlichen Verschiebung des Handels in Richtung der weniger komplexen Q&A-Runde führt. Diese Verschiebung ist ein erster Hinweis darauf, dass die Finanzmärkte auf sprachliche Komplexität kontextspezifisch reagieren. Diese Argumentation wird durch die Beobachtung verstärkt, dass Pressekonferenzen, die UMPM enthalten, vom Wortlaut her weniger Ähnlichkeiten mit früheren Konferenzen aufweisen.

Infolgedessen glauben wir, dass die Finanzmarkthändler auf neuartige komplexe Informationen in einleitenden Erklärungen zu UMPM reagieren. Die anschließende Diskussion und Erläuterung des kognitiv kostspieligen Inhalts während der Q&A-Runde mildert den Effekt ab und verlagert so den Handel von der Phase der Einführungsstatements in die Q&A-Phase.

Der letzte Aufsatz befasst sich mit der Quantifizierung von Zentralbank-Kommunikation, d.h. er untersucht, wie geldpolitische Texte effektiv zusammengefasst und analysiert werden können. Martin Baumgärtner und ich schlagen ein neuartiges Sprachmodell vor, das auf maschinellem Lernen aufbaut.
Die Notwendigkeit und Durchführbarkeit, die Kommunikation von Zentralbanken auf diese Weise zu messen, ergibt sich aus zwei wichtigen Entwicklungen im Bereich der Geldpolitik und des maschinellen Lernens in den letzten zwei Jahrzehnten. Zum einen hat die Kommunikation der Zentralbanker, sowie deren Analyse, erheblich zugenommen. Dieser Fortschritt erfordert eine Art von Quantifizierung der qualitativen Komponenten, ein Forschungsthema, das von Wörterbuchansätzen dominiert wird. Andererseits ermöglichten Fortschritte an der Schnittstelle zwischen Linguistik und Informatik den zunehmenden Einsatz von maschinellem Lernen. Die so trainierten Sprachmodelle können die Mehrdimensionalität und Kontextabhängigkeit von Sprache adäquat erfassen. Die daraus resultierenden Modelle sind in der Regel frei zugänglich. Aufgrund des Fachjargons der Zentralbanker sind sie jedoch für den Einsatz in der Praxis im Allgemeinen ungeeignet. In diesem Aufsatz wenden wir Computer-linguistische Forschung auf die Geldpolitik an, indem wir ein Sprachmodell entwickeln, das ausschließlich auf die Kommunikation von Zentralbanken trainiert ist.
Zu diesem Zweck erstellen wir einen umfangreichen und vielfältigen Textkorpus, mit welchem wir eine Reihe von modernsten Algorithmen für maschinelles Lernen vergleichen. Wir wählen den vielversprechendsten aus und entwickeln ein zentralbankspezifisches Sprachmodell.
Anhand mehrerer Anwendungen wird die breite Anwendbarkeit dieses Sprachmodells aufgezeigt. Zunächst schlagen wir eine neue Methode zum Vergleich von Zentralbanken vor, die verdeutlicht, dass die Ähnlichkeit durch gemeinsame Ziele bedingt ist. Anschließend konstruieren wir einen Zeitreihen-Index, der die Bereitschaft der EZB widerspiegelt, als letzte Refinanzierung-Instanz zu agieren. Der Index deutet darauf hin, dass eine ähnliche Kommunikation wie Mario Draghis "whatever it takes"-Rede die Finanzmärkte in Zeiten großer Un-

sicherheit beruhigen kann. Die dritte Anwendung unterstreicht das Vorliegen von Vorurteilen, selbst in dem technischen Jargon von Notenbankern. Wir zeigen, wie sich soziale Muster, z.B. die Verteilung der Geschlechter im Beruf, in der Kommunikation der Zentralbanker widerspiegeln. Die letzte Anwendung ist eine Prognoseübung, welche darauf hindeutet, dass Notenbankreden möglicherweise genauere Vorhersagen ermöglichen, als die bisherige Forschung vermuten lässt.

Contents

List of Figures **xxiii**

List of Tables **xxv**

List of Acronyms **xxvii**

1 What is on the ECB's mind? Monetary policy before and after the global financial crisis **1**

1.1 Introduction . . . 2
1.2 The ECB's potential interest rate determinants . . . 4
1.3 Estimation approach . . . 10
1.4 Data . . . 18
1.5 Results . . . 20
1.6 Robustness checks . . . 28
1.7 Conclusion . . . 30

2 Above, but close to two percent. Evidence on the ECB's inflation target using text mining **33**

2.1 Introduction . . . 34
2.2 Literature review . . . 36
2.3 Econometric model . . . 38
2.4 Data . . . 41
2.5 Results . . . 49
2.6 Robustness checks . . . 55
2.7 Conclusion . . . 58

3 Complexity of ECB Communication and Financial Market Trading **61**

3.1 Introduction . . . 62
3.2 Central Bank Communication and Financial Markets . . . 64
3.3 Sample and Descriptive Analysis . . . 67
3.4 Regression Analysis . . . 76

3.5 Robustness of Results . . . 83
3.6 Conclusion . . . 89

4 Whatever it takes to understand a central banker - Embedding their words using neural networks 91
4.1 Introduction . . . 92
4.2 Related literature . . . 94
4.3 Methodology . . . 97
4.4 Evaluation of language models . . . 107
4.5 Applications . . . 113
4.6 Conclusion . . . 126

A Appendix: What is on the ECB's mind? 129
A.1 Taylor rule estimate . . . 129
A.2 Textual analysis: Most mentioned bigrams in ECB communication 129
A.3 Summary of variables . . . 130
A.4 Ten top models (pre- and post-crisis) . . . 131
A.5 Robustness checks . . . 132

B Appendix: Above, but close to two percent. 134
B.1 Sample text . . . 134
B.2 Similarity index . . . 135
B.3 Descriptive speaker summary . . . 136
B.4 Unemployment target illustration . . . 137
B.5 Robustness Check Regressions . . . 138

C Appendix: Complexity of ECB Communication. 139
C.1 UMPM Press conferences . . . 139
C.2 Discriptive summary . . . 140

D Appendix: Whatever it takes to understand a central banker. 141
D.1 Graphical illustrations of text corpus . . . 141
D.2 Language Model specifications . . . 142
D.3 Additional evaluation . . . 142

D.4 Applications - Robustness checks 145

References **158**

List of Figures

1.1 Comparison of the actual and Taylor rule-approximated ECB interest rate. . . . 5
1.2 Comparison of the shadow rate by Wu and Xia (2016) and the EONIA rate. . . . 20
1.3 Ten Top Models (whole period). . . . 22
1.4 Relative Frequency of Selected Terms in ECB Communication. . . . 24
1.5 Model Sizes of the Different Time Horizons. . . . 26
1.6 Comparison of BMA-derived Approximation vs. Actual Interest Rate. . . . 28
2.1 Similarity of ECB speeches. . . . 42
2.2 Monthly sentiment score. . . . 45
2.3 Speaker Sentiment 46
2.4 Bin-Scatter Plot of Output Regression (3) 52
2.5 Bin-Scatter Plot of Robustness Check Regression (1) 57
3.1 Complexity distribution of the ECB's communication 71
3.2 Excess trading pattern 73
3.3 Temporal distribution of trading volumes 74
3.4 Trading volumes and complexity 75
4.1 How to get a language model 98
4.2 Illustration of frequency of used terms between Federal Open Market Committee (FOMC) and European Central Bank (ECB) 100
4.3 Graphical illustration of Mikolov, Yih, et al.'s (2013) Word2Vec model. . . . 103
4.4 Graphical illustration of Le and Mikolov (2014)'s Doc2Vec model. . . . 104
4.5 Central banks' similarity 113
4.6 Similarity of all European Central Bank speeches to the "Whatever it takes" speech. . . . 116
4.7 Relative frequency of hawkish and dovish speeches preceding an ECB press-conference. . . . 122
4.8 Regression results of rolling window approach. . . . 124

A.1 Wordcloud: ECB communication 129
A.2 Top models: Pre crisis. 131
A.3 Top models: Post crisis. 131
B.1 Bin-Scatter Plot of Unemployment Regression (6) 137
D.1 Descriptive summary of the corpus 141
D.2 Illustration of frequency of used terms between ECB other central banks. 141
D.3 Evaluation of Embeddings 143

List of Tables

1.1 Regression results . . . 21

1.2 Model Fit Evaluation . . . 27

2.1 Statistic summary of Sentiment Index . . . 44

2.2 Regression Results . . . 50

3.1 Overview of variable definitions . . . 77

3.2 Analysis of trading volume . . . 78

3.3 Analysis of the temporal distribution of trading volume . . . 80

3.4 Similarity analysis . . . 82

3.5 Robustness check - Time horizon . . . 84

3.6 Robustness check - Vague talk . . . 85

3.7 Definitions of Alternative Complexity Measures . . . 86

3.8 Coefficients for alternative measures of complexity . . . 87

3.9 Additional Variables for Quantifying Communication - Related Aspects . . . 87

3.10 Coefficients for complexity measures based on factor analysis . . . 88

4.1 Corpus Summary . . . 98

4.2 Model Overview . . . 107

4.3 Evaluation results of algorithms. . . . 109

4.4 Intrinsic Evaluation: Similarity in selected word embeddings. . . . 111

4.5 Intrinsic Evaluation: Homonym across language models. . . . 112

4.6 Inflation target regression . . . 115

4.7 Regression results: Whatever it takes . . . 117

4.8 Academic profession association by gender . . . 120

4.9 Regression results - Gender Bias . . . 121

4.10 Regression results: Altavilla et. al. (2019) shocks . . . 123

4.11 Out of sample accuracy Altavilla surprises. . . . 125

A.1 Regression results: Taylor rule . . . 129

A.2 Overview of used data and transformations. . . . 130

A.3 Robustness: Prior on the model space . . . 132

A.4 Robustness: Prior on the parameter space . . . 132

A.5 Robustness: Starting date of the crisis 133
A.6 Robustness: EONIA 133
B.1 Summary statistics per speaker 136
B.2 Regression Results Robustness Checks 138
C.1 ECB's GCM press conferences included in the sample 139
C.2 Descriptive summary 140
D.1 Hyperparameter Settings for Evaluation 142
D.2 Additional Intrinsic Evaluation: Homonym across language models. 144
D.3 Application 2: Whatever it takes - Full table 145
D.4 Regression results - Gender Bias 146

List of Acronyms

AIC	Akaike information criterion
BIC	Bayesian Information Criterion
BMA	Bayesian Model Averaging
BRIC	Benchmark Prior
BoE	Bank of England
BIS	Bank of International Settlement
BoJ	Bank of Japan
CISS	Composite Index of Systemic Stress
EONIA	Euro OverNight Index Average
ECB	European Central Bank
EU	European Union
Fed	Federal Reserve
FOMC	Federal Open Market Committee
GDP	Gross Domestic Product
HICP	Harmonized Index of Consumer Prices
HP	Hodrick-Prescott
KNN	K-Nearest-Neighbor
LDA	Latent Dirichelet Allocation
MSE	Mean Squared Error
MRO	Main Refinancing Operations
NLP	Natural Language Processing
OLS	Ordinary Least Squares
PIP	Posterior Inclusion Probability
RIC	Risk Inflation Criterion
RND	Relative Norm Distance
RTD	Real Time Database
SPF	Survey of Professional Forecasters
UIP	Unit Information Prior
UMPM	Unconventional Monetary Policy Measures
ZLB	Zero Lower Bound

1 What is on the ECB's mind? Monetary policy before and after the global financial crisis[*]

Jonas Gross[a] and Johannes Zahner[b]

Abstract

This paper analyzes the monetary policy of the European Central Bank (ECB) both before and after the outbreak of the global financial crisis in 2008. In the literature, researchers typically select *one* Taylor rule-based model to analyze monetary policy of central banks and to derive determinants for the interest rate setting. However, uncertainty about the choice of this respective model is typically neglected. In contrast, we apply a Bayesian model averaging (BMA) approach to extend the Taylor rule to account for model uncertainty driven by heterogeneity in the ECB's decision-making body, the Governing Council. Our results suggest the following: First, the ECB focuses on the inflation rate when setting interest rates. Second, economic activity indicators were in the focus of the ECB before the financial crisis. Third, over the last decade, the role of economic activity decreased, indicating that inflation is the main driver of monetary policy decisions in the post-crisis period. Fourth, when setting interest rates, central bankers appear to consider more than one model.

Keywords: Taylor rule, Bayesian model averaging, Model uncertainty, Monetary policy, ECB.

JEL classification: C11, E43, E58.

[*]This essay is published as Gross, J., & Zahner, J. (2021). What is on the ECB's mind? Monetary policy before and after the global financial crisis. *Journal of Macroeconomics*, 68, 103292.

[a]Faculty of Law, Business & Economics, International Economics and Finance, Universität Bayreuth, Germany.

[b]Corresponding author. School of Business and Economics, Institutional Economics Research Group, Philipps-Universität Marburg, Germany.

1.1 Introduction

> *"The challenge for monetary policy in practice is to retain the virtues of rule-based policy-making, while taking into account the complex, uncertain and constantly evolving environment facing monetary policy-makers."*
>
> — European Central Bank (2001, p. 38)

Due to its pivotal role in monetary policy, the interest rate setting of central banks is one of the most debated topics in the field of macroeconomics. When setting its interest rates, the ECB advocates a rule-based approach while retaining room for discretionary interventions. Generally, researchers apply the so-called Taylor rule in order to analyze the interest rate setting (see Taylor, 1993). The Taylor rule proposes that inflation and economic activity can well approximate the short-term interest rate. Various studies suggest that this simple rule is a powerful tool to approximate central bank interest rates under special conditions (see e.g. Sturm and Wollmershäuser, 2008).

However, there are various reasons why applying a standard Taylor rule yields misleading policy implications. First, it seems far-fetched to restrict decision-makers to a monetary policy rule that captures sufficient information about the real economy in only two variables. Vítor Constâncio, former vice president of the ECB, argues that

> "*the environment in which monetary policy-makers have to act is much more complex than what is assumed in model-based analysis of policy rules. [...] A simple rule that responds to one or two macroeconomic variables and ignores all other indicators of price developments is not able to account for the complexities of the real world."*
>
> — Vitor Constancio (2017)

Second, monetary policy decisions are based on incomplete information and uncertainties about the actual state of the economy. Hence, central bankers might analyze a variety of different economic variables and indicators – besides the inflation rate and economic activity – to obtain more accurate information about the state of the economy (see Milani, 2008; Lee, Olekalns, et al., 2013; Lee, Morley, et al., 2015).

Third, in reality, central bankers might have more than one model in mind of how the economy functions. The interest rate setting body of the ECB, the Governing Council, consists of the presidents of the euro area national central banks and the Executive Board. These central bankers have different backgrounds that might affect their attitudes towards relevance of certain variables such as inflation and economic activity and the importance of indicators such as bond yields to affect those variables. Although we remain agnostic about the source of heterogeneity, central bankers differ with respect to their social, political, and academic backgrounds. However, aside from potential heterogeneity between the *different* members of the Governing Council, it can further be assumed that heterogeneity exists within *each* central banker, i.e., that more than one reaction function is in her mind. Such heterogeneity leads to different concepts about the transmission of shocks and the interaction of economic agents that might yield deviating policy implications and interest rate recommendations. Therefore, we argue that the standard Taylor rule should be extended to draw more precise inferences for monetary policy.

We contribute to the current monetary policy literature by shedding light on the ECB's monetary policy decisions and analyze the potential shift of priorities due to the global financial crisis in 2008. Our analysis focuses on the following two key factors: Firstly, uncertainty about the form of the central bankers' reaction functions, and secondly, uncertainty about the magnitude of the coefficients included in the specific reaction functions. We base our analysis on real-time data and insights derived from textual analysis of ECB press conference statements.

Employing an empirical Bayesian Model Averaging (BMA) approach allows us to consider variables besides the ones included in the classical Taylor rule, and to evaluate $\sim$33,000 model combinations of potential monetary policy determinants. We consider a variety of variables and evaluate all model combinations with respect to the observed data to determine the most likely models. Thereby, we derive the ECB's most likely interest rate determinants. Applying BMA in the context of the ECB's monetary policy is – to the best of our knowledge – a novel approach and addresses a gap in the current literature.

Our key findings are as follows: First, the ECB focuses its decisions mainly on the inflation rate measured by the Harmonized Index of Consumer Prices (HICP). Second, economic activity seems to be a key priority for the central bank before the financial crisis. Third, our results suggest that the importance of economic activity for the ECB's monetary policy decisions decreased over the last decade. Fourth, when setting interest rates, central bankers from the ECB tend to consider more than one model. This finding supports the necessity to use model averaging techniques in order to take model uncertainties in the context of monetary policy into account.

Our paper is structured as follows: In section two, we identify the variables potentially influencing the ECB's monetary policy by a literature review and a textual analysis of ECB communication. Furthermore, we motivate why one should consider model uncertainty in the context of monetary policy. In section three, the BMA approach is discussed. In section four the data used is discussed. In sections five and six, the estimation results are explained in detail, and robustness checks are conducted. The last section concludes the paper.

1.2 The ECB's potential interest rate determinants

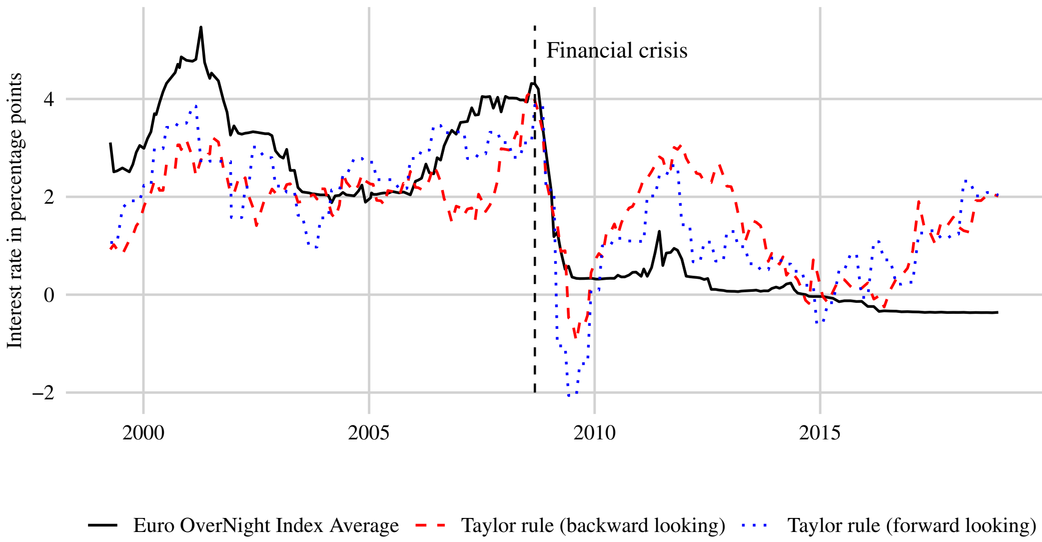
In 1993, John Taylor proposed that the central bank-set short-term interest rate i can be approximated according to the following rule:

$$i_t = Z_t'\beta + \varepsilon_t \tag{1.1}$$

where β is a vector of reaction coefficients, Z a matrix of macroeconomic variables, and ε an error term. In its initial specification, Taylor proposed the inclusion of the macroeconomic variables inflation and a measure for economic activity captured in the matrix Z.[1] Quite surprisingly, this standard Taylor rule provided a relatively good fit for the Federal Reserve (Fed)'s and ECB's actual interest

[1] Note that Taylor (1993) assumes a constant real interest rate. However, in most advanced economies there is growing evidence for a decline in real interest rates (see Del Negro, Giannone, et al., 2019). We address this decline by dividing the sample period and presume a constant real interest rate within those subsamples.

Figure 1.1: Comparison of the actual and Taylor rule-approximated ECB interest rate.



rate path (see Gerdesmeier and Roffia, 2003; Sturm and Wollmershäuser, 2008). This fit is illustrated in Figure 1.1, where we compare the actual ECB interest rate and its Taylor rule-based approximations for the period from 1999 until 2018.[2] We measure the ECB's interest rate i by the Euro OverNight Index Average (EONIA) rate (black line) and use two Taylor rule-based interest rate approximations as comparisons: one measured using current inflation and output (red dashed line) and one using expected inflation and expected output (blue dotted line).[3] Figure 1.1 provides anecdotal evidence for the following three conclusions. First, both Taylor rules approximate the short-term interest rate (relatively) well until the outbreak of the financial crisis in 2008. Second, after the outbreak, deviations became more severe, suggesting a more restrictive monetary policy stance than actually conducted. The divergence between the actual interest rate and the Taylor rule-based approximations indicates that a single reaction function based on the standard Taylor rule might not be sufficient to approximate the actual short-term interest rate. Therefore, it seems unlikely that two variables

[2] The corresponding regression table can be found in Section A.1.

[3] Note that conclusions derived from Taylor rules based on actual data (*backward-looking*) or based on expected data (*forward-looking*) can vary and, therefore, provide different policy implications (see e.g. Svensson, 2003; Gerdesmeier and Roffia, 2003). The implications of forward- and backward-looking specifications are further discussed in Section 1.2.1.

are sufficient to properly approximate the short-term interest rate.

Consequently, we argue in line with Milani (2008) that the central bankers in charge do not seem to follow one single model when setting interest rates. Instead, the ECB's interest rate setting can be best approximated by using a variety of models. Considering different models is equivalent to assume that each central banker has a model about the economy in mind and her model is assigned some probability of being chosen to represent the aggregated Governing Council decision.

We use textual analysis of the ECB's communication as an additional instrument to analyze the ECB's monetary policy reaction functions. To be precise, we analyze introductory statements of the ECB's press conferences, the official communication instrument of the central bank to the general public. Using textual analysis on these statements, we provide evidence to include further variables in the reaction function. A word cloud of the introductory statements (see Section A.2) shows that the ECB discusses myriads of different variables and indicators, each, at least implicitly, a potential indication for some reaction by central bankers to the respective variable. Even though the ECB mentions 'inflation' and 'output' in the introductory statements frequently, those two variables do not encompass all of the attention. Also other variables, e.g., related to financial stability and commodity prices, are discussed by the ECB frequently. Therefore, the introductory statements indicate that additional variables might play a role for the ECB.

We do not choose our variables with a view to the Taylor rule, but follow an 'unbiased' approach by using a three-step procedure to identify potential determinants. In a first step, we derive these variables from the monetary policy literature. In a second step, we eliminate variables that were not mentioned in papers analyzing the ECB's monetary policy in a rule-based environment at least once. In a last step, we identify variables that the ECB presidents referred to in their introductory statements. An overview of all potential determinants can be found in Section A.3.

Note that two aspects are beyond the scope of this paper: First, we do not examine whether these additional variables enter the ECB's monetary policy framework on their own or as an instrument, i.e., by affecting other variables such as inflation and economic activity. Therefore, the derived variables from the estimated reaction functions do not necessarily reflect the determinants of the central banks' objective function. As a result, causal interpretation should be regarded with caution. Second, we do not claim to estimate the 'true' ECB reaction function since the Governing Council's monetary policy discussions are not provided to the general public. By using exclusively public information, we analyze the public perception of an ECB monetary policy reaction function rather than a private – ECB internal – reaction function.

1.2.1 Business cycle variables

Inflation and output According to its official mandate, the ECB's primary objective is to maintain price stability (see Article 127 §1 of the Treaty on the Functioning of the European Union). Since 2003, price stability is defined as the medium-term annual growth of the HICP of below but close to two percent (see European Central Bank, 2019a). A second, subordinated, objective of the ECB is to support the general economic policies of the European Union (EU) (see Article 127 §1 and Article 3 of the Treaty on the Functioning of the European Union). This objective is often interpreted as considering economic activity, i.e., output and unemployment, when setting interest rates. Due to their pivotal roles, it is not surprising that the terms 'inflation' and 'output' are used frequently in the communication of the ECB. These terms account for more than 1.3% of all words used in press conference statements, i.e., 'inflation' is mentioned almost 2000 times and 'production' and 'output' 142 times. Therefore, we are confident that inflation and output are relevant for the ECB's reaction functions.

Note that the actual indicators measuring inflation and economic activity are discussed heavily. While various Taylor rules use actual data for inflation and output, Svensson (2003) argues that a standard Taylor rule focusing on backward-looking data is not optimal as those variables affect monetary policy with a lag. Therefore, he suggests to use inflation and output expectations as applied in Sauer and Sturm (2007) and Gerlach (2007). Further, it is ambiguous whether the ECB focuses on HICP inflation, core inflation (see Gerlach, 2007), or commodity prices.

In our analysis, we use both backward-looking and forward-looking inflation and output as well as core inflation, and commodity prices.

Unemployment Article 3 of the Treaty on the Functioning of the EU specifies full employment as one objective of the EU. Therefore, the euro area-wide unemployment rate could be relevant for the ECB's interest rate decision. Furthermore, the state of the labor market can be interpreted as an indicator for economic activity (see Molodtsova and Papell, 2013). Textual analysis supports the importance of unemployment. Terms containing the word 'employment' are mentioned more than three times as often (429 times) as the terms 'production' or 'output'.

1.2.2 Financial markets variables

One of the most prominent extensions of the standard Taylor rule is with respect to the stability of financial markets (see Peek, Rosengren, et al., 2016). We identified various channels through which financial stability could influence the ECB's interest rate setting (see Kaefer, 2014).

Credit measures Credit growth is suspected to impact financial stability via two channels: Firstly, asset bubbles could emerge if credit is massively invested in asset markets. Secondly, credit-financed consumption and investment could lead to a non-sustainable level of debts, which negatively influence economic activity. The term 'credit' is mentioned more than 600 times in the ECB's communication.

Euro exchange rate If a currency appreciates, exporters become less competitive, driving a decline in output and inflation. This appreciation might lead to capital inflows potentially creating asset bubbles. Holders of the appreciated currency experience financial gains, and consumers adjust inflation expectations. The relevance of the exchange rate is emphasized by the ECB mentioning the exchange rate more than 100 times.

Euro area government bond yield A decrease in the government bond yield eases refinancing cost for countries and is expected to work as a stimulus for the respective economy. Furthermore, Roskelley (2016) stresses the forward-looking component of government bond yields. Terms such as 'government debt', 'government deficit', and 'government bond' are mentioned more than 100 times in the ECB press conferences.

Stock prices The inclusion of asset prices, such as stock prices, follows a similar argumentation as the inclusion of the exchange rate.[4] An increase in asset prices can lead to a rise in output and inflation through increased consumption (consumption smoothing) and investments. Thereby, stock prices include relevant information about future inflation. Terms related to stock and asset prices appeared more than 170 times in ECB communication. Additionally, in the current literature, researchers use market volatility as a measurement for financial stability (see Albulescu, Goyeau, et al., 2013; Bleich, Fendel, et al., 2013). Since we analyze a wide information set, we include stock prices as well as volatility measures.

Financial stress indicator The global financial crisis revealed that instabilities in the financial sector can impact financial stability. Hollo, Kremer, et al. (2012) constructed an indicator measuring systemic risk and contemporaneous stability in the financial system, the so-called Composite Index of Systemic Stress (CISS). The index aggregates information beyond the above-mentioned channels about the current instability of the financial system. 'Stress' is mentioned almost 30 times in ECB communication.

Money supply One important component of ECB's policy strategy is the monetary analysis. In the long run, an expansion in the monetary base is expected to drive inflation (see Friedman, 1963; Belke and Klose, 2010). Money supply ('M3') is mentioned more than 500 times in the ECB communication analyzed.

[4] Note that we do not include property prices as property prices typically adjust slowly due to market frictions and infrequent valuations since transactions are conducted seldom. As a result, the response to central banks' monetary policy measures takes place with a considerable lag. Therefore, property prices may reflect long-term interest rate expectations (e.g., forward guidance) and do not solely react to actual short-term monetary policy changes.

1.2.3 Further variables

Economic policy uncertainty Political and economic uncertainty could encourage individuals to postpone investment and consumption decisions. Aastveit, Natvik, et al. (2013) argue that, in the presence of uncertainty, individuals react more cautiously to interest rate decisions. In order to ensure effective monetary policy measures, central banks would have to act more aggressively to achieve their objectives. Furthermore, Philip Lane, the chief economist of the ECB, argues that uncertainty can affect the wage-setting of companies (see Lane, 2019). Due to the relevance of uncertainty (mentioned 280 times), we include Baker, Bloom, et al.'s (2016) economic policy uncertainty index for the euro area.

Trade deficit Obstfeld and Rogoff (2005) and Ferrero, Gertler, et al. (2010) argue that trade flows capture information on future currency movements. Hence, the trade balance can be suspected to be a source of inflationary pressures. The ECB mentions 'trade' almost 70 times.

Interest rate smoothing Note that we do not account for interest rate smoothing as we assume that the central bank monitors a variety of different indicators and variables when setting interest rates. Therefore, the central bank reacts to different economic signals – both positive and negative – that provide mixed interest rate implications indicating a smooth response as positive and negative signals yield an inert interest rate response.[5] Milani (2004) argues that *"the explicit introduction of a wider information set is [...] itself a cause of interest rate smoothness"*.

1.3 Estimation approach

The main objective of this paper is to analyze the ECB's monetary policy under the assumption that central bankers consider a multitude of models containing a wide information set. In this section, we briefly summarize the methodology

[5]This hypothesis is later supported when analyzing BMA-based monetary policy reaction functions.

of Bayesian Model Averaging (BMA).[6] Using a BMA approach allows us to account for model uncertainty and to assess and evaluate every feasible model combination that can be constructed from a predefined dataset based on the potentially relevant variables identified in Section 1.2.

In previous publications, BMA was already applied to account for model uncertainty in economic contexts, e.g., by analyzing determinants of Gross Domestic Product (GDP) growth (see Koop, 2003; Fernandez, Ley, et al., 2001b), and exchange rate crises (see Cuaresma and Slacik, 2009). Milani (2008), Lee, Olekalns, et al. (2013), and Lee, Morley, et al. (2015) were the first to use a BMA approach in the context of monetary policy.

Lee, Olekalns, et al. (2013) estimate a 'meta Taylor rule' – i.e., they characterize monetary policy with reference to the Taylor rule – for the United Kingdom using data from 1972 until 2010 and for Australia from 1972 until 2011. Lee, Morley, et al. (2015) use a similar approach for deriving monetary policy reaction functions for the United States, analyzing data from 1972 until 2008. Both papers apply the BMA approach in the context of monetary policy to account for uncertainties with respect to the duration of monetary policy regimes and the specification of the Taylor rule. While both do not consider variables beyond the standard Taylor rule, i.e., inflation and economic activity, they also address regime uncertainty – an aspect we do not focus on in our analysis.[7] Whereas Lee, Olekalns, et al. (2013) and Lee, Morley, et al. (2015) estimate a three-digit number of models, we evaluate approximately 33,000 different models and, therefore, consider a wider range of different variables and resulting models. A wide range of variables is also used in Milani (2008) making his approach most similar to ours. In his setup, also variables beyond the 'classical' Taylor rule are included as potential determinants – different to Lee, Olekalns, et al. (2013) and Lee, Morley, et al. (2015), but

[6] We refer the interested reader to more comprehensive BMA literature such as Fernandez, Ley, et al. (2001a), Milani (2008), Zeugner and Feldkircher (2015), and Moral-Benito (2015). We mainly follow the notation of Moral-Benito.

[7] Lee, Olekalns, et al. (2013) and Lee, Morley, et al. (2015) specifically analyze the impact of monetary policy regimes, e.g., related to the stability of the policy responses to economic conditions, policy horizons, or the presidents of the respective central bank. In this paper, we differentiate between the period before the global financial crisis (1999–2008) and the period during and after the crisis (2008–2018). Therefore, we consider two different 'regimes', but do neither address the role of different presidents, nor changes independent of the global financial crisis.

in line with our approach. However, Milani (2008) applies BMA for the United States considering data from 1979 only until 2001.

In contrast to Milani (2008), Lee, Olekalns, et al. (2013), and Lee, Morley, et al. (2015), we analyze the ECB's monetary policy of the last two decades and take structural changes around the financial crisis and the emergence of the Zero Lower Bound (ZLB) into account. Furthermore, we integrate elements from textual analysis to allow for more systematic insights. Our paper is – to the best of our knowledge – the first to use BMA in the context of monetary policy around the financial crisis and, generally, the first to analyze ECB's monetary policy using BMA.

We implement the BMA approach in the following way: First, we specify the ECB's short-term interest rate as the dependent variable y. Second, we choose a set of independent variables potentially influencing the dependent variable (Z). The matrix Z consists of all potential regressors that were identified and discussed in Section 1.2. Third, we perform the actual BMA estimation and evaluate all possible linear combinations of these regressors.[8] Uncertainty about the (subjective) choice of the 'true model' vanishes as not only one single model is analyzed, but rather all possible model combinations. For every model, the regression coefficients captured in the vector $\hat{\beta}$ are estimated via Bayesian techniques. As a last step, we compute the average coefficient weighted by the respective likelihood of the model. This vector of average coefficients $\hat{\beta}_{BMA}$ can be expressed as

$$\hat{\beta}_{BMA} = \sum_{j=1}^{2^k} \hat{\beta}^j w^j, \tag{1.2}$$

where $\hat{\beta}^j$ is the coefficient estimate of model j and w^j is the respective weight. In the following, $\hat{\beta}$ is referred to as the posterior probability and w as the posterior model probability.

[8]Note that the integration of interaction terms or other non-linear parameters is still an open research topic in BMA and, therefore, excluded from our analysis.

1.3.1 Bayesian model averaging

The following linear model is considered:

$$y_t = Z_t'\beta + \varepsilon_t \tag{1.3}$$

$$\varepsilon_t \sim N(0, \sigma^2)$$

where y_t is the interest rate and Z_t is a vector of k explanatory variables at time t. β is the k-dimensional vector of regression coefficients and ε_t is a vector of error terms, which follows an univariate normal distribution with zero mean and variance σ^2. With k being the amount of possible regressors, we observe model space $M = \{M_j : j = 1, \ldots, 2^k\}$ and coefficients $\beta = \{\beta^j : j = 1, \ldots, 2^k\}$ with each model j having k individual beta coefficients. By applying Bayes rule, the posterior probability – the distribution of the estimated coefficient vector conditional on one specific model j and the underlying data – can be derived as follows:

$$p(\beta^j|y, M_j) = \frac{p(y|\beta^j, M_j)p(\beta^j|M_j)}{p(y|M_j)}. \tag{1.4}$$

Equation (1.4) states that the posterior probability $p(\beta^j|y, M_j)$ is calculated by multiplying the likelihood $p(y|\beta^j, M_j)$ with the probability $p(\beta^j|M_j)$ divided by the marginal likelihood $p(y|M_j)$.[9] Note that the marginal likelihood is constant over all models and is, therefore, a multiplicative term. Thus, the marginal likelihood is not in the focus of our analysis since it does not depend on β, which we seek to examine in this paper.

The inclusion of own information – so-called priors – is one of the key features of Bayesian modeling. In this way, a researcher defines distributional assumptions, e.g., of a coefficient or a model space, before observing the data. Note that the prior choice expresses subjective beliefs and has to be set with caution. In this paper, $p(\beta^j|M_j)$ is referred to as the prior on the parameter space expressing our belief about the probability distribution of β^j. In Section 1.3.2, different model prior specifications are discussed.

[9]Note that $p(y|M_j) = \sum_{s=1}^{2^k} p(y|\beta^s, M_s)p(\beta^s|M_s)$.

As a next step, we aggregate the posterior distributions over the whole model space using an aggregation weight w^j. The posterior model probability $p(M_j|y)$ is used as a weight since it indicates the degree of support for model M_j. Applying Bayes rule yields this posterior model probability:

$$w^j = p(M_j|y) = \frac{p(y|M_j)p(M_j)}{p(y)} = \frac{p(y|M_j)p(M_j)}{\sum_{s=1}^{2^k} p(y|M_s)p(M_s)}. \tag{1.5}$$

Equation (1.5) states that the posterior model probability – the probability of selecting model j – depends on the marginal likelihood $p(y|M_j)$, the marginal probability $p(M_j)$, and the integrated likelihood $p(y)$. Note that the integrated likelihood does not vary across models and is, therefore, only a multiplicative term. To compute the posterior model probability, a second prior needs to be introduced ($p(M_j)$). This prior specifies the distribution over the model space and expresses our belief about the probability of choosing model M_j before observing the data.

As a next step, a weighted average of all individual posteriors probabilities is computed to obtain one full posterior distribution. As a weight, the posterior model probability from Equation (1.5) is used. Hence, the full posterior distribution can be expressed as follows:

$$p(\beta_{BMA}|y) = \sum_{j=1}^{2^k} p(\beta^j|y, M_j)p(M_j|y). \tag{1.6}$$

This full posterior distribution allows us to analyze coefficients across all models. We examine economic relevance of the included variables by estimating the expected value $E(\beta_{BMA}|y)$ and the variance $V(\beta_{BMA}|y)$ of each coefficient. Both moments can be derived from Equation (1.6) as follows (see Moral-Benito, 2015; Koop, 2003):

$$E(\beta_{BMA}|y) = \sum_{j=1}^{2^k} E(\beta^j|y, M_j)p(M_j|y) \tag{1.7}$$

$$V(\beta_{BMA}|y) = \sum_{j=1}^{2^k} V(\beta^j|y, M_j)p(M_j|y) + [E(\beta^j|y, M_j) - E(\beta^j|y)]^2 p(M_j|y) \tag{1.8}$$

To identify relevant variables, we analyze the Posterior Inclusion Probability

(PIP) and rank the most relevant regressors based on their fit. The PIP for variable h can be computed as follow:

$$PIP_h = p(\beta_h \neq 0|y) = \sum_{\beta_h \neq 0} p(M_j|y). \tag{1.9}$$

A variable with a high PIP indicates that the variable is included in a variety of relevant models and can, thus, be considered robust. We define variables with a PIP > 0.15 as robust.

1.3.2 Definition of priors

In this section, we specify the priors on the parameter space $p(\beta^j|M_j)$ from Equation (1.4) and on the model space $p(M_j)$ from Equation (1.5).

Prior on the parameter space We follow Koop (2003) and specify a normal distribution with zero mean for the distribution of the coefficient to put as less (subjective) information on the distribution as possible before observing the data. For the variance, we apply the g-prior proposed in Zellner (1986), who introduced an additional parameter g into the variance structure.[10] The vector of the estimated coefficients β of model j, therefore, follows the following normal distribution:

$$\beta^j|\sigma^2, M_j, g, X \sim N(0, \sigma^2 g(X_j' X_j)^{-1}) \tag{1.10}$$

In line with the BMA literature, the standard deviation parameter σ is assumed to be equal in all models and is set as an uninformative prior, as proposed in Fernandez, Ley, et al. (2001a) and Zeugner and Feldkircher (2015)[11]:

$$p(\sigma) \propto \frac{1}{\sigma} \tag{1.11}$$

[10]The popularity of the g-prior variance specification is mainly due to the facts that (1) a closed-form solution for the posterior distributions exists reducing computational issues, (2) the variance of the coefficient only depends on the scaling parameter g as σ is equal in all models (see Moral-Benito, 2015), and (3) a penalty for large models is included.

[11]Note that the prior choice p(σ) does not influence the estimation results since σ is equal in all models and, therefore, has the same implications for every model (see Koop, 2003).

Therefore, the expected value of β^j can be expressed as follows (see Zeugner and Feldkircher, 2015):

$$E(\beta^j|M_j, g) = \frac{g}{1+g}\hat{\beta}^j_{OLS} \tag{1.12}$$

Equation (1.12) shows that by using the g-prior, the expected value of the coefficients can be expressed as a convex combination of the Ordinary Least Squares (OLS) estimator $\hat{\beta}^j_{OLS}$ and the prior mean (zero). By specifying g, the researcher indicates how much importance she puts on the prior belief. A small g expresses a high weight of the prior mean. In the case $g \to 0$, the expected value of the coefficient converges to the prior mean (zero). In the case $g \to \infty$, the estimated value approaches the OLS estimator, neglecting the prior completely. This prior structure yields a likelihood $p(y|M_j, X, g)$ that is similar to R^2 and includes a penalty for large models.

In the literature, three different possibilities for the specification of g are considered primarily, namely the Unit Information Prior (UIP) proposed by Kass and Raftery (1995) ($g = t$), the Risk Inflation Criterion (RIC) by Foster and George (1994) ($g = k^2$) and the Benchmark Prior (BRIC) by Fernandez, Ley, et al. (2001a) ($g = max(t, k^2)$). In our estimation, we apply the BRIC prior as it combines the UIP and the RIC by using the maximum g of both candidates. We thereby put as little importance on the prior as possible. In Section 1.6.1, we show that our results are robust to different prior specifications.

Prior on the model space The prior on the model space $p(M_j)$ defines the expectation about the number of regressors the researcher believes to be included in the true model. For example, if the prior is set to five the researcher expects the dependent variable to be most accurately explained by five independent variables. Hence, the researcher can use the prior on the model space to express a preference for smaller or larger models. We assume that the model size Ξ follows a binomial distribution, specified as $\Xi \sim Bin(k, \theta)$, where θ is the prior inclusion probability for each variable. Therefore, model M_j with k regressors has a prior model probability of

$$p(M_j) = \theta^{k_j}(1-\theta)^{k-k_j} \tag{1.13}$$

In the following, we consider two approaches for implementing this prior on the model space, namely through a binomial distribution and a binomial-beta distribution. Both priors have the advantage of being easy to implement but have the disadvantage of neglecting multicollinearity issues, i.e., that the probability that a regressor will be included in a model is observed separately.[12] In other words, the inclusion or elimination of one variable does not change the probability of any other regressors being included. Our approach to address the multicollinearity issue is discussed in Section 2.5.

(a) Binomial (uniform) model prior. Using a binomial approach, the expected model size can be expressed as:

$$E(\Xi) \equiv m = k\theta \tag{1.14}$$

Milani (2008) sets $\theta = 1/2$, allocating the highest probability (and the expected value) to models that contain $k/2$ variables. An alternative is to set $\theta < 1/2$. This specification increases the likelihood of smaller models and could take limitations in human cognitive abilities into account. In other words, it can be suspected that central bankers tend to consider primarily small models. We implement several specifications for θ accounting for a lower probability of large models. Note that our results indicate a tendency towards small models, independent of the specification of θ.

(b) Binomial-beta prior. Ley and Steel (2009) suggest the use of a hyperprior on the inclusion probability θ. This specification makes θ random, in contrast to the binomial distribution, where θ is fixed. Ley and Steel suggest to use a beta distribution for the hyperprior, i.e., $\theta \sim Beta(a, b)$. The use of such a binomial-

[12]Multicollinearity can lead to outcomes that highly correlated variables are included in the models causing biased estimation results.

beta prior leads to an expected model size of

$$E(\Xi) \equiv m = \frac{a}{a+b}k \tag{1.15}$$

Ley and Steel (2009) propose $a = 1$ and $b = (k - m)/m$ expecting the researcher to specify – similar to the binomial prior – only the expected model size m. Choosing $m = k/2$ leads to $a = b = 1$, which yields the following (flat) model size probability distribution:

$$p(M_j) = \frac{1}{k} \tag{1.16}$$

Equation (1.16) indicates that the selected binomial-beta distribution results in a posterior probability that is equal for each model size. Thereby, it reduces the subjective influence concerning the expected model size which minimizes the impact of the prior choice by the researcher. While the probability distribution is centered around the expected model size for a binomial distribution, in the case of a binomial-beta distribution, the probability distribution of the posterior is flat for all models.

1.4 Data

In this section, we discuss the data used in our BMA analysis. Note that we focus on real-time data, which allows a more comprehensive understanding of the perceived central bankers' reaction to macroeconomic data. Orphanides (2001) provides evidence that monetary policy implications based on revised (not real-time) data are inaccurate as the data used by policy-makers and researchers does not align. He argues that revisions constitute new information about previous data points, not available to policy-makers at that point in time. Since the ECB acknowledges that macroeconomic variables were subject to significant revisions (see European Central Bank, 2010), it is essential to use real-time information. We obtain such real-time data from the ECB's own Real Time Database (RTD). Due to the following three factors, we argue that the RTD represents an accurate information set available to ECB central bankers. First, using the RTD allows to merge low frequency macroeconomic data such as GDP or inflation with higher granular financial data such as asset prices or exchange rates. Second, it seems reasonable to assume that policy-makers primarily consider own information. Third, the RTD contains the latest macroeconomic data, relevant for the central bankers, as revisions are usually published on the day before the monetary policy decisions.

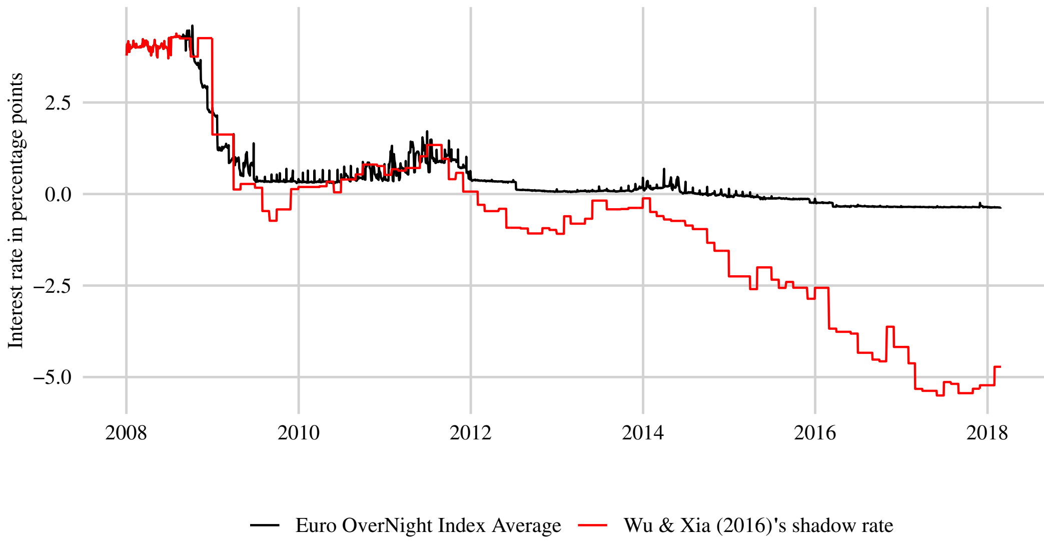
However, the RTD does not contain data for all the variables discussed in the previous sections. In particular, expectation data of macroeconomic variables, such as inflation, economic activity, and unemployment, are not included. We obtain expectation data from the ECB's Survey of Professional Forecasters (SPF). The remaining data is obtained from various sources (see Section A.3).[13]

The interest rate set by the ECB's Governing Council is the Main Refinancing Operations (MRO) rate. However, the presence of the Zero Lower Bound (ZLB) and the ECB's unconventional monetary policy measures do not favor the MRO as an appropriate representation of the short-term interest rate. Instead, we use two alternatives: First, we consider the EONIA rate. The EONIA rate is applied in Taylor rules, e.g., in Fendel and Frenkel (2006) and Castro (2011). Second, we consider the Wu and Xia shadow rate for the euro area (see Wu and Xia, 2016) since it captures additional information from unconventional monetary policy measures such as forward guidance or asset purchase programs.

Figure 1.2 highlights only minor deviations between the EONIA and the shadow rate from 2008 until 2011. However, the divergence between the two rates increased from 2011 onwards – with the EONIA being restricted by the ZLB – and reached a gap of more than five percentage points from 2017 onwards. The shadow rate (s) is applied as the dependent variable for the period after the outbreak of the global financial crisis when unconventional monetary policy measures were introduced. We set the date for the outbreak of the financial crisis to September 15th, 2008. On that day, the bankruptcy of Lehman Brothers was officially declared. Therefore, the short-term interest rate y is comprised of the

[13] For most data, the only transformation was to compute annual growth rates to reduce stationarity. One exception is the output gap. We calculated the output gap by applying the Hodrick-Prescott (HP) filter on real-time data for real GDP. Note that this is different from the method applied in Taylor (1993). Instead of real-time data, he uses an end-of-sample measure and constructs his data of the output gap from the deviation of actual real GDP from a constant trend real GDP.

Figure 1.2: Comparison of the shadow rate by Wu and Xia (2016) and the EONIA rate.



EONIA rate i and the shadow rate s, i.e.:

$$y_t = \begin{bmatrix} i_t \\ s_t \end{bmatrix} \begin{matrix} \text{if } t < \text{September 2008} \\ \text{if } t \geq \text{September 2008} \end{matrix} \tag{1.17}$$

1.5 Results

In the following section, we discuss the main results of our BMA analysis. We apply the BRIC g-prior on the parameter space and a flat binomial-beta distribution on the model space. These specifications are chosen in order to put as little (subjective) information on the priors as possible. As described in Section 1.4, the EONIA rate and the Wu and Xia shadow rate have been combined as the dependent variable. As we discuss in the next chapter, our main findings are robust across different prior specifications. We run the regressions in R, using the BMS package by Zeugner and Feldkircher (2015). To account for multicollinearity, we use Cuaresma and Slacik's (2009) approach and neglect all regressors with a correlation[14] higher than $|0.6|$.[15] In total, we estimated $2^k = 2^{15} \approx 33{,}000$ different

[14] We recognize the limitation of using bivariate correlations as a representation for multivariate relationships. For further discussions, see Hayo (2018).

[15] Actual unemployment, expected inflation, and money growth were excluded from the regression due to strong correlation. We have conducted a range of robustness checks to see if

Table 1.1: Regression results

	1999 – 2018		*1999 – 2008*		*2008 – 2018*	
	PIP	Post Mean	PIP	Post Mean	PIP	Post Mean
HICP Inflation Rate	1.000	1.095 (0.169)	0.736	0.313 (0.216)	0.941	0.776 (0.273)
Unemployment (expected)	1.000	-0.666 (0.111)	0.239	-0.068 (0.136)		
Output Gap (expected)			0.986	1.708 (0.438)		
Output Gap (actual)			0.967	0.500 (0.179)		
Observations		216		114		102

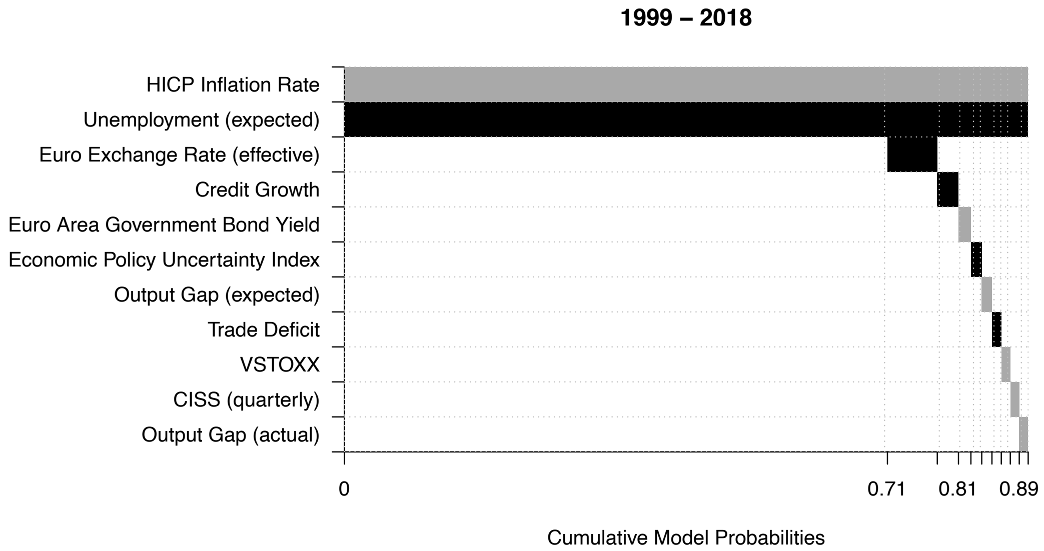
Note: Only robust variables with a PIP ≥ 0.15 are presented.

models for the time period from April 1999 until March 2018 aggregating the respective model-specific estimation results to obtain one average effect for every regressor.

1999–2018 The main results of the regression are shown in Table 1.1. The first column shows the regressor, the second the PIP – the aggregated probability of the models including the respective variable –, and the third the post mean – the average marginal effect – with the standard deviations denoted in brackets.
The main results for the period from 1999 until 2018 are the following: First, the HICP inflation rate is included in all relevant models, indicated by a PIP of 100%. Therefore, our results suggest that the current inflation rate is indeed primarily considered when the ECB sets interest rates. The central bank reacts to an 1 percentage point increase in the annual inflation rate by increasing the interest rate on average by 1.1%. However, the inflation coefficient's standard error equals 0.17. Thus, the Taylor principle cannot be confirmed with reasonable statistical significance. Second, expected unemployment has a PIP of 100% as well. The coefficient is negative, indicating that the central bank reacts to an increase in the expected unemployment rate with expansionary monetary policy. Third, no further variables are considered in the majority of the models, i.e., no variable

our choice has altered the outcomes. The primary findings appear to be selection independent.

Figure 1.3: Ten Top Models (whole period).



has a PIP > 15%.

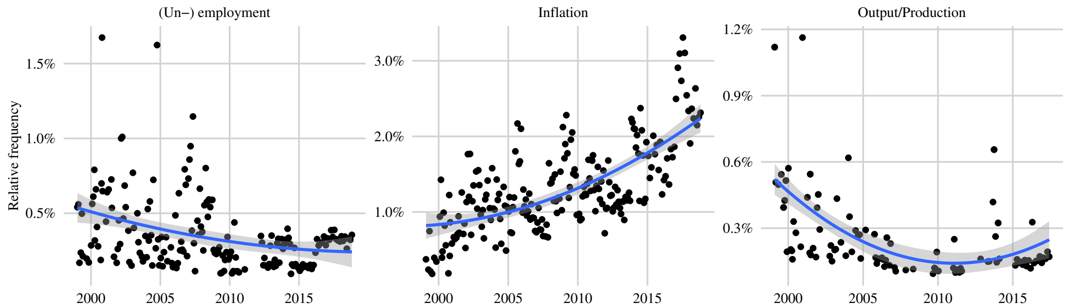
Next, we consider individual models. Figure 1.3 provides an overview of the ten 'best-performing' models ranked by their respective model inclusion probabilities. The higher the model inclusion probability, the higher the likelihood that this specific model represents the 'true' model. The cumulative model probabilities – an aggregated likelihood of models considered – are shown on the horizontal axis and the relevant regressors on the vertical axis, i.e., the variables that are included in the models with the highest likelihood. Gray color indicates a positive sign of the respective coefficient, while black color indicates a negative sign. The model that can explain the data in the most precise manner (the 'top model') has a model probability of 71%, i.e., the best individual model out of 2^{15} models has a likelihood of 71%. This specific model includes only the two (robust) variables, namely inflation and expected unemployment with the expected signs. The model with the second highest likelihood additionally includes the effective Euro exchange rate that yields a likelihood of approximately 7%. The next best model with the inflation rate, unemployment, and credit growth has a model likelihood of approximately 3%. Thus, the figure indicates that model probabilities are not evenly distributed across a large number of heterogeneous models. Restricting the analysis to one single model – a common approach in classical model-selected

Taylor rules – would neglect relevant model probabilities. Therefore, using our ten top models instead of a two variable Taylor rule increases the cumulative model probability from 71% to almost 90%.

1999–2008 A crucial question is whether the ECB altered its monetary policy strategy after the financial crisis. One approach to detect a potential systemic change in the perceived ECB reaction function is to observe the communication of the central bank. Figure 1.4 illustrates the relative frequency of terms related to inflation and economic activity mentioned in ECB press conferences over time. Analyzing the relative frequency, we find evidence for a gradual shift from unemployment and output towards inflation. To account for this potential shift, we separate the time periods during and after the crisis explicitly from the period before the financial crisis.[16] The main results for the pre-crisis period are displayed in Table 1.1 in columns four and five: First, the output gap seems to be the main determinant of the ECB's interest rate. Both the expected output gap – based on expectations – and the actual output gap – based on previous data – are robust and significant. Both variables are included in almost all models with the expected sign: An increase in the output gap leads, on average, to an increase in the interest rate. The coefficient for the expected output gap (1.7) is higher than the coefficient for the actual output gap (0.5). This reaction is in line with the ECB's official objective to support economic activity in the euro area. Quite interestingly, the coefficient for the actual output gap equals Taylor's initial calibration from 1993. However, Taylor includes only one measure of economic activity – a key difference to our empirical results. Note that the focus on the output gap is in line with the communication of the ECB. As discussed previously, the fraction of output-related terms decreases over time (see Figure 1.4).
Second, the HICP inflation rate is again robust (PIP = 0.74), although with a lower coefficient. Note that for the pre-crisis period the Taylor principle can be significantly rejected. Third, the expected unemployment rate is included in a minority of the models (PIP = 0.24) with the expected sign. Unemployment is only included in two out of the ten top models (see Figure A.2). Therefore, our

[16] Note that we use the term 'post-crisis' for the time period after the financial crisis although it actually includes both the period of the crisis itself as well as the post-crisis period.

Figure 1.4: Relative Frequency of Selected Terms in ECB Communication.



results suggest that some central bankers – prior to 2008 – saw unemployment as an important determinant to consider in the context of monetary policy but the majority favored the output gap as a measure for economic activity. Note that, overall, the unemployment rate coefficient is not statistically significant. Fourth, further indicators, such as the exchange rate or stock market prices, only enter the minority of the models (PIP ≤ 0.15) and are, thus, not considered robust.

To summarize, our results suggest that the main determinants of the ECB's monetary policy for the period before the financial crisis are related to the business cycle. These findings are in line with the official mandate of the ECB to account for both the development of inflation and economic activity. However, the Treaty on the Functioning of the European Union specifies that the two objectives are hierarchical in a sense that the ECB mainly focuses on inflation and only subordinately on the business cycle. Our result do not confirm this hierarchy. We find a stronger focus on economic activity than on inflation, indicated by a higher posterior mean, a higher PIP, and that the Taylor principle cannot be confirmed. Focusing on individual models (see Figure A.2), the top model includes the (actual and expected) output gap and the HICP inflation rate (PIP = 31%). The ten top models yield a cumulative model probability of 62%. Therefore, the results provide again evidence to favor averaging techniques when analyzing the monetary policy of the ECB.

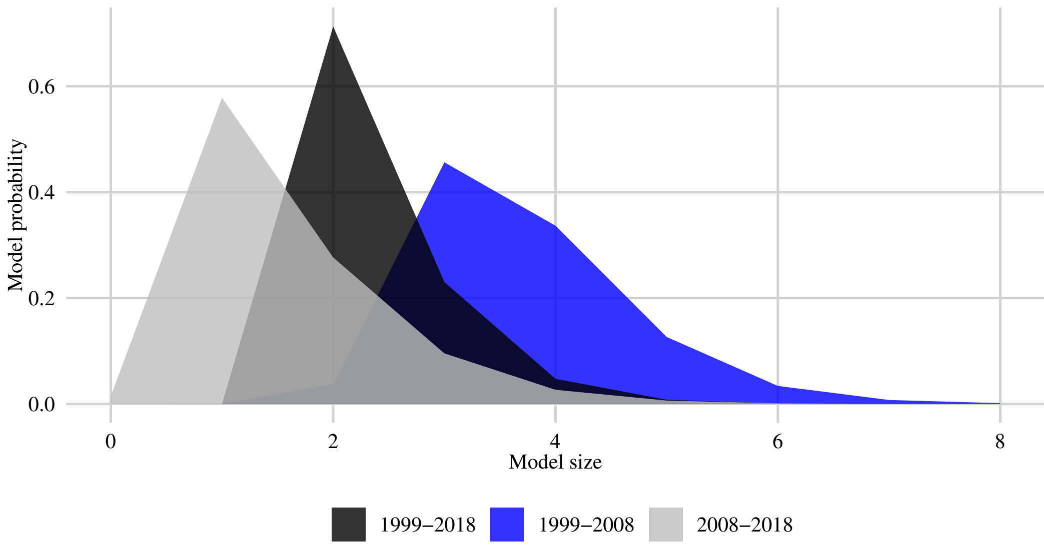
2008–2018 Next, we analyze the ECB's monetary policy in a post-crisis context. The main results for the post-crisis period are displayed in Table 1.1 in columns six and seven. The findings suggest a shift in the ECB's monetary pol-

icy strategy. On the one hand, inflation stays highly robust. In the post-crisis period, the PIP is even higher and indicates that HICP inflation is included in almost all models. The sign of the coefficient is again positive and the magnitude of the coefficient is higher compared to the pre-crisis period. Therefore, our results suggest that the ECB shifts its focus towards the inflation rate after the bankruptcy of Lehman Brothers. For this period, we can neither reject nor confirm the Taylor principle. On the other hand, none of the economic activity measures – actual and expected output gap, and unemployment – are robust for the post-crisis period (PIP ≤ 0.15).
The results suggest that the focus of the ECB shifted from considering both inflation and economic activity in the pre-crisis period to solely considering the HICP inflation rate in the post-crisis period. This finding is supported by evidence from the ECB's communication, showing that, in the post-crisis period, the ECB mentions terms related to inflation more often and terms related to economic activity less often. The ten top models have a cumulative model probability of 80%. The top model only includes HICP inflation and yields a likelihood of 55%.

Short summary To sum up all periods, our results cannot confirm that the ECB does not account for the inflation rate. Both in normal times and in times of crisis, inflation is included in the majority of the relevant models. Furthermore, we find evidence that the ECB reacts to the expected unemployment rate and the output gap. To be precise, our results suggest that the ECB seems to focus its monetary policy decisions after Lehman bankruptcy mainly on the HICP inflation rate, while before the financial crisis both inflation and economic activity measures seem to be relevant. However, the size of the inflation coefficients provide evidence that the Taylor principle might not be fulfilled.

Model size Next, we draw inferences on the number of included regressors in the interest rate setting by evaluating the distributions of the posterior model sizes. Figure 1.5 plots the posterior model sizes for the different time periods. The black density plot refers to the period from 1999 until 2018 (whole period), the blue one from 1999 until 2008 (pre-crisis), and the gray one from 2008 until

Figure 1.5: Model Sizes of the Different Time Horizons.



2018 (post-crisis).

The figure indicates that the distributions of the model sizes vary slightly with respect to their means. Depending on the specification, the average number of determinants in the monetary policy reaction functions are 2.4 (whole period), 3.7 (pre-crisis), and 1.6 (post-crisis). Hence, these results correspond to our previous findings in Table 1.1, i.e., that the ECB seems to consider fewer variables after the financial crisis when setting interest rates. However, independently of the time period considered, large models with more than five regressors seem unlikely. Hence, the ECB's monetary policy is best approximated using medium-sized models with between one and five variables.

Model fit Next, we analyze the model fit of the ten top models via an in sample prediction. The main results are presented in Table 1.2 and can be summarized as follows: First, an improvement in explanatory power for the whole period can be reached by using BMA. If we compare the BMA-derived ten top models with the two standard Taylor rules (see Section 1.2), the goodness of fit improves according to the Mean Squared Error (MSE) between 28% and 33%. This improvement could potentially be caused by overfitting. However, it is unlikely that overfitting is the sole driver of the better fit since we demonstrated that the de-

Table 1.2: Model Fit Evaluation

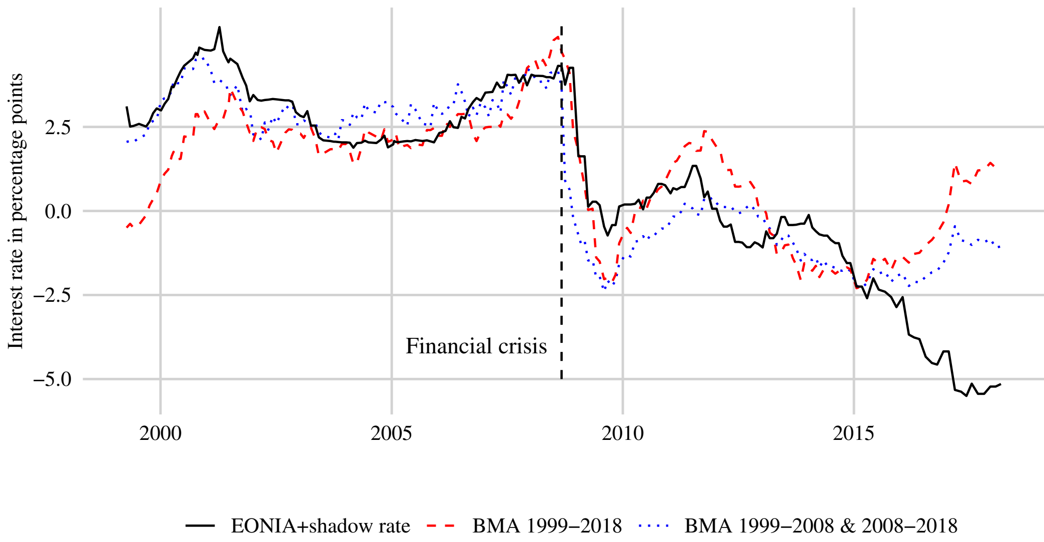
	MSE	AIC	BIC
Taylor rule (backward looking)	5.30	376.72	390.31
Taylor rule (forward looking)	4.95	361.38	374.97
BMA 1999 – 2018	3.55	302.16	339.54
BMA 1999 – 2008 & 2008 – 2018	1.89	163.16	200.54

rived ten top models include only a few variables. To further examine overfitting, we evaluate our results using the Akaike information criterion (AIC) and the Bayesian Information Criterion (BIC) as alternative measures for goodness of fit. For both criteria, we find similar improvements.[17] Therefore, we argue that the improvement in fit is predominantly driven by increased information provided by the additional regressors.

Second, the separation between pre- and post-crisis periods leads to a further substantial improvement in fit as indicated in Table 1.2, with a decrease in the MSE by more than 60%. Moreover, we find similar improvements when controlling for the higher amount of variables in both the AIC and BIC. Figure 1.6 compares the BMA-derived interest rate approximations with the actual short-term interest rate captured by the EONIA and the shadow rate. Until 2015, both BMA approximations provide similar results with little variation. Afterwards, the two BMA approximations diverge substantially. The increase in the BMA-derived interest rate for the whole period (red dashed line) suggests more restrictive monetary policy measures in that period. The specification separating the whole period in the pre- and post-crisis period (blue dotted line) supports the expansionary monetary policy stance to a higher degree. The separated specification advocates an interest rate higher than the shadow rate, but still below the ZLB. At the end of our observation period, the gap between our two BMA approximations is more than 2%. Note that inflation as the main determinant in the post-crisis period has steadily increased since the end of 2015 leading to a divergence from the shadow rate.

[17] AIC and BIC are more appropriate criteria compared to the MSE. AIC and BIC introduce a trade-off between overfitting and underfitting by including a penalty increasing with the number of regressors. To analyze the robustness of our results, we conservatively estimate both measurements with the maximum number of possible variables instead of the actual maximum model size.

Figure 1.6: Comparison of BMA-derived Approximation vs. Actual Interest Rate.



1.6 Robustness checks

In this section, we discuss the robustness of our results. Robustness checks are conducted with respect to (1) prior modifications, (2) a varying date of the beginning of the financial crisis and (3) the use of a different dependent variable. We conclude that our results are robust across a multitude of model specifications.

1.6.1 Priors

In a first robustness check, we modify the priors. As mentioned in Section 1.3.1, priors are chosen subjectively by the researcher. Therefore, it is essential to conduct robustness checks to different prior specifications.

Prior on the model space In the benchmark case, we apply a flat binomial-beta model prior. However, as described previously, one could assume that central bankers favor small models. Therefore, as a robustness check, we apply a different binomial-beta prior, thereby putting a higher weight on small models ($m = 2$). Results are illustrated in Section A.5. The findings are similar to the benchmark case. If the whole time period is considered, inflation and the expected unemployment rate are again the only robust variables and the respective posterior means are almost identical. When separating the pre-crisis and post-crisis period, the findings with respect to the PIPs and coefficient estimates do not vary substantially, as well. To summarize, if a higher weight is put on smaller model sizes, the key findings remain the same.

Prior on the parameter space Next, we alter the prior on the parameter space. In the benchmark case, we apply the BRIC g-prior. Note that this prior combines the RIC and the UIP. As long as $k^2 > N$, the BRIC equals the RIC. In our case ($k = 15$ and $N = 216$), the RIC dominates the UIP. Hence, applying the RIC rather than the BRIC does not alter the outcomes. The results for a robustness check using the UIP prior are shown in Section A.5. The findings are (again) almost identical with respect to the PIP and the posterior means compared to the benchmark case. To summarize, applying different priors on the parameter space leads to similar results.

1.6.2 Starting date of the crisis

In our benchmark case, the beginning of the financial crisis is set to the day of Lehman's bankruptcy, where also macroeconomic data indicated an upcoming recession. However, one can also date the beginning of the financial crisis 13 months earlier, precisely on August 9, 2007. On this day, the inter-banking market in the euro area broke down and the ECB provided additional funding for banks (see Stark, 2010). We, therefore, perform another robustness check altering the starting date of the financial crisis. Results are shown in Section A.5. In the pre-crisis period, the robust variables are almost identical to the benchmark case and most of the PIPs and post means are similar, with two exceptions: Firstly, the expected unemployment rate is not robust anymore. Secondly, the inflation coefficient decreases and becomes insignificant. For the post-crisis period, the PIP of inflation increases from 0.94 to 1.00 and the coefficient increases almost twofold to 1.22. To summarize, redefining the starting date of the crisis pronounces our main result for the post-crisis period even more, namely that inflation becomes the most relevant determinant in the post-crisis period as the PIP and the coefficient size increases.

1.6.3 Endogenous variable

In this robustness check, we substitute Wu and Xia's (2016) shadow rate with the EONIA rate in the post-crisis period, i.e., we use the EONIA rate for the whole period from 1999–2018 as our endogenous variable. The findings of the robustness check are shown in Section A.5. For the whole period, the results are qualitatively similar. However, the coefficients are – due to the lower variance in the endogenous variable – smaller. Note that the results for the pre-crisis period remain the same. For the post-crisis period, the findings from the benchmark case seem to be confirmed qualitatively, as inflation still appears to be a main determinant of the ECB's monetary policy. Besides the inflation rate, the expected unemployment rate is robust and significant. Few other variables tend to be robust – the output gap, commodity prices and core inflation –, but are neither economically relevant nor statistically significant. In summary, our benchmark results are robust and qualitatively independent of the selection of the endogenous variable.

1.7 Conclusion

Over the last decade, the standard Taylor rule, using inflation and economic activity to approximate the ECB's short-term interest rate, has lost substantial explanatory power. This divergence might indicate that central bankers consider other variables beyond those suggested in Taylor (1993) and employ different models when setting interest rates. In this paper, we mainly attribute this divergence to model uncertainty. We analyze a wide arrow of potential determinants that we derive from the literature and textual analysis of the ECB's press conference statements. Using a BMA approach enables us to assess and evaluate every feasible model combination constructed from the variables identified. This approach has – to the best of our knowledge – not been applied previously in the context of the ECB's monetary policy and, therefore, addresses a gap in the current literature. Our derived reaction functions aim to provide clarity for understanding the ECB's monetary policy and allow both researchers and the general public to draw conclusions about potential determinants of the ECB's monetary policy. By separating pre-crisis and post-crisis periods, we account for a potential shift in the central bank's monetary policy strategy.

Our key findings are the following: First, our results indicate that, irrespective of the period analyzed, the inflation rate is a robust determinant. However, our analysis suggests that the inflation coefficient is not statistically significantly different from one. Therefore, we do not find evidence that the Taylor principle is fulfilled. Second, the importance of inflation in terms of robustness and coefficient magnitude increased over time. In fact, for the last decade, inflation appears to be the only robust determinant. This result seems to be in accordance with the communication of the ECB. Third, we find that the robustness of the output gap has decreased over the last decade. Fourth, small single-digit models, including between two and seven variables, approximate the ECB's monetary policy most precisely. Fifth, the distribution of model probabilities shows that no single model can sufficiently explain the observed data. This finding reaffirms using model averaging methods when evaluating monetary policy of central banks rather than selecting one single – e.g., standard Taylor rule-based – model. Nonetheless, we can explain most of the variation in the interest rate by analyzing only ten models.

This paper provides a first analysis of applying model averaging techniques for the ECB interest rate setting. Future research could extend the model averaging approach as follows: First, dilution priors could be incorporated into BMA applications as an alternative approach to account for multicollinearity issues. Second, whereas BMA has already been applied to monetary policy of other central banks (e.g. Lee, Morley, et al., 2015), our approach incorporates novel techniques – such as textual analysis and analyzing a broader range of variables – and enables the determination and comparison of monetary policy reaction functions of different central banks. Applying this approach to the interest rate setting of other central banks, such as the Fed or the Bank of England, could provide an interesting comparison of similarities and differences of monetary policy strategies of global central banks. Third, future research could put a higher focus on accounting for heterogeneity, i.e., by considering macroeconomic developments of specific euro area countries and not euro area aggregates. Finally, in our analysis, we neglect regime uncertainty in the spirit of Lee, Morley, et al. (2015). Accounting for different regimes may add another dimension to our understanding of how central banks conduct monetary policy.

2 Above, but close to two percent. Evidence on the ECB's inflation target using text mining

Johannes Zahner[b]

Abstract

Due to its official mandate, the European Central Bank (ECB) is assumed to maximize an implied objective function that leads it to pursue inflation with a subordinate focus on economic activity. This objective is – by its very nature – difficult to quantify. My paper tries to decipher information regarding the ECB's objective through the use of text mining on all public speeches between 2002 and 2020. The findings of my analysis suggest a concave objective regarding the inflation rate. The implied inflation target is best summarized as an inflation rate of *'above, but close to 2%'*. Deviations from this target lead to a reduction in the optimism of the institutions' communication. Furthermore, my results suggest a convex objective towards output growth and a linear objective towards the unemployment rate, with a preference for higher GDP growth and employment independently of the current level. The hierarchical order in the ECB's mandate does not always appear to be consistent with my findings. Deviations from its primary objective, the inflation rate, appear to be of no greater concern than deviations in its subordinate objective. Finally, in periods of heightened uncertainty, there is an additional decrease in sentiment.

Keywords: Sentiment Analysis, ECB, Monetary Policy, Public Perception.

JEL classification: E53, E58, E61.

[b]Corresponding author. School of Business and Economics, Institutional Economics Research Group, Philipps-Universität Marburg, Germany.

> *"We communicate what we are trying to achieve, and how we go about achieving it. In practical terms, this means that communication revolves around providing a narrative about the economy and the outlook for price stability relative to our objective."*
>
> — Peter Praet (2014)

2.1 Introduction

Over the last decades, the relevance of narratives and perception in explaining macroeconomic phenomena gains importance in the economic discourse (Shiller, 2017). Monetary policy is thereby no exception. As stated in the quote at the beginning of this paper, the ECB recognizes that the role of monetary policy is not limited to actions, but that words are a powerful tools for anchoring expectations and self-enforcing a desired equilibrium path. Since the effectiveness of monetary policy depends on the belief of the general public in the inflation target (Diron and Mojon, 2005), the central bank invests enormous efforts to shape expectations around this institution. This effort is related to the ECB's success in maintaining its target and mandate of price stability.

However, in the current environment of persistently low inflation, the potential of abandoning the asymmetry in the current target (Blot, Creel, et al., 2019) or the 2% reference and raising the inflation target in an effort to anchor inflation expectations at a higher level are frequently discussed (e.g. Blanchard, Dell'Ariccia, et al., 2010; Ball, 2014). The ECB's internal evaluations of its monetary policy strategy reveal that the central bank is not indifferent to these debates. In May 2003, the central bank clarified its inflation objective by adding the component *"close to [2% inflation]"* to its existing definition of price stability, and in June 2021, it moved to a *"symmetric 2% inflation target"* (European Central Bank, 2021, p. 7).

However, there may remain ambiguity about the central bank's objective function. According to Miles, Panizza, et al. (2017), the euro area's low inflation expectations may be due to the ECB's target's *double asymmetry*: On the one hand, because of the de facto inflation ceiling before 2021 and the resulting unlikely anchoring of inflation around 2%, and on the other hand, because of the vagueness surrounding what the term "close to" actually means. Furthermore, while it is widely assumed that the function has some relationship towards inflation and economic activity, specific quantitative measurements of the function's shape are unknown. In this paper, I attempt to shed light on the history of the ECB's objective functions by building on the work of Shapiro and Wilson (2019) in approximating it using semantic information from the central bank's public communication.

The ECB asserts that in recent years, communication of its monetary policy has become increasingly more important, both to experts and to the general public (European Central Bank, 2021). The central bank has various communication channels that target such a broader audience. Alongside its blog, podcast, and 'ECB Listens'-events, the primary means of ECB communication to a wider audience are still public speeches. Consequently, over the last two decades, members of the ECB Executive Board held more than 2,000 speeches, and thereby provide an abundance of qualitative information of central bankers elucidating their objectives to the general public. The aim of this paper is to quantify this qualitative data through the use of text mining, which allows for the approximation of otherwise intangible objectives.

Taking advantage of the heterogeneity in economic conditions – i.e. inflation and economic activity – allows me to approximate the communicated objective function of the ECB. Ultimately, I will be able to identify the respective bliss point towards the variables in this function. This means that speeches become more pessimistic (optimistic), above and below the maximum (minimum) – implying an articulated target. My findings suggest that the ECB communicates an inflation target that is *"above, but close to 2%"*. I find further evidence that the subordinate objective of promoting the general economic policy of the EU is best aligned with a relationship where the central bank favors better economic conditions, unconditionally of the present level. Furthermore, my findings suggest that speeches respond equally strongly to changes in economic activity as to changes in the inflation rate. Moreover, speeches become more pessimistic over time, even when controlling for macroeconomic and financial variables.

The contribution of this work is to be found in the areas of text mining, monetary policy, and public perception. Using text mining techniques, this paper contributes to a better understanding of the perception of monetary policy in the euro area. It is, to my best knowledge, the first work on the monetary policy objective function, using text analysis on speeches the euro area.
This paper is structured in the following order. The second section provides a brief literature overview of text mining applications in the central bank communication literature. Next, the empirical approach to estimate the latent inflation target is introduced. The text-mining approach used to derive the endogenous variable, as well as the underlying datasets for the exogenous variables, are discussed in depth in the fourth section. The results and their implications are addressed in the fifth section before the final section concludes this paper.

2.2 Literature review

The relevance of communication as a monetary policy instrument has increased substantially in central banking. Blinder, Ehrmann, et al. (2008) identify four rationalizations that require communication by a central bank, including the degradation of asymmetric information between the general public and the central bankers. The content of this asymmetric information could originate from a variety of sources. It may regard the outcome of previous policy votes (Meade, 2005), the content of committee members' deliberations (Hansen, McMahon, and Prat, 2018), or the central bankers' risk balance assessment (Hanson and Stein, 2015). As a result, there is a myriad of applications where tone variations in monetary policy communication are used to evaluate market responses.[18] Reactions to changes in sentiment were found by asset price markets (Schmeling and Wagner, 2019), sovereign yield spreads (Falagiarda and Reitz, 2015), and short-term interest futures (Rosa and Verga, 2008). Furthermore, improvements in the predictability of future monetary policy decisions, using sentiment analysis (Apel, Blix Grimaldi, et al., 2019; Baranowski, Bennani, et al., 2021), have been found.

[18]The terms 'tone' and 'sentiment' are used substitutable throughout this paper. In Section 2.4.1, sentiment will be defined as a measure of optimism in one's language.

The number of applications using of text mining in analyzing central bank communication has increased substantially thanks to the rise in computational power and the abundance of available text. Inherent in those papers – as in mine – is the assumption that market participants believe a central bank obeys a coherent systemic approach that allows for the inference with respect to its parameters. However, while earlier text mining literature assumed central bankers would communicate coherently and and observe market reactions to changes in communication, my analysis concentrates on changes in communication in response to variation in economic conditions. To be precise, my paper concentrates on the institutionalized objective function underlying the ECB and thus the communicated objectives towards inflation and economic activity that it implies.

The closest papers in this respect are by Shapiro and Wilson (2019), Paloviita, Haavio, et al. (2020), and Fraccaroli, Giovannini, et al. (2020). Shapiro and Wilson provide the econometric model – shown in the next chapter – applied in this paper. The authors use internal meeting transcripts of FOMC meetings to estimate the central banks' inflation objective using sentiment analysis. The FOMC had no explicit inflation target before 2012. Nevertheless, it has long been assumed to be close to ECB's. However, Shapiro and Wilson (2019) find an implicit inflation objective of 1.5% for the FOMC, substantially lower than the expected 2%. Their estimated inflation target prevails at 1.5%, even when considering public speeches and a dual mandate. Finally, the authors find that the central bankers' sentiment is heavily affected by output growth – rather than changes in the unemployment rate – and stock market performance. However, they remain agnostic whether this result is driven by the perceived predictive power of financial markets or by central bankers' reactions to financial variables in general.

The main difference between Shapiro and Wilson (2019) and my paper is the underlying text corpus, the resulting magnitude of the implied inflation coefficient as well as the factors influencing the sentiment of central bankers. Whereas the former focuses on transcript of FOMC meetings, my analysis uses public information from the Executive Board members of its European counterpart. Their estimated inflation target prevails at 1.5%, even when considering public speeches and a dual mandate – the most analogous regression to mine. Contrary, my findings suggest that the ECB – which had an explicit inflation objective since its beginning – maintains a communicated inflation target above 2%. Moreover, my results suggest that sentiment of the ECB Executive Board is primarily driven by economic conditions (the unemployment rate *and* output gap), whereas financial variables do not have any significant affect.

Paloviita, Haavio, et al. (2020) replicate the results of Shapiro and Wilson (2019) employing a different empirical approach[19] for the euro area relying on the ECB's introductory statements. The authors find an implicit 1.7% inflation target. However, when allowing for a flexible inflation target and including the output gap, the implicit inflation target rises to 2.3%. My paper differs primarily in the methodology, as well as – again – in the underlying text corpus and the magnitude of the implied inflation target. In Section 2.4.1, I demonstrate that, in contrast to speeches, the use of these introductory statements has the disadvantage of increasing similarity over time, which may explain the discrepancy in results. In terms of magnitude, my results consistently indicate a target that is significantly higher than 2%.

Finally, Fraccaroli, Giovannini, et al. (2020) employs sentiment analysis and topic modeling in ECB, FOMC, and Bank of England (BoE) parliamentary hearings. They find that the respective central banks' objectives strongly influence the topics discussed, and that the current unemployment rate, in particular, is a strong predictor of sentiment during those hearings.

2.3 Econometric model

This paper aims to empirically identify the latent inflation target that is communicated by the ECB by the use of sentiment analysis. It thereby follows the econometric approach for identification presented by Shapiro and Wilson (2019). In short: the method approximates the central bank's loss by a semantic index. Assuming a traditional New-Keynesian loss function, allows me then to estimate the central banks' implied inflation target using current inflation data.

[19] The authors estimate a symmetric piecewise linear loss function and find the implied inflation target via a grid search.

Following the textbook literature (e.g. Gali, 2015; Walsh, 2017), a central bank is modelled to minimize a loss function. Assuming that its members share one function, the utility loss L_t at time t can be described as

$$L_t = \hat{\pi}_t^2 + \phi \hat{y}_t \tag{2.1}$$

where $\hat{\pi}_t$ ($\hat{y}_t$) represent deviations in inflation (economic activity) and ϕ the relative weight on economic activity. Note that both, the inflation gap and the gap in economic activity, are latent variables, i.e. can not be directly observed. They are defined as the difference of the observable current inflation π_t (current economic activity y_t) and the target inflation π^* (target economic activity y^*) – the variable at interest in this study. Consequently, Equation (2.1) can be written with $\hat{\pi}_t = \pi_t - \pi^*$ and $\hat{y}_t = y_t - y^*$.

The inclusion of an inflation and an economic activity term in the monetary policy loss function stems from the official mandate ECB set out in the Treaty on the Functioning of the European Union. The Treaty specifies two objectives that the ECB must adhere to in hierarchical order. Article 123 §1 sets out the primary objective of maintaining price stability in the medium term. A subordinate objective of promoting the general economic policies of the European Union is laid out in Article 3. This is usually interpreted as an objective regarding output. However, as one of the European Union's explicit targets relates to employment levels, the unemployment rate is sometimes used as an appropriate indicator (e.g. Molodtsova and Papell, 2013). In Section 2.6, I present findings with respect to both hypotheses, a measure for output and unemployment as economic activity terms, demonstrating irrelevance of this choice to the overall results.

One essential presumption in the following analysis is that public statements of the central bank contain information that can be approximated by the tone of those discourses. In the communication literature on monetary policy, this principle has been applied numerous times. For instance, Bennani and Neuenkirch (2017) demonstrate that central bankers' sentiment in public communication is affected by current economic conditions and Bennani, Fanta, et al. (2020) provide evidence that this sentiment could be used as a predictor for future interest rate decisions. My paper continues with this practice of suggesting that speeches are a communication device used to inform the public about a central bank's objective and, thus, the words chosen in a speech can approximate that objective. Assuming that the loss function from Equation (2.1) can be approximated by a sentiment index $S_{i,t}$, the sentiment for speaker i at time t can be written as follows:

$$\begin{aligned} S_{i,t} &= \gamma_i \ L_t \\ &\approx \beta_{0,i} + \beta_1 \ \hat{\pi}_t^2 + \beta_2 \ \hat{y}_t + \epsilon_t. \end{aligned} \tag{2.2}$$

The introduction of a speaker-specific constant allows for personal and unique communication nuances between the speakers that are unrelated to the objective. This constant also ensures that the Executive Board's composition does not affect the estimated loss function. The necessity of this speaker specific term is further elaborated in Section 2.4.1, where the heterogeneity of the Executive Boards' communication is addressed. Disassembling the latent variables into its components, Equation (2.2) can be reformulated as:

$$\begin{aligned} S_{i,t} &= \beta_{0,i} + \beta_1 \ (\pi_t - \pi^*)^2 + \beta_2 \ (y_t - y_t^*) + \epsilon_t \\ &= \Phi_i + \beta_1 \ \pi_t^2 + \Omega \ \pi_t + \beta_2 \ y_t + \epsilon_t \end{aligned} \tag{2.3}$$

with $\Phi_i = \beta_{0,i} + \beta_1 \pi^{*2} + \beta_2 y^*$ and $\Omega = -2\beta_1 \pi^*$. Thus, the communicated inflation target can be effectively calculated as:

$$\pi^* = -\frac{1}{2} \frac{\Omega}{\beta_1} \tag{2.4}$$

Assuming that the perceived quadratic loss function can be approximated by the sentiment index, Equation (2.4) enables me to calculate the global minimum or maximum in the loss function with respect to the inflation rate. In other words, if the members of the ECB's Executive Board communicate an objective, which is, say, concave to the inflation rate, the sentiment in its speeches should be the highest – the most positive – if the current inflation is equal to π^*. At this point, the deviation from the inflation target is effectively zero, which in turn minimizes

Equation (2.1).

2.4 Data

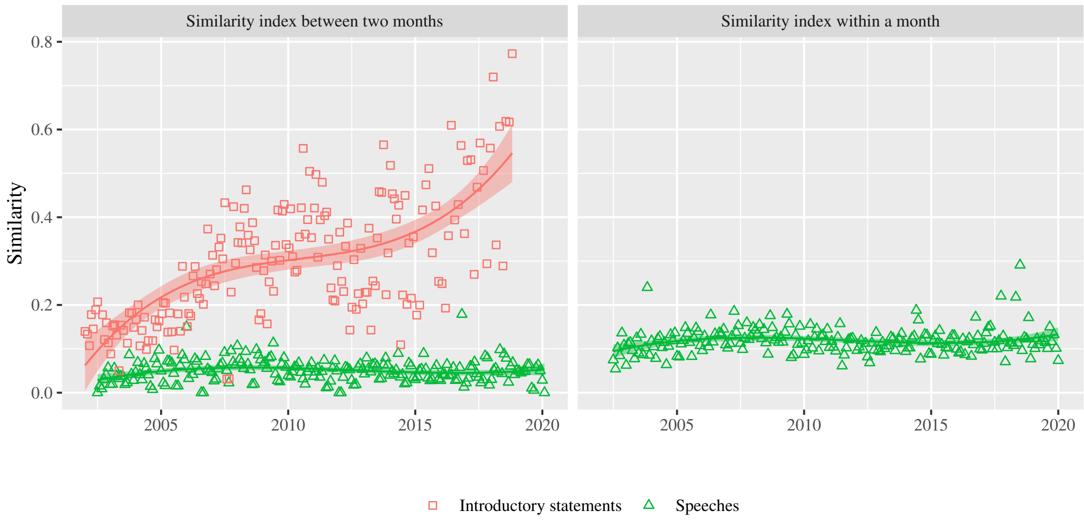
Following from the previous section, a sentiment indicator ($S_{i,t}$) as well as data on the objectives – inflation (π_t) and economic activity (y_t) – is required. Hence this section's focus is twofold. The estimation of the dependent variable, i.e. the sentiment index, is discussed first. The second part dicusses on the right-hand side macroeconomic variables.

2.4.1 Sentiment Index

Studies quantifying central bank communication differ mainly in two aspects: the underlying text corpus and the method of quantifying the information in the text. With respect to the underlying text corpus, most research on the ECB focuses on the organization's introductory statements to the monthly press conference following its monetary policy decision. While this permits for coherent observations – with constant time, location and communicator – it has two shortcomings: few observations and increasing similarity. Using exclusively press conference statements disregards much of the potentially relevant information the central bank provides to market participants. To be explicit: There are ten times the number of speeches compared to press conferences throughout my observation period. This larger amount of observation renders the quantification simpler. In addition, speeches remain heterogeneous in their language throughout the observation period as opposed to the press conference texts. Amaya and Filbien (2015) find the similarity of the ECB's press statements increased about fivefold between 1998 and 2014. A graphical illustration of both observations is provided in Figure 2.1, where each red squares present one ECB's introductory statement and each green triangle a speech. It is apparent that the public speeches between 2002 and 2020 do not exhibit such non-stationarity behavior in the similarity index.

Although not as abundant as studies on press conferences, there exists literature that investigates the ECB's communication beyond press statements. Significant contributions were made through the analysis of media coverage (Bennani, Fanta, et al., 2020), tweets (Masciandaro, Romelli, et al., 2020) and speeches

Figure 2.1: Similarity of ECB speeches.



Notes: The plot above illustrates three similarity indices estimated by Amaya and Filbien's (2015) measurement. Similarity is estimated as the occurrence of the same bigrams in two consecutive months (left plot) and within a month in different speeches (right plot) against time. Consequently, a high similarity (near one) is a sign for homogeneous texts, and the similarity is low (near zero) for heterogeneous text. The appendix provides a brief overview as to what bigrams are and a summary of Amaya and Filbien's (2015) index. The author's calculations for all three indices are available upon request.

(Hayo, Kutan, et al., 2008; Bennani and Neuenkirch, 2017; Gertler and Horvath, 2018). My work is closely related to the latter two studies, however, I use a novel dataset that enables me to evaluate all of the central bank's Executive Board public speeches. The speech corpus used in this analysis is larger in size, covers a longer time period and is uniform in terms of the organization represented. Thus differences in tone nuances can be attributed to the individual, rather than the speaker's organisation.

For the calculation of the sentiment index, I rely on the speech dataset provided by the ECB itself (European Central Bank, 2019b). The corpus contains (all) public speeches made by members of the Executive Board between January 1998 and January 2020. The dataset provides information on 2150 speeches, including the name of the speaker, the title, the subtitle, the date, and, of course, the text of the speech itself. Due to the missing observation in the dataset, I am forced to discard speeches held before January 2002. In addition, restrictions on word-, sentence- and speech-level are implemented.

To begin with, I convert all words to lower case and remove numbers and special signs. On the sentence level, I follow the convention of Shapiro and Wilson (2019) to include only sentences concerning economic content. Therefore, each sentence must contain at least one economic word or phrase as defined in the Oxford Dictionary of Economics (Black, Hashimzade, et al., 2017). This restriction intends to remove the non-economic parts of the speech – like obligatory thank-you remarks in the beginning. Although the dictionary is quite exhaustive, with more than 3,500 words and phrases, almost one-third of the sentences in the corpus do not pass this restriction. On the speech level, my restriction follows Bennani and Neuenkirch (2017), whereby each speech must contain at least 25 of the 'economic sentences' defined above. Imposing these restrictions yields a corpus of 1,915 speeches and roughly 170,000 sentences from 21 board members over a time horizon of 18 years.

There are different ways to reduce the dimensionality of such a dataset – i.e. to quantify text. The interested reader is referred to more comprehensive literature on textual analysis such as Gentzkow, Kelly, et al. (2019) and Bholat, Hansen, et al. (2015). The three most common methods are based on manually rating text by hand, enabling a computer to rate text according to a predefined dictionary and unsupervised machine learning techniques. This paper relies on the dictionary approach – often called a 'bag-of-words' – to calculate a sentiment index, using a predefined lexicon to count positive and negative terms. In recent years, the 'bag-of-words' approach on central bank communication text has been applied to the ECB (Tobback, Nardelli, et al., 2017), the Bundesbank (Tillmann and Walter, 2018), the FOMC (Apel, Blix Grimaldi, et al., 2019) and the Riksbank (Apel and Grimaldi, 2014), among others.

Regarding the lexicon, Loughran and McDonald (2011) provide evidence that the frequently applied Harvard dictionary causes misleading results when applied in the financial sector. Instead, they developed their lexicon, identifying around 2,700 words that are classified as either positive or negative. Although their dictionary is heavily biased towards negative terms – the updated dictionary contains 2,355 negative terms and only 354 positive terms – I will later show that the average sentiment estimated with this lexicon is not significantly different

Table 2.1: Statistic summary of Sentiment Index

	Number of sentences	Negative terms	Positive terms	Sentiment score $S_{i,t}$
Mean	91.14	47.78	36.66	-0.06
Std. dev.	(51.76)	(38.69)	(23.73)	(0.36)

Note: Descriptive statistics of our variables based on a total of 1,915 observations. The Loughran and McDonald (2011) dictionary is used to determine positive and negative terms. Sentiment is estimated according to Equation (2.5).

from zero.[20]

Following the literature on central bank communication (Bholat, Hansen, et al., 2015) the sentiment index $S_{i,t}$ is estimated by the relative fraction of positive minus negative terms. After removing positive and negative words that are preceded by "not" – as recommended by Loughran and McDonald (2016) – the sentiment for each speech is estimated to be the average sentiment of all sentences j in that speech:

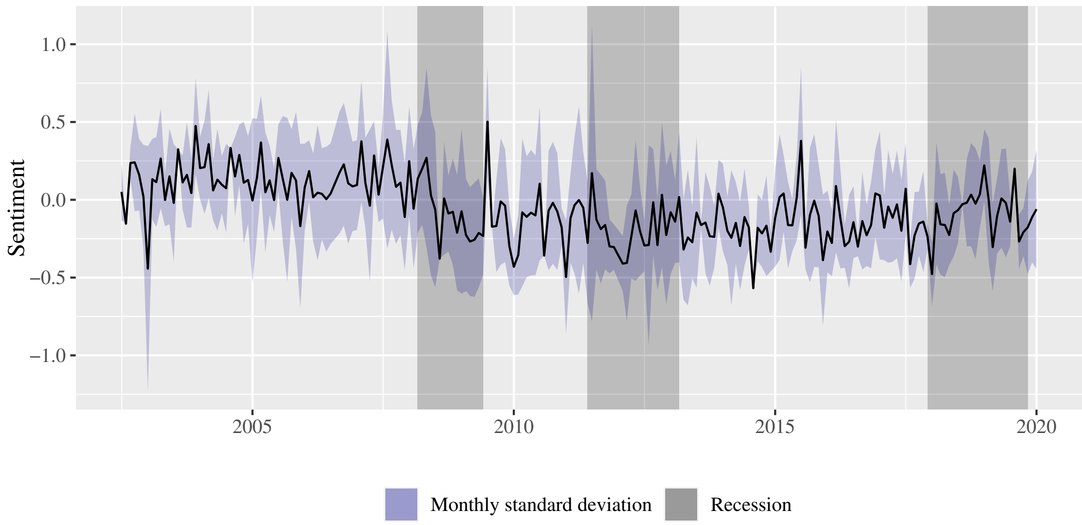
$$S_{t,i} = \frac{1}{n} \sum_{j=1}^{n} \frac{\#positive_j - \#negative_j}{\#positive_j + \#negative_j} \tag{2.5}$$

Using Loughran and McDonald's (2011) dictionary, the average speech contains approximately 48 negative terms and 37 positives. The resulting mean sentiment score is therefore -0.06. However, with a standard deviation of 0.36, the average sentiment is not significantly different from zero. The index's statistical summary can be found in Table 2.1.

In order to assess time-dependent developments, the average monthly sentiment and its standard deviation are graphically illustrated in Figure 2.2. Two noteworthy observations arise. First, there seems to be no adjustment in the sentiment during economic downturns (grey bars). One might expect the sentiment of those speeches to reflect the economic condition during the crisis, however, there is no evidence in this graphical representation. Second, there appears to be a shift from predominantly positive communication in the 2000s to largely negative communication in the 2010s. Given that monetary policy in the 2010s was dominated by the presence of the effective lower bound and inflation rates that are considerably below the explicit target, this observation provides first anecdotal evidence that

[20] An example text for illustration purposes is provided in Section B.1.

Figure 2.2: Monthly sentiment score.



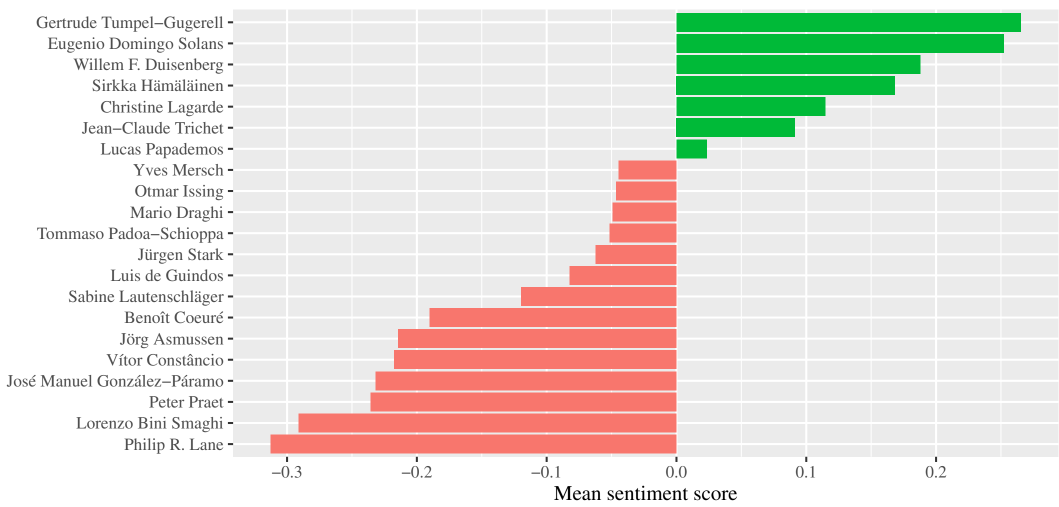
Notes: The plot above shows the monthly average sentiment index over time. In addition the standard deviation for each month is illustrated by the shading around the curve. Recessions are illustrated according to the organization of Economic Development (OECD) based recession indicator for the euro area. The author's calculations are available upon request.

the sentiment index measures the institutions' objective and therefore contains useful information. This interpretation should be regarded with caution, as given the small number of observations per month this shift is not statistically significant.

Although the index appears to be stationary, a unit root rest (Dickey and Fuller, 1979) on the sentiment scores cannot be rejected. To control for such non-stationarity, I include a delayed sentiment term into the regression. Since speeches do not occur on a regular basis (e.g. once every month), there are two possibilities to include such a term. In the baseline regression, the sentiment term from the previous speech is included, averaging the sentiment score when more than one speech was held on a single day. However, suspecting that a central bankers sentiment in a particular speech might be affected by the 'general' mood of her peers during speech writing, I also control for the lagged sentiment variable with the average sentiment of the previous month in one robustness check.

Before moving on, it is important to highlight three potential shortcomings of the text corpus. Table B.1 in the appendix provides an extended summary statistic for each Executive Board member, outlining the first two. First, the number of speeches each speaker delivered in the underlying text corpus differs greatly.

Figure 2.3: Speaker Sentiment



Notes: The plot above is an illustration of Table B.1 in the Appendix. The author's calculations are available upon request.

Some Executive Board members (in particular Jean-Cleaude Trichet) deliver far more speeches than others. Second, the speaker's average sentiment varies considerably across speakers as illustrated in Figure 2.3. This may suggest that there are members in the ECB's Executive Board who are generally more positive (or negative) in their communication. Note that the statistical rule of mean reversal appears to adhere to the sentiment scores, since the average score of the three most frequent speakers (Jean-Claude Trichet, Benoit Coeure, and Mario Draghi) can be found at the center of the distribution. Consequently, this indicates that variations in sentiment between speakers may be due to limited observations, and not to speaker's heterogeneity. Nevertheless, with those limitations in mind, the deviations are considerable. The most extreme (Philip Lane with -0.31 and Gertrude Tumpel-Gugerell with 0.27) are almost 0.6 index points apart. In absolute terms, this would imply a difference of 57 positive expressions between those two in an otherwise average speech. Note that the difference is substantial with an average of only 37 positive terms per speech (see Table 2.1). Therefore, in order to control for such diversity in communication, I include a speaker-specific constant in the regression model (see Equation (2.2) in the previous section) and thereby measure only the deviation from their mean.

Finally, it is not clear whether the information provided is of relevance. In particular, it is important to question whether the speaker provides information on the objective function of the institution. Hayo and Neuenkirch (2013) argue that the setting of a public speech might prevent a speaker from revealing true preferences, but adjust their language to the respective regional auditorium.[21] However, de Jong and van Esch (2014) argue that a public event may force the speaker to express her opinion on the official role she represents. That is particularly important for a central banker trying to shape expectations. Therefore, the sentiment index may represent the objective function as perceived by the employee. Implicit in this argument is the assumption that the members of the ECB Executive Board have a coherent function – and therefore an inflation objective – in mind when communicating, and only differ in word choice due to individual semantic preferences and their perception of the shared objective. The result would be the heterogeneity of speeches for which I control.

2.4.2 Inflation, Output and Unemployment

This second subsection focuses on the variables used to represent the objectives – i.e. inflation, output, and unemployment – and the need for real-time data in the analysis. In combination with the day-specific sentiment scores, this enables me to replicate accurately the information set available to the general public when listening to a speech.

Orphanides (2001) provides evidence that monetary policy regressions with revised data sets can lead to inaccurate findings. Since the ECB has acknowledged significant revisions of macroeconomic variables (European Central Bank, 2010), it is important to use unrevised data – real-time data – for the question at hand. I utilize real-time information provided by the ECB through its Real Time Database. The database collects revisions of macroeconomic variables in the Monthly Bulletin. By treating each revision as new information to the general public, I am able to create a real time dataset for the macroeconomic variables of interest. This approach has been used in other work with an interest in the

[21] It is crucial to emphasize, however, that Federal Reserve Fed presidents operate in a different institutional setting: they are expected to represent the special interests of their respective districts.

perception of ECB monetary policy (e.g. Gross and Zahner, 2021).
The HICP, as stated in the official objective,[22] expressed in annual growth rates, is used for the underlying analysis to measure the inflation rate. Economic activity is expressed either in terms of output, through the euro-wide annual GDP growth rate and the output-gap – calculated using a Hodrick-Prescott (HP) filter –, or in terms of the seasonally adjusted unemployment rate.
One assumption, at least implicitly, is that this macro-economic dataset is exogenous to the speech corpus. This assumption is likely to be true, since macroeconomic information collection is done independently from speech writing. However, the prospect of a revision of a macro-economic variable in the context of the forthcoming speech can not be completely excluded.
The regression analysis includes a number of control variables. As mentioned before, the speaker's specific sentiment as well as lagged sentiment is controlled for. In addition, one regression incorporates a time variable, capturing the development in the general sentiment as highlighted in the previous section.
Moreover, the presence of a stability objective, independent of inflation and economic activity, as a tertiary objective reemerged following the central banks' response during the financial crisis (e.g. Peek, Rosengren, et al., 2016; Kaefer, 2014). Using text-mining techniques on congressional hearings in the US, Wischnewsky, Jansen, et al. (2021) find evidence that deteriorating financial stability affect the monetary policy of FOMC. Consequently, to control for such a potential tertiary objective, three additional variables are included in the analysis. First, the EURO STOXX 50 Volatility (VSTOXX) is added in linear and square terms to measure the implied volatility for the euro area stock market. Second, the european uncertainty index of Baker, Bloom, et al. (2016) is included as a benchmark in logarithmic terms for political instability. Finally, the 3-month standard deviation of the EURO STOXX 50 is added to capture medium-term volatility in the asset market.

[22] https://www.ecb.europa.eu/mopo/strategy/pricestab/html/index.en.html (accessed 2020-05-01)

2.5 Results

This section presents the regression results, where the ECB's speeches are used to quantify the central banks' objective, allowing for the calculation of its implied inflation target. The regressions are estimated versions of Equation (2.3), using around 1.800 speeches from 21 speakers between January 2002 and January 2020. All results can be found in Table 2.2, with the implied inflation target estimates as calculated in Equation (2.4) and a 95% confidence interval at the bottom of the table.[23]

As discussed in the preceding section, two different potential subordinate objectives are being considered as representations for the ECB's goal of promoting the general economic policies of the European Union: output and unemployment. The discussion starts with the output growth rate as subordinate objective. The regression results can be found in Table 2.2 in the first three columns. Following that, the findings for a central bank with an unemployment subordinate objective are addressed. Columns four through six contain the results. Finally, the seventh column presents the findings of the ECB targeting both output and unemployment. Robustness checks with respect to (1) the exponentiation of the economic activity term, (2) an alternative economic activity terms, (3) the lagged sentiment indicator, and (4) the resorts of the central bankers conclude this section. My findings appear to be robust.

2.5.1 Inflation and output

To begin with, a naive model is estimated, including only the inflation rate and output growth as inputs, neglecting speaker heterogeneity and other controls like the tertiary financial stability objective. The results are shown in column one. There are several noteworthy observations by this naive model. First, the inflation coefficient is statistically significant and economically relevant. The coefficient suggests that an increase in the inflation by 1 percentage point leads to

[23]The inflation target and the confidence intervals were estimated using a non-parametric case resembling bootstrap method, sampling the regression 5000 times with replacement. The implied inflation target represents the average of the estimated implied inflation targets. The 95% confidence interval of these estimates is presented in this paper. Note that few extreme outliers were removed (<0.1%).

Table 2.2: Regression Results

	Sentiment Index $S_{i,t}$						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
π	0.118**	0.071**	0.079***	0.053	0.034	0.028	0.052*
	(0.050)	(0.031)	(0.025)	(0.043)	(0.031)	(0.027)	(0.030)
π^2	−0.019	−0.014	−0.015**	−0.010	−0.008	−0.006	−0.011
	(0.012)	(0.009)	(0.007)	(0.012)	(0.009)	(0.008)	(0.008)
y_t	0.049***	0.041***	0.042***				0.030**
	(0.010)	(0.012)	(0.010)				(0.012)
u_t				−0.040***	−0.028**	−0.034***	−0.024***
				(0.012)	(0.011)	(0.007)	(0.009)
S_{t-1}	0.117***	0.079***	0.059***	0.110***	0.083***	0.060***	0.052**
	(0.022)	(0.022)	(0.021)	(0.022)	(0.023)	(0.021)	(0.021)
Year			−0.012*			−0.015***	−0.016***
			(0.007)			(0.005)	(0.005)
EPU_{Europe}			−0.089***			−0.079***	−0.070***
			(0.022)			(0.015)	(0.016)
Speaker Dummy	No	Yes	Yes	No	Yes	Yes	Yes
Inflation target	3.4	2.9	3.2	2.8	2.2	2.1	2.2
CI	[3.3, 3.5]	[2.7, 3.1]	[3.0, 3.5]	[2.2, 3.4]	[1.9, 2.4]	[1.7, 2.5]	[2.0, 2.5]
Observations	1,749	1,749	1,749	1,749	1,749	1,749	1,749
Adjusted R^2	0.057	0.212	0.225	0.063	0.210	0.224	0.228

Note: Coefficients are estimated using an OLS regression. The variable y_t denotes log GDP. Financial market controls, as well as a time trend, are included in specifications (3), (6), and (7), as defined in Section 2.4.2. EPU_{Europe} is the uncertainty index by Baker, Bloom, et al. (2016). Robust standard errors (clustered by speaker) are displayed in parentheses. ***, **, * indicate significance at the 1, 5, and 10 per cent level, respectively.

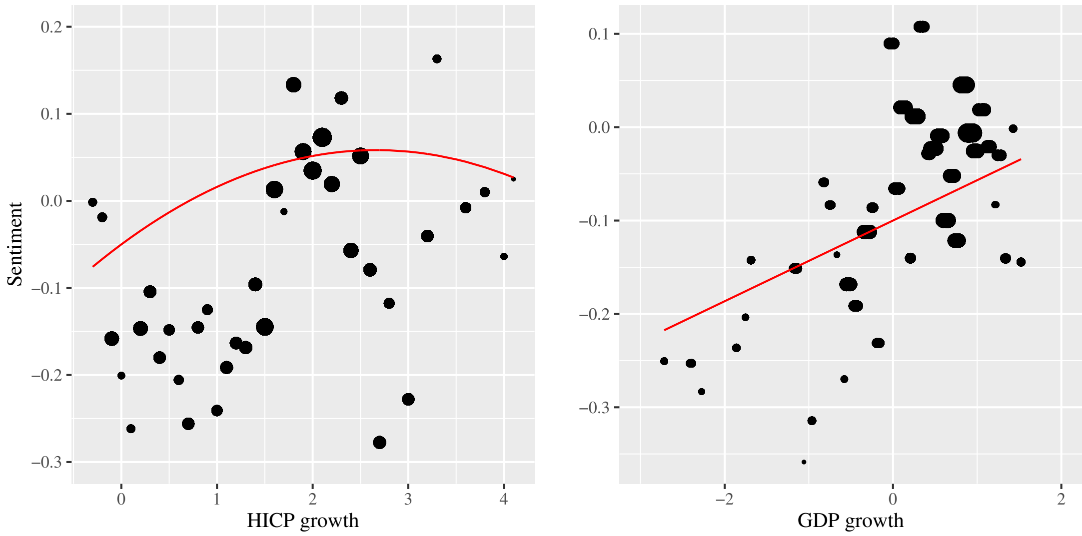
an increase of the sentiment index by 0.12. This translates to around 10 more positive words throughout an ordinary speech, although the negative quadratic term may partly offset this effect. The positive sign of the inflation coefficient, combined with the negative sign of the squared inflation coefficient indicates a concave relationship towards inflation. Importantly, the concavity in the inflation relationship is consistent across all specifications in this paper. The existence of a such concave inflation target implies the existence of a global maximum for that objective. Based on the results in the naive regression model, the communicated inflation target of the ECB is more than 3%. This is well above the official target of 'close, but below 2%'.

Second, the output growth coefficient is significant and suggests a positive correlation between production and sentiment, i.e. one additional percentage point of GDP growth increases an average speech by approximately four positive terms. Therefore, when comparing only the linear coefficients, the hierarchical order of the ECB's objective appears to be manifested in its speeches sentiment. However, this might be driven by the heterogeneity communication between speakers, as we will see below. Finally, the tone of the previous speech – delayed sentiment index – is a highly statistically significant and relevant approximate, even if the speakers differ, indicating persistence in tone over time. The Durbin-Watson test suggests that with the inclusion of this delayed index, no auto-correlation remains in the residual.

Taking into consideration that much of the variation in the sentiment of a particular speech may arise from the semantic differences between the speakers, I will next correct for this. In the second column of Table 2.2, the regression results controlling for speaker-specific nuances are presented. The main findings are the following. First, the speakers generally enter the regression results with the anticipated signs. The most pessimistic speaker is Lorenzo Smaghi (-0.135***), whereas Gertrude Tumpel-Gugerell (0.419***) is the most optimistic speaker after controlling for economic conditions. The inclusion has significant implications for the delayed sentiment index, which decreased by a third in magnitude. Second, controlling for speaker-specific sentiment has a significant impact on the magnitude of the implied inflation target, which decreases to 2.9%. This suggests that the previous finding may genuinely have been affected by the heterogeneity of the members in the ECB's Executive Board. Using bootstrapping on the perceived inflation target, the 95% confidence interval can be estimated at [2.7%, 3.1%]. This interval still clearly rejects the null hypothesis that the target is articulated to be zero. It also rejects the interpretation that the implied inflation target does not significantly exceed 2%.

Third, the output coefficient remains positive, statistically significant and decreases only to a small degree in magnitude. As a result, the ECB appears to

Figure 2.4: Bin-Scatter Plot of Output Regression (3)



Notes: The illustration of a 'bin-scatter-plot' represents a practical alternative to the more conventional scatter plot. The data points are grouped into bins, and each bin is averaged. In addition, the size of each point is proportional to the number of data points within the respective bin. The objective (red continuous line) is illustrated as of regression results (3). The author's calculations are available upon request.

be only slightly more responsive to deviations in inflation than production, the difference is not high. A back-on-the-envelope calculation shows that an increase in GDP by 1 percentage point has less of an effect on the sentiment index than an increase in inflation from 1% to 2%.[24] Fourth, the regression's explanatory power increases more than five-fold, suggesting the necessity to incorporate speaker-specific information in the regression model. Note that to economize on space the only measure for the goodness-of-fit provided here is the adjusted R^2 but the AIC confirms these findings.

Finally, I include the additional measures of financial stability and the time-dependent dummy. The results are qualitatively and quantitatively consistent with the previous ones. The inflation rate becomes highly significant – as does the squared inflation term – and the implied inflation target increases slightly to 3.2%. A graphic illustration of the objective, as estimated in this regression, is provided in Figure 2.4. The left bin-scatter plot illustrates the perceived inflation

[24]Raising the inflation rate from 1% to 2% increases the sentiment by $2\times0.071-2^2\times0.014-(1\times0.071-1\times0.014)\approx0.03$, whereas an increase in GDP growth by 1% increases the sentiment by ≈ 0.04.

target. The concavity of the communicated objective itself (red line) is visually highlighted. The estimated relationship towards output is illustrated on the right side of the above mentioned bin-scatter plot. The positive correlation between sentiment and GDP growth rates is evident. The plot provides further evidence on the stronger reaction of sentiment to changes in production than inflation. GDP growth remains a highly significant variable.

In addition a potential tertiary objective towards financial stability, as identified for the FOMC (Wischnewsky, Jansen, et al., 2021), is added to the regression. However, neither the VSTOXX, nor the standard deviation of the euro area stock market have a significant impact. Only Baker, Bloom, et al.'s (2016) uncertainty index is statistically significant. A 1% increase in the index decreases the respective speech by about six positive words, indicating that the discourse is becoming extra pessimistic in times of heightened uncertainty. Furthermore, there appears to be a negative trend in sentiment over time. This finding is significant and economically relevant, with an average of almost two fewer optimistic terms used per speech every year. The decline in sentiment over time is consistent with the anecdotal evidence presented in Figure 2.2 and may be indicative of the central bankers' increasingly challenging role of maintaining inflation in the euro area while being restricted by the effective lower bound.

In my findings, the adjusted R^2 is substantially higher than that by Shapiro and Wilson (2019) on their public communication dataset of the FOMC. As mentioned previously, the analysis of Shapiro and Wilson (2019) focuses on the FOMC's non-public communication. There, up to a quarter of the variation can be explained by their regression ($R^2 \approx 0.24$). However, the model's explanatory power decreases to $R^2 \approx 0.06$ when using speech data. Therefore, with regard to explanatory power, my findings ($R^2 \approx 0.22$) on public speeches by the ECB seem to be in line with their results on the private correspondence of the FOMC.

Overall, the results of the regressions suggest that the objective is communicated with a concave relationship to the inflation rate and a linear increasing relationship to the output growth rate. This finding implies the presence of an inflation rate, which minimizes the loss function of the ECB. According to my results, the implied inflation target is around 3%. Above and below this price increase, the institution's sentiment in speeches becomes more negative. Importantly, my findings appear to imply that deviations in output growth, rather than inflation, appear to be the most relevant factor in explaining variations in sentiment. This finding contrasts with the results of Shapiro and Wilson (2019) and Paloviita, Haavio, et al. (2020). In addition, speaker-specific effects have a huge impact on the sentiment score and financial uncertainty tends to influence the speeches. In the next step, I will substitute the GDP growth rate with the unemployment rate in order to test the consistency of the findings.

2.5.2 Inflation and unemployment

As discussed in Section 2.4.2, there remains ambiguity regarding the subordinate objective of ECB. In the previous subsection, the findings when interpreting the aim of promoting the general economic policies in the euro area as an explicit production growth target were presented. As mentioned in Section 2.2, the unemployment rate is an alternative metric for this subordinate objective. Qualitatively, my results on the inflation target appear to be independent of the choice in the economic activity objective. However, using employment reduces the statistical power of the inflation coefficients. The results can be found in Table 2.2 in columns four to six.

Embedding the unemployment rate in the regression instead of the production growth rate yields several findings. First, the implied inflation target does diverge from previous findings but may still be best described as *'above, but close to 2%'*. Again, it is noteworthy that the implied inflation target decreases, once I control for the respective speaker. Although they deviate in quantitative terms, the findings of the preceding subsection can be qualitatively confirmed. The implied inflation targets are consistently between 2% and 3%. A graphical illustration of the inflation objective – including speaker-specific effects, financial stability controls and time effects – can be found in the appendix.

Second, the indicated relationship towards unemployment is negative and highly statistically significant, indicating that the ECB communicates a preference towards lower unemployment. Analogous to my previous findings, the reaction of the sentiment to changes in unemployment is economically relevant and dominates the effect of changes in the inflation rate. An increase in the unemployment rate by 1% decreases an average speech by up to three positive terms. Finally, the integration of financial stability and a time index yield findings that parallel those in the output-case. The depreciation of sentiment over the observation horizon is now highly statistically significant and, in terms of magnitude, very similar to what was previously found.

2.5.3 Inflation, output, and unemployment

In the last regression, GDP growth is embedded together with the unemployment rate in the regression. The results are presented in the last column in Table 2.2. Interestingly, both economic activity coefficients – GDP growth and unemployment rate – remain highly statistically significant, but decrease slightly in magnitude. The inflation rate yields only weakly significant results and is further dominated by the economic activity coefficients in terms of economic relevance. The back-of-the-envelope calculation (ignoring significance levels) shows that GDP growth would have to increase by only 0.5% (or the unemployment rate decrease by 1%) in order to have the same impact on the communicated loss as an increase in the inflation rate from 1% to 2%.[25] This further indicates that central bankers are reacting to changes in GDP and unemployment rather than inflation and may be more concerned with these variables. Nevertheless, the implied inflation target remains above, albeit close to 2%. Also its confidence interval [2.0, 2.5] still rejects the 'below 2% inflation' target.

2.6 Robustness checks

This final subsection covers a set of robustness checks conducted for my analysis. To economize on space, the regression results are shown in Section B.5. The first two robustness checks concern the presence of a non-linear economic activity term (columns one and two) and the choice of output measurement (column three). This section concludes with a robustness check of an alternative delayed

[25] Raising the inflation rate from 1% to 2% increases the sentiment by $2 \times 0.052 - 2^2 \times 0.011 - (1 \times 0.052 - 1 \times 0.011) \approx 0.02$, whereas an increase (decrease) in GDP growth (unemployment rate) by 1% increases the sentiment by ≈ 0.03 (≈ 0.024).

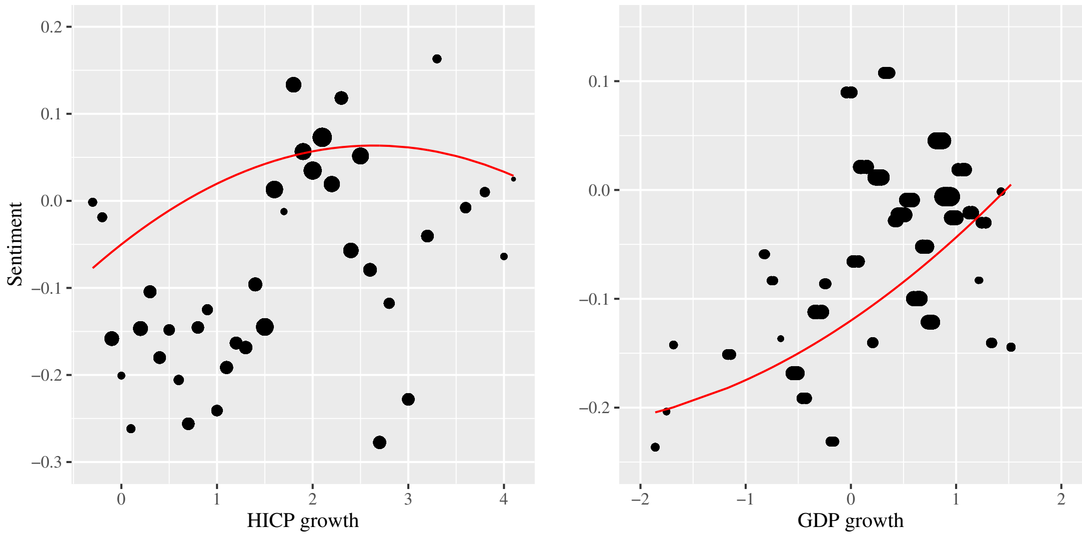
sentiment index (columns four and five) and the inclusion of task-specific central banker controls (last two columns).

In Equation (2.1), following standard textbook literature (e.g. Walsh, 2017; Barro and Gordon, 1983), $\hat{y}_t$ enters the objective function with a linear term. However, one could also use non-linear version (e.g. Walsh, 2017; Shapiro and Wilson, 2019; Paloviita, Haavio, et al., 2020). In this subsection I validate my results using a squared economic activity in addition to the linear one, demonstrating that the results qualitatively hold independently of this assumption. The presence of this non-linear term has the interesting side effect of allowing me to calculate the optimal level of output and employment, i.e. the implicit bliss point for the economic activity term.

Including the non-linear term yields the following results: There is only a slight variation in the implicit inflation target, being 2.6%. The non-linear output coefficient is positive and statistically significant (as is the linear coefficient), indicating a convex production objective for ECB. The implicit minimum is a growth rate of -3%. This growth rate is below the lowest reported GDP growth rate in the euro area. In other words, the Executive Board at the ECB communicates more pessimistically until production deteriorates by 3% year-on-year, i.e. the central bankers seem to favor a positive output-gap, independently of the current economic situation. Note that the implication is therefore very similar to the linear case. The objective is illustrated in Figure 2.5. When the squared unemployment rate is included, the coefficients do not yield the same statistical significance anymore. Since u_t^2 and u_t are highly correlated ($\sim 99\%$), this insignificance is likely to be driven by multicollinearity. Hence, for the unemployment rate, the linear specification seems preferable. Finally, in both robustness tests, the remaining variables join with similar sign, magnitude and significance level as they did in the baseline regression.

The second robustness reviews the validity of GDP growth as representation for the economic activity objective. The use of the output-gap is an alternative to the previously applied growth rate. The output-gap is estimated using a HP-filter and replaces the growth rate in this regression. While inflation becomes a highly statistically significant variable, the output-gap is not significant. In addition,

Figure 2.5: Bin-Scatter Plot of Robustness Check Regression (1)



Notes: The illustration of a 'bin-scatter-plot' represents a practical alternative to the more conventional scatter plot. The data points are grouped into bins, and each bin is averaged. In addition, the size of each point is proportional to the number of data points within the respective bin. The objective (red continuous line) is illustrated as of the robustness regression results (1). The author's calculations are available upon request.

the response of the sentiment index to uncertainty as measured by the Baker, Bloom, et al.'s (2016) index increases. The inflation target in this regression can be calculates as 2.9% inflation.

The third robustness check concerns the delayed sentiment variable. The sentiment of the previous speech was included in the baseline regression. However, in order to capture the general mood when giving the speech and test the robustness of this choice, I substitute this term with the average sentiment of the previous month. The findings remain the same. The result indicates – similar to the previously used lagged variable – a significant positive impact. Both, output and unemployment, remain significant variables in this regression, while inflation remains significant in the output specification.

The final robustness check addresses the speaker specific effects. One might argue that the resorts (departments) held by a central banker influence the sentiment of this banker's speech, since a central banker speaking about, say, "Risk Management", may be forced to choose a more pessimistic language than her peer speaking about the "New ECB Premises Project". Therefore, I reconstruct the resorts held by each central banker at the time of her speech and include the corresponding dummy in the regression. Although most resorts do not yield significant impact on sentiment, the hypothesis of irrelevance can be rejected by a joint significant test. Nevertheless, the results indicate that the inclusion of speaker-specific dummies is preferable in terms of explanatory power.

2.7 Conclusion

With its price stability mandate in the euro area, the ECB is tasked to pursue inflation with a subordinate goal to stabilize economic activity. The general public's perception of this objective function is important for the effectiveness of the central bank's monetary policy, which in turn incentives the ECB to clearly communicate it. In light of the current internal evaluation of the central bank's monetary policy as well as the ongoing debate over the specific functional form of its objective, this paper employs a text mining approach proposed by Shapiro and Wilson (2019) to approximate the objective function and its determinants. Using a sentiment index, the approach quantifies the central bank's satisfaction with the current economic state. Taking all public speeches delivered by the ECB's Executive Board between 2002 and 2020, I contribute to a growing body of academic work focusing on quantifying qualitative text data.

First, the findings of my analysis suggest a concave objective regarding the inflation rate. The implied optimum for inflation – the ECB's inflation target – is best described as *'above, but close to 2%'*. Deviations from this value lead to a reduction in the sentiment of the institution's communication. This finding contrasts the results obtained by text analysis on press conference statements by Paloviita, Haavio, et al. (2020), who found a communicated target of $\sim 1.6\%$. However, as pointed out in this paper, the non-stationarity in similarity in the press-conference statements, in comparison to the speeches, makes me confident in the presented results. Furthermore, my results appear to be robust across a wide range of model-specifications and determinants.

Second, including output growth or the unemployment rate as a subordinate aim of promoting the European Union's general economic policies yields qualitatively similar results. My findings suggest a potential concave relationship between the central bankers sentiment and output, as well as a linear relationship towards the unemployment rate. Independent of the specification, my findings suggest a communicated preference for higher GDP growth and lower levels of unemployment, regardless of the current economic condition. The hierarchical order in the ECB's mandate does not always appear to be consistent with my findings. Deviations from its primary objective inflation appear to be of no greater concern than deviations in its subordinate objective. In fact, both output and unemployment are consistently significant variables, in contrast to inflation.

Third, in contrast to the findings of the FOMC, financial market variables have no significant effect on the central bankers' sentiment. Fourth, over the last two decades, the tone of speeches has deteriorated, even when controlling for economic condition. This might be indicative for the increasing challenging role of the ECB after approaching the effective lower bound. Fifth, during times of crisis and heightened uncertainty, central bankers appear extra pessimistic. Finally, the explanatory power of public communication of the ECB is surprisingly high, especially when compared to previous studies on the FOMC.

This paper provides a first analysis of the ECB's objective function using text mining on the institutions' speeches. Future research may further examine the tremendous amount of information provided by the speeches and the heterogeneity of the speakers. One possibility would be to extend the analysis beyond its sentiment by incorporating meta-information provided by the text of the speech. Evidence of the impacts of complexity and length of press conference statements have previously been observed (Smales and Apergis, 2017b; Hayo, Henseler, et al., 2020). Further potential might be in the heterogeneity of communication between central bankers or even national central banks (Hayo and Neuenkirch, 2013; Tillmann and Walter, 2018). This could provide an interesting comparison of the differences across countries. In addition, it would be one way of testing whether there is evidence for the hypothesis that a central banker communicates the inflation target of a central bank, or vice versa.

3 Complexity of ECB Communication and Financial Market Trading

Bernd Hayo,[c] Kai Henseler,[d]
Marc Steffen Rapp[e] and Johannes Zahner[b]

Abstract

We empirically examine how complexity of ECB communications affects financial market trading based on high-frequency data from European stock index futures trading. Our sample covers ECB press conferences between January 2009 and December 2017, during which unconventional monetary policy measures (UMPM) substantially increased communication complexity. Analysing the linguistic complexity of the introductory statements and differentiating between press conferences with and without UMPM-announcements, we find more complex communication, i.e. high linguistic complexity and UMPM-announcement, is associated with a lower level of contemporaneous trading activity. Moreover, complex communication leads to a temporal shift in trading activity towards the subsequent Q&A session, which suggests that Q&A sessions facilitate market participants' information processing. Finally, we document a relatively lower similarity of unconventional monetary policy statements and argue that this might explain our findings.

Keywords: ECB, Central Bank Communication, Textual Analysis, Linguistic Complexity, Readability, Financial Markets, European Stock Markets.

JEL classification: D83, E52, E58, G12, G14.

---

[c]School of Business and Economics, Macroeconomic Research Group, Philipps-Universität Marburg, Germany.
[d]School of Business and Economics, Accounting and Finance Area, Philipps-Universität Marburg, Germany.
[e]Corresponding author. School of Business and Economics, Accounting and Finance Area, Philipps-Universität Marburg, Germany.
[b]School of Business and Economics, Institutional Economics Research Group, Philipps-Universität Marburg, Germany.

3.1 Introduction

Over the last decade, central bank communication has become a key component of the central bankers' toolkit (see, e.g., Dell'Ariccia, Rabanal, et al., 2018; Kuttner, 2018).[26] To effectively steer the expectations of the private sector with the aim of enhancing monetary policy transmission, central banks in several countries have institutionalised monetary policy communication (see, e.g. Blinder, Ehrmann, et al., 2008). Announcements about current monetary policy decisions, assessments of the economic outlook, and the expected consequences of monetary policy have become an important tool of central banks' communication strategy (see, e.g. Hansen, McMahon, and Prat, 2018; Kohn and Sack, 2003)

However, the increase in complexity of monetary policy during and after the financial crisis creates significant challenges for central bank communication (see, e.g. Bulir, Cihak, and Jansen, 2013; Bulir, Cihak, and Smidkova, 2013; Hernández-Murillo and Shell, 2014). As Peter Praet, former chief economist of the ECB, put it *'[i]n normal times, central banks adapted their monetary policy stance by influencing the level of one short-term interest rate. In unconventional times, communication has had to cope with the new challenge of explaining the complementarities between policy tools, as non-standard monetary policy has become multidimensional. [...] In this context, it is perhaps no coincidence that the complexity of the introductory statements delivered at the ECB's press conferences, as measured by common indices of text readability, has also increased'.*[27]

While there is a large number of studies exploring central bank communication (see the surveys by Blinder, Ehrmann, et al., 2008; de Haan and Sturm, 2019), there is still little understanding as to how financial markets are influenced by the complexity of central bank communication. We address this gap in the literature

[26]Communication is "a process by which information is exchanged between individuals through a common system of symbols, signs, or behaviour" (Merriam Webster Dictionary, available at: https://www.merriam-webster.com/dictionary/communication (accessed: 05 Mar 2019)). Central banks can utilise communication to reduce asymmetric information and share their private information to guide expectations. Central banks' private information may stem from a myriad of sources, such as the outcome of previous policy votes (see, e.g., Meade, 2005), the discussions at the meeting (see, e.g. Hansen, McMahon, and Prat, 2018), or risk balance evaluations (see, e.g. Hanson and Stein, 2015).

[27]See https://www.ecb.europa.eu/press/key/date/2017/html/ecb.sp171115.en.html (accessed Dec 2, 2020).

by studying the European Central Bank's (ECB) press conferences following the Governing Council meeting and the impact of linguistic and content-related complexity of the introductory statement for contemporaneous trading behaviour in financial markets. After interest rates reached the effective lower bound following the 2008 financial crisis, many central banks around the world embraced unconventional monetary policy measures (UMPM), such as quantitative easing and forward guidance (e.g., Bowdler and Radia, 2012). Arguably, the complexity – and potential ambiguity – associated with these novel monetary policy tools demands a more disciplined and coherent communication strategy, especially since the effectiveness of monetary policy potentially seems to increase in the degree of comprehension of financial market participants (see, e.g. Cœuré, 2018; Lucca and Trebbi, 2009; Praet, 2017).[28]

To date, central banks in all major economies conduct regular press conferences following the meetings of their monetary policy committees. The ECB instituted its press conferences after the Governing Council Meetings (GCM) right from its establishment. Each press conference begins with a prepared introductory statement and ends with a Q&A session attended by journalists. In the light of the ECB's unique communication design, we address the question of whether higher complexity of central bank communication, causes financial market participants to delay trading and whether the generally less complex Q&A sessions may mitigate the effect.

We measure market trading using high-frequency data from European stock index futures trading and study ECB press conferences in the aftermath of the 2008 financial crises, during which unconventional monetary policies substantially increased communication complexity. We proceed in four steps. First, we analyse the linguistic complexity of the introductory statements. Differentiating between press conferences with and without the announcement of UMPM, we find no difference in their linguistic complexity. Second, examining the overall effect of linguistic complexity on trading volume, we find – in contrast to Smales and Apergis (2017a) who study the Federal Open Market Committee (FOMC) –

[28] An extensive discussion on how financial market participants themselves evaluate the success of these policies is provided in Hayo and Neuenkirch (2015a) and Hayo and Neuenkirch (2015b).

no effect in our sample period. Third, differentiating between press conferences with and without the announcement of UMPM, we find that higher linguistic complexity of introductory statements is associated with a lower level of contemporaneous trading activity for UMPM-announcements. Moreover, we find that increasing complexity leads to a shift in trading activity towards the subsequent Q&A session, which suggests that Q&A sessions facilitate market participants information processing. Finally, drawing on Ehrmann and Talmi (2020), we analyse the similarity of introductory statements and infer that the observed effect of UMPM-announcements is due to their 'unconventionality' that is, their degree of novelty. Specifically, we document that UMPM communication is, on average, less similar and, therefore, more likely to transmit a higher degree of potentially complex new information.

The remainder of this paper is structured as follows: The next section develops the central research question and presents our hypotheses. Section 3.3 describes the dataset and provides the descriptive analysis. Section 3.4 illustrates our empirical design and presents the regression results. Section 3.5 discusses the robustness of the results and Section 3.6 concludes.

3.2 Central Bank Communication and Financial Markets

Economic theory suggests that trading decisions depend on 'news', i.e. novel information (see, e.g., Stigler, 1961), which is swiftly incorporated by efficient financial markets (see, e.g., Fama, 1979). Central bank communication often contains such relevant news about future economic developments, with consequences for the macroeconomy, specific industries, and individual companies (see, e.g. Bernanke and Kuttner, 2005; Funke and Matsuda, 2006). Consistent with that view and Cukierman and Meltzer's (1986) hypothesis, Andersson (2010) and Nakamura and Steinsson (2018) find evidence that unexpected information (i.e. surprises) in central bank communication has an immediate effect on financial markets.

Most studies analysing the informational content of central bank communication focus on well-defined signals from the central bank, such as monetary policy announcements (see, e.g. Blinder, Ehrmann, et al., 2008). In an attempt to minimise omitted variable bias and endogeneity, these studies commonly take an event-study approach (see, e.g. Rosa, 2011a). The dependent variables employed typically include some short-term reactions by financial markets around monetary policy announcements (see, e.g. Boguth, Grégoire, et al., 2019; Brand, Buncic, et al., 2010; Gürkaynak, Sack, et al., 2005; Hussain, 2011; Rosa, 2011a; Rosa, 2011b; Schmeling and Wagner, 2019). Other studies quantify the content of these announcements through text-mining techniques and investigate communication of various central banks, e.g., the ECB (Picault and Renault, 2017), the FOMC (Shapiro and Wilson, 2019), the Bundesbank (Tillmann and Walter, 2018), and the Riksbank (Apel and Grimaldi, 2014).

In general, this stream of research considers information to be a rather simple construct, easily understood and comprehended by market participants. However, several studies question this assumption and emphasise the possibility of (1) variations in the degree of understanding and interpretation of information (see, e.g. Grossman and Stiglitz, 1976; Harris and Raviv, 1993) and (2) heterogeneity in the speed of information processing. Kandel and Pearson (1995), for instance, suggest that different ex ante opinions may rationalise dispersion in interpretation, that is, while all market participants receive the same information, their assessment is heterogeneous. Alternatively, the (lack of) general comprehensibility of the information could be the cause for the differential interpretation of information (see, e.g. Loughran and McDonald, 2016; Smith and Taffler, 1992; You and Zhang, 2009). That is, all market participants receive the same information but decode it differently and/or at a different speed due to the contents' complexity. Hong and Stein (1999) argue that private information may be required to transform public news into an opinion and heterogeneity in private information may result in gradually updated opinions and, thus, an underreaction of the market to public news.

Regarding central bank communication, Ehrmann and Talmi (2020) report substantial similarity in press releases announcing monetary policy decisions. They find that similarity of press releases of the Bank of Canada is negatively associated with market volatility. Examining FOMC statements, Hernández-Murillo and Shell (2014) document that these statements have become more complex since the beginning of UMPM. Smales and Apergis (2017a) and Smales and Apergis (2017b) investigate in two studies the effect of linguistic complexity of FOMC statements and find that complexity positively affects daily trading volume. The authors rationalise their finding with heterogeneity in beliefs and opinions because of the complexity of information in light of Harris and Raviv (1993) and Kandel and Pearson (1995).

In this paper, we extend the analysis of Smales and Apergis (2017a) and Smales and Apergis (2017b) along two dimensions. First, we are interested in the dynamics of information processing and trading behaviour in financial markets. In light of Hong and Stein (1999), we argue that at a given level of cognitive ability and private information, the time to process news is positively correlated with the complexity of the text containing that news. Hence, we expect that the market underreacts to more complex central bank communication and that contemporaneous trading volume is negatively correlated with complexity:

> H1: Complexity of central bank communication has a negative impact on contemporaneous trading behaviour.

Second, we argue that it is not only the linguistic complexity of the transcripts that matters, but also the complexity of the context and content that matters. Following Peter Praet, former chief economist of the ECB, who argues *'[a] multi-instrument policy toolkit [UMPM] is more complex because it adds a further dimension to the central bank reaction function'*,[29] we posit that announcements of UMPM are more complex in context and content. Hence, in the case of UMPM-events, we expect the underreaction of the market to be even more pronounced.

> H2: Complexity of central bank communication has a more negative impact on contemporaneous trading behaviour, when communication refers to unconventional monetary policy measures.

Finally, we shed some light on the question of whether the unique communication design of the ECB, where each press conference begins with a prepared introductory statement and ends with a Q&A session attended by journalists,

[29] See https://www.ecb.europa.eu/press/key/date/2017/html/ecb.sp171115.en.html (accessed Dec 2, 2020).

may mitigate the underreaction of the market. Arguing that communication in Q&A sessions is less formal and thus less complex, we hypothesise that Q&A sessions may be helpful for reducing heterogeneity in information processing and opinions and thus attenuate the underreaction of the market to complex news.

> H3: There is a positive relationship between ECB communication complexity and a temporal shift of trading activity to the Q&A session.

3.3 Sample and Descriptive Analysis

To test the three hypotheses, we analyse the effect of complexity in introductory statements of the ECB press conferences using high-frequency trading volume data from European stock index futures. Our core sample contains all press conferences from January 2009 to December 2017.[30] It covers the aftermath of the 2008 financial crises, when the ECB started conducting UMPM on a recurring basis. Specifically, it covers the announcement of ECB's first covered bond purchase programme on 07 May 2009 (see, e.g. Henseler and Rapp, 2018).

3.3.1 Introductory Statements to ECB Press Conferences

The main decision-making body of the ECB is the Governing Council, which assesses economic and monetary developments and conducts monetary policy decisions on a regular basis at the ECB's premises in Frankfurt am Main, Germany.[31] After GCMs, the ECB issues a press statement at 13:45 CET on its interest rate decision, followed by a press conference, where the monetary policy decisions are explained in detail by the ECB's president, sometimes supported by other members of the Executive Board.

A typical GCM press conference proceeds as follows. After the official start at 14:30 CET, the ECB's president reads a prepared introductory statement, which

[30] In the robustness checks, we extend the sample to cover the January 2003 to December 2017 period. Our results remain unaffected (see Section 5.1).

[31] A detailed and comprehensive description of the Governing Council's responsibilities can be found at: https://www.ecb.europa.eu/ecb/orga/decisions/govc/html/index.en.html (accessed: 17 Feb 2019).

covers the GCM's decisions, the underlying rationale, and a monetary policy outlook. This introductory statement takes between 10 and 20 minutes, with mean and median at 15 minutes (for our sample). Subsequently, a 40- to 60-minute Q&A session is held, starting at around 14:50 CET. During this, local participants (usually press representatives) ask questions, which are answered by the president. The Q&A session is explicitly intended to make the correspondence of the ECB as clear as possible (see, e.g. Cœuré, 2018). The press conference concludes between 15:30 to 15:50 CET.

Searching the ECB webpage, we identify all GCM press conferences during our sample period. For each press conference, we download transcripts of the introductory statement and save it in a separate text file.[32] We opt for analysing the GCM press conference introductory statements, since they represent an important and standardised part of ECB communication (e.g. Hayo, Henseler, and Rapp, 2019): important, as it embodies the communication as intended by the ECB, and standardised, as the statements exhibit a common structure and duration. Still the statements differ in content and, hence, provide an appropriate basis for comparative text analysis. Overall, our sample covers 95 introductory statements.

3.3.2 Unconventional Monetary Policy Measures

In a detailed content analysis, we assess the introductory statements with regard to the disclosure of Asset Purchase Programmes, Swap Agreements, Allotment Policy, and/or Forward Guidance. If at least one of these topics is discussed substantively, a dummy variable UMPM is coded 1 and 0 otherwise. A comprehensive list of the resulting 34 press conferences can be found in Table C.1 in the Appendix.

3.3.3 Measuring Complexity of Introductory Statements

To quantify the latent dimension of comprehensibility, we follow the linguistic approach of and Hernández-Murillo and Shell (2014), Smales and Apergis (2017a),

[32]ECB press conference transcripts with introductory statements and Q&A sessions are available at: https://www.ecb.europa.eu/press/pressconf.

and Smales and Apergis (2017b) and use the Flesch-Kincaid Grade Level (Kincaid, Fishburne Jr., et al., 1975) to measure complexity in the introductory statements.[33]

The Flesch-Kincaid Grade Level score (FK) is a linear function in the average sentence length and the average word length measured in syllables. Technically, for a document i it is calculated as:

$$FK_i = 0.39\frac{total\ words_i}{total\ sentences_i} + 11.8\frac{total\ syllables_i}{total\ words_i} - 15.59 \tag{3.1}$$

It is supposed to be equivalent to the US grade level of education and indicates the required years of education to be able to understand the respective text. The Flesch-Kincaid grade level approach can be applied to documents of arbitrary length. Consider for instance, the following – rather complex – sentence from Mario Draghi's introductory statement to the ECB press conference on 4 September 2014: *'The Eurosystem will purchase a broad portfolio of simple and transparent asset-backed securities (ABSs) with underlying assets consisting of claims against the euro area non-financial private sector under an ABS purchase programme (ABSPP)'.*[34] With 37 words and 68 syllables, the Flesch-Kincaid grade level score of this sentence is $0.39 \times 37/1 + 11.8 + 68/37 - 15.59 = 21$, suggesting that a person needs to be a professional reader for full comprehension.

We calculate the Flesch-Kincaid Grade Level score for all introductory statements using the quanteda package in R (Benoit, Watanabe, et al., 2018). To reduce the potential influence of outliers, we define the variable Complexity as the log of this score.

3.3.4 Measuring Trading Volume

To proxy financial market trading activity, we use trading volume of the EURO-STOXX-50 futures, since futures are highly liquid trading instruments that react quickly to new information (see, e.g. Kuttner, 2001; Bomfim, 2003). The underly-

[33] This approach is also applied in other fields of finance. For example, Smith and Taffler (1992), You and Zhang (2009), and Miller (2010) investigate the effect of complexity in corporate reports on subsequent trading volumes and stock-price movements. Loughran and McDonald (2016) discuss the use of textual analysis and linguistic measures in accounting and finance.

[34] https://www.ecb.europa.eu/press/pressconf/2014/html/is140904.en.html (accessed: 29 Aug 2020).

ing stock index, the EURO-STOXX 50 (ISIN: EU0009658145) with 50 large-cap constituents from the euro area, is one of the leading European stock indices. The corresponding future (ISIN: DE0009652388) is traded on the EUREX and, with a tick-size of 10, is widely considered the most liquid European stock index future.[35]

We retrieve trading volume at a 1-minute frequency from PortaraCQG and calculate $Volume_{Intro}$ as the natural logarithm of the mean trading volume (per minute) over the 15-minute window from 14:30 - 14:45 CET. This period reflects the start of the press conference and the average time span needed to read the introductory statement. Correspondingly, we define $Volume_{Q\&A}$ as the natural logarithm of the mean trading volume (per minute) measured during the roughly 60 minutes long Q&A-session (14:50 - 15:50 CET) and $Volume_{Conf}$ as the natural logarithm of the mean trading volume (per minute) over the period of the whole press conference (14:30 - 15:50 CET).

3.3.5 Descriptive Analysis

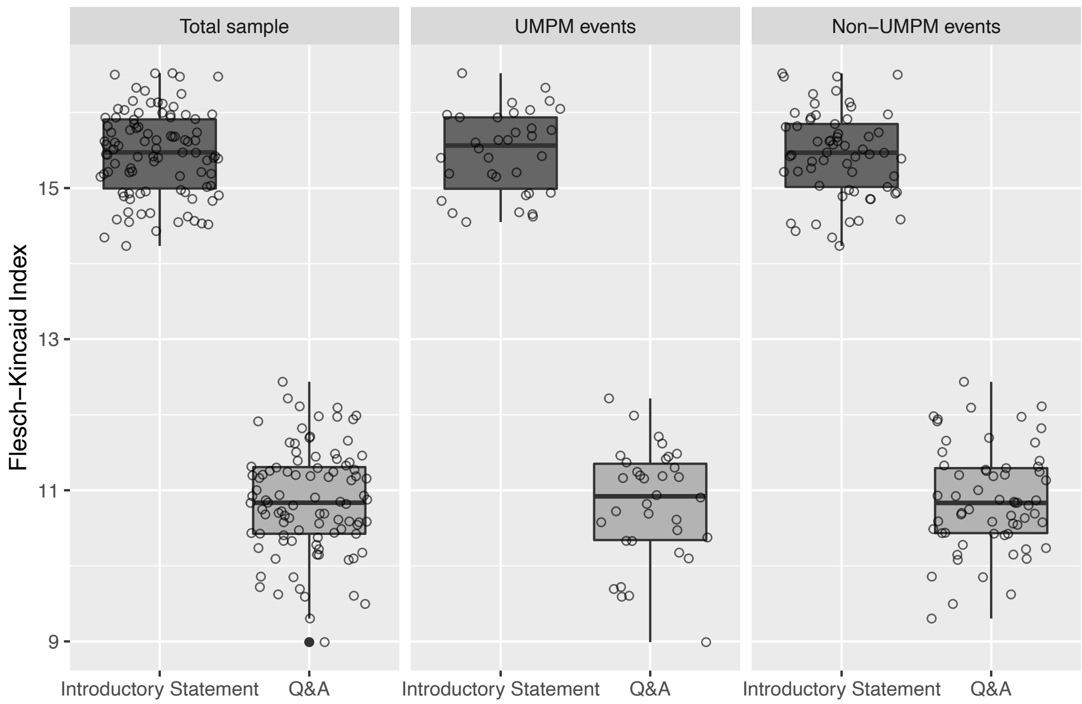
In this subsection, we provide descriptive statistics about the linguistic complexity of the introductory statements, as well as stylised facts and anecdotal evidence demonstrating the relevance of our hypotheses. Specifically, with regard to the later, we analyse (1) the trading activity around the press conferences, (2) the trading activity's temporal distribution, and (3) its relationship with respect to complexity of the introductory statements.

Linguistic Complexity of Introductory Statements Calculating the Flesch-Kincaid Grade Level score for every introductory statement in our sample, we find a mean score of 15.4 with a standard deviation of 0.6. This can be roughly interpreted as 15 years of education are required to comprehend and follow an

[35]According to Eurex Daily Statistics from 30 December 2016 and 29 December 2017, the average annual trading volume of the EURO-STOXX-50 futures was roughly 328 million contracts, corresponding to 10,474bn Euro and an average daily trading volume of 1.35 million contracts (Source: https://www.eurex.com/ex-en/data/statistics/trading-statistics, accessed 4 December 2020). In our sample, covering 2009-2017, the average trading during an introductory statement is some 4,600 contracts per minute, which is significantly more than the (time-of-the day pattern adjusted) average trading volume per minute (see Figure 3.1).

average introductory statement of the ECB. For all statements, the observed minimum and maximum values for the Flesch-Kincaid Grade Level are 14.2 and 16.5, respectively. These statistics correspond with the findings of Coenen, Ehrmann, et al. (2017) and demonstrate that the level of linguistic complexity of introductory statements is consistently high. A descriptive summary of the Flesch-Kincaid Grade Level is provided in Table C.2 in the Appendix. Interestingly, we do not find a significant difference between UMPM-events (15.5) and non-UMPM events (15.4).

Figure 3.1: Complexity distribution of the ECB's communication



Notes: Boxplot of Flesch-Kincaid Grade Level across introductory statements and Q&A sessions, with observed values illustrated as jitter plot. Differentiation between UMPM-events and non-UMPM events according to Table C.1 in the Appendix.

To illustrate the disparity in complexity between introductory statement and Q&A session, we also calculate the Flesch-Kincaid Grade Level score for the transcripts of the Q&A sessions. With an average Flesch-Kincaid Grade Level of above 15 for the introductory statement (independent of the type of event) and below 11 for the Q&A session (again, independent of the type of event), we discover a difference of more than 4 years of required education between the two forms of communication. Figure 3.1 shows these differences for all press conference, UMPM-events, and non-UMPM events, respectively.

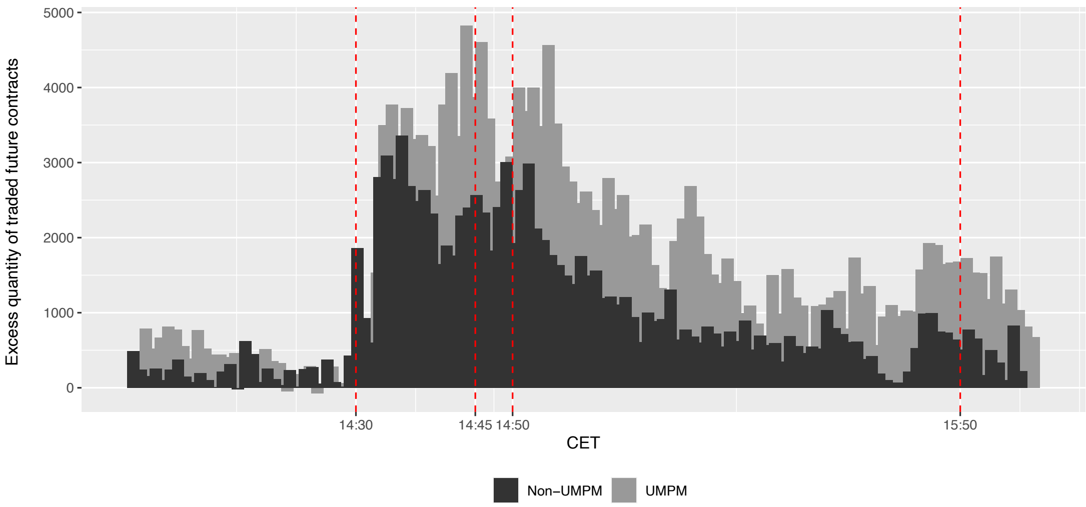
We can see that in all three samples, statements are clearly more linguistically complex than Q&A sessions. Moreover, the most complex Q&A session is less complex than the least complex introductory statement. This supports our argument that communication in Q&A sessions is less formal and thus less complex, and thus may help to improve the flow of information and encourage trading.

Excess Trading Patterns Next, we examine the EURO-STOXX-50 future trading activity during all GCM press conferences in our sample to understand its extent and temporal distribution. To exclude the effects stemming from common time-of-the-day patterns, we calculate excess trading volumes, defined as the difference between the mean trading volume per minute from all event days (i.e., press conference days) minus the mean volume per minute from non-event days (i.e., days without an ECB press conference). Figure 3.2 illustrates the mean excess trading volume for UMPM-events and non-UMPM events.

Three observations stand out. First, mean excess trading volume in stock index futures increases significantly a few minutes after the beginning of the GCM press conferences (14:30 CET). The pattern is consistent with previous work on the effects of ECB communication on financial markets (see, e.g., Andersson, 2010), and the view that the introductory statement conveys relevant news for financial markets.[36] Second, the mean excess trading volume remains high at the beginning of the Q&A session (at around 14:50 CET). This suggests that the Q&A session provides additional information to financial market participants. From 15:00 CET onwards, trading volumes slowly decrease until the end of the Q&A session around 15:50 CET, when trading activity reverts to near normal levels. Third, Figure 3.2 highlights considerable differences between trading volume during UMPM-events (grey) and non-UMPM events (black). In addition, the following conclusions can be drawn: (i) trading volume tends to be higher during UMPM-events, (ii) trading peaks later during UMPM events, and (iii) during the Q&A session, excess trading

[36]Note that these spikes in trading volumes are unlikely due to reactions to the Governing Council's interest rate decision, as the interest rate decision is communicated prior to the press conference at 13:45 CET.

Figure 3.2: Excess trading pattern

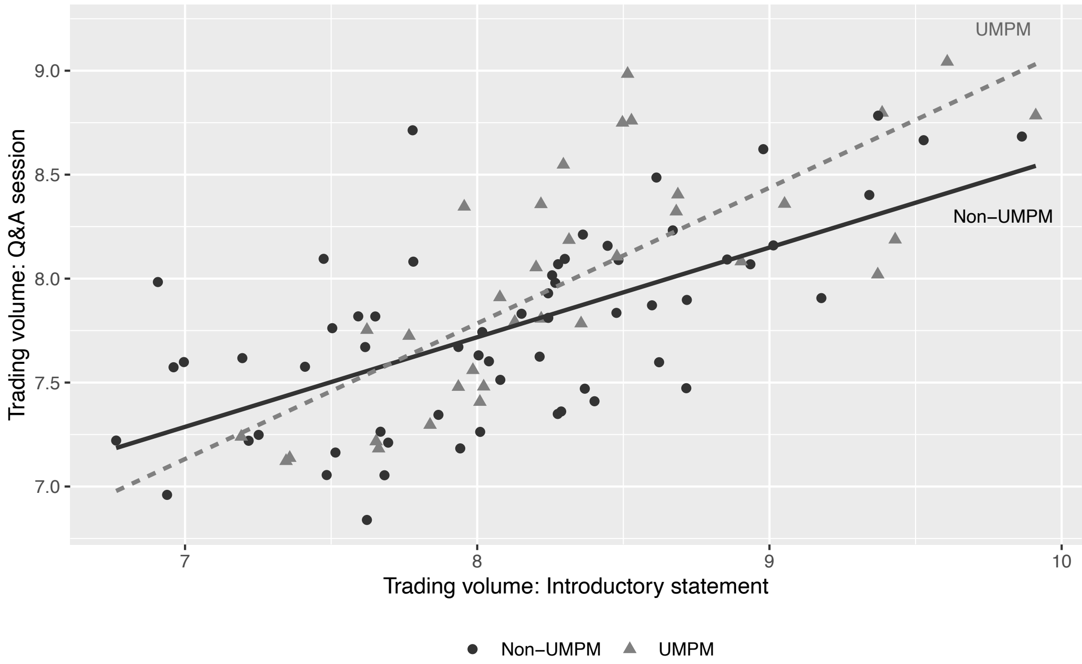


Notes: Mean excess trading volume in stock index futures during the analysed GCM press conferences. Calculation using mean excess trading volume for 1-minute intervals of the EURO-STOXX-50 Future across all GCM days between January 2007 and December 2017. Excess trading volume computed as mean EUREX trading volume across all GCM days minus mean EUREX trading volume on non-meeting days over the same period. Differentiation between UMPM-events and non-UMPM events according to Table C.1 in the Appendix.

volumes slow down faster for non-UMPM than for UMPM-events. This pattern is consistent with the view that financial market participants find UMPM-events relatively more difficult to understand than non-UMPM events, which is why they temporarily underreact. The hike in trading activity especially during UMPM-events suggests that the less complex Q&A session provides valuable information for market participants too.

Temporal Distribution of Trading Volumes Next, we assess whether the temporal distribution of trading activity illustrated in Figure 3.2 is representative for all events in our sample or whether it is simply a product of aggregation over time. For each of our UMPM-events and non-UMPM events, Figure 3.3 plots the (logarithm of the) average trading volume during the Q&A session versus the (logarithm of the) average trading volume during the introductory statement as well as the corresponding event-specific regression lines. Three main patterns are evident from Figure 3.2. First, there is a positive correlation between the trading volumes in the two periods. This relationship is statistically significant in both cases. Second, qualitatively, we find a steeper slope for the regression line

Figure 3.3: Temporal distribution of trading volumes

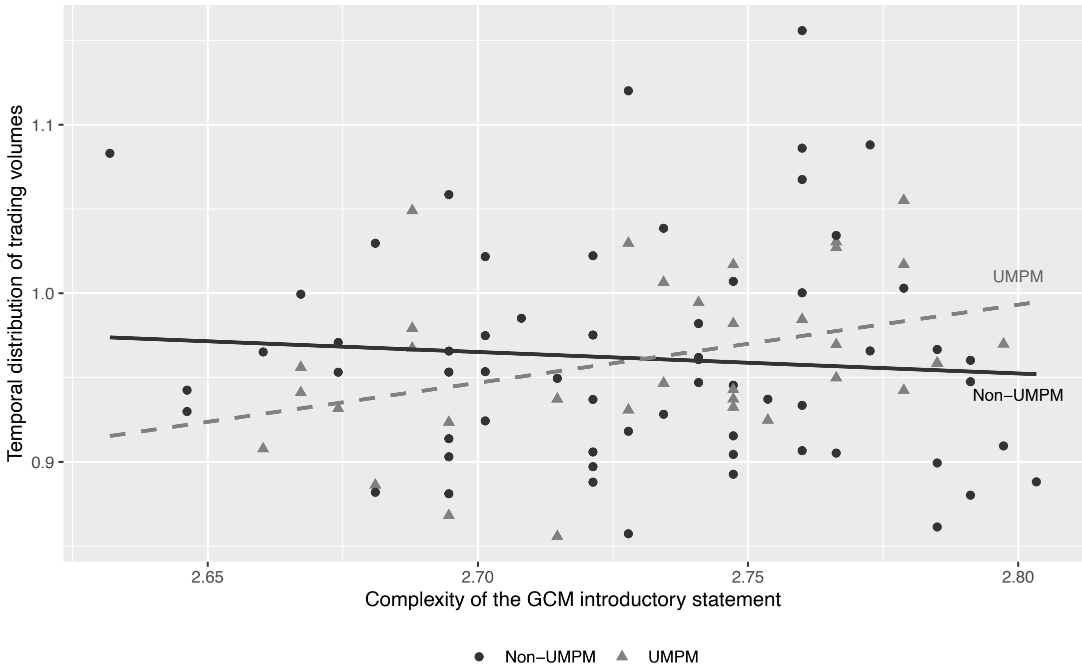


Notes: Cross-plot of trading volumes across introductory statements and Q&A sessions, with observed values and a fitted regression line. Calculated using the natural logarithm of average (mean) minute trading volume of the EURO-STOXX-50 Future on EUREX all GCM days between January 2009 and December 2017. Separation between 14:30-14:45 introductory statement and 14:50-15:50 Q&A session. Differentiation between UMPM-events and non-UMPM events according to Table C.1 in the Appendix.

representing UMPM-events, suggesting that financial markets react with delayed trading in case of UMPM events. Third, the quantitative relationship between trading volume during the introductory statement and during the Q&A session can vary substantially. This indicates that the relationship is not perfectly linear and, therefore, further variables appear to be relevant.

Trading Volumes and Complexity Finally, we shift our focus to the ease of understanding of the transmitted information. We analyse whether there is a relationship between the temporal distribution of trading volumes and the linguistic complexity of introductory statements. In order to capture the temporal distribution of trading volumes in a single variable, we calculate the difference of the logarithm of the average trading volume per minute during the introductory statement, $Volume_{Intro}$, to the respective value for the Q&A session, $Volume_{Q\&A}$.

Figure 3.4: Trading volumes and complexity



Notes: Temporal distribution of trading volumes and complexity. Calculation based on the ratio of average (mean) minute trading volumes of the EURO-STOXX-50 Future on EUREX over GCM days with during the 14:50 - 15:50 Q&A session divided by the 14:30 - 14:45 introductory statement, with the natural logarithm applied to the fraction. Communication complexity of GCM introductory statements is measured by the natural logarithm of the Flesch-Kincaid Grade Level. Differentiation between UMPM-events and non-UMPM events according to Table C.1 in the Appendix.

Figure 3.4 plots this log difference against the linguistic complexity for each event. For UMPM-events, the relationship between the complexity of the introductory statements and the temporal distribution of trading volumes is positive, whereas it is slightly negative for non-UMPM events. This pattern is consistent with the view of a positive association between complexity and delayed trading. In other words, we find evidence of an underreaction of the market in the case of UMPM-events.

3.4 Regression Analysis

3.4.1 Empirical Design

To formally test our hypotheses, we estimate versions of the following regression:

$$\begin{aligned} V_t = & \alpha + \beta_1 \times Complexity_t + \beta_2 \times Complexity_t \times UMPM_t \\ & + \beta_3 \times UMPM_t + \gamma \times Controls_t + \epsilon_t, \end{aligned} \tag{3.2}$$

where V_t measures trading behaviour at the event t (i.e., the ECB press conference following the Governing Council Meetings), $Complexity_t$ the linguistic complexity of the event's introductory statement, and $UMPM_t$ the type of event. α represents a constant, $Controls_t$ a vector of control variables, and ϵ_t the error term. Table 3.1 provides further details of our variable definitions. The β_i's are our coefficients of interest. Specifically, arguing that UMPM-events are more complex in context and content, it is β_2, the coefficient of the interaction term, which is geared to reflect our argument that it is not only the linguistic complexity of the transcripts that matters, but also the complexity of the context and content.

We add three control variables to our regression. First, based on Kuttner (2001), we capture the surprise effect in conventional monetary policy by long-term Bond Returns. We use the log-return of the 10-year BUND future as traded on EUREX during 13:44 - 14:29 CET. Second, we use a Rate Change Dummy, which indicates whether the ECB announced a change in its de-posit facility rate at 13:45 CET. Third, we include Δ *Shadow Prime Rate*, which captures monetary tightening as conveyed in the ECB's communications. In line with Hayo, Henseler, and Rapp (2019), we calculate this measure using the Wordscores approach (Laver, Benoit and Garry, 2003), calibrated by using introductory statement transcripts of GCM press conferences from 1999 - 2006 and corresponding changes in the deposit facility rate.

3.4.2 Regression Analysis

To assess the first two hypotheses, we regress statement complexity on trading volume. Table 3.2 reports the results for two measures of trading behaviour,

Table 3.1: Overview of variable definitions

Dependent variables		
V_t	$Volume_{Intro}$	ln(mean minute volume$_{14:30-14:45}$)
	$Volume_{Q\&A}$	ln(mean minute volume$_{14:50-15:50}$)
	$Volume_{Conf.}$	ln(mean minute volume$_{14:30-15:50}$)
D_t	$Volume_{Q\&A-to-Intro}$	ln((mean minute volume$_{14:50-15:50}$) /(mean minute volume$_{14:30-14:45}$))
	$Volume_{Q\&A-to-Conf.}$	ln((mean minute volume$_{14:50-15:50}$) /(mean minute volume$_{14:30-15:50}$))
Independent variables		
$Complexity_t$		Flesch-Kincaid Grade Level for GCM introductory statements, calculated as: 0.39 × WS + 11.8 × SW - 15.59 WS = Total number of words divided by total number of sentences SW = Total number of syllables divided by total number of words
Control variables		
$Controls_t$	Bond Return	ln(Price[14:29]/Price[13:44]), of EUREX traded EURO-BUND Futures
	Rate Change Dummy	Deposit facility rate change announced at 13:45 (yes=1/no=0)
	ΔShadow Prime Rate	Calculated using Wordscores, calibrated based on introductory statement transcripts of GCM press conferences in 1999-2006 and corresponding changes in the deposit facility rate

Note: A descriptive summary of all variables can be found in Table C.2 in the Appendix.

$Volume_{Intro}$ and $Volume_{Conf}$, where $Volume_{Intro}$ measures trading volume during the introductory statement and $Volume_{Conf}$ during the aggregate press conference.[37] For each of the volume measures, we estimate three specifications. The results can be summarised as follows. First, Specification (1) and (4) reveal that linguistic complexity of introductory statements is negatively (but insignificantly) correlated with contemporaneous trading activity. While this is consistent with our first hypothesis (H1), the coefficients are far from significant. Essentially, the results from these two specifications suggest that overall for the period 2009-2017 linguistic complexity of introductory statements uncorrelated with contemporaneous trading activity, which is in contrast to the findings of

[37]We confirm the results presented here in unreported tests, where we use (i) corresponding measures of excess trading volume and (ii) an alternative measure for $Volume_{Intro}$, which we define as the logarithm of average trading volume defined over the period 14:35-14:45 CET aiming to get rid of potential noise trading and make sure we capture the effect of the introductory statement only.

Table 3.2: Analysis of trading volume

	Dependent variable:					
	$Volume_{Intro}$			$Volume_{Conf}$		
	(1)	(2)	(3)	(4)	(5)	(6)
Complexity	−1.23	2.68	2.73	−0.97	1.45	1.46
	(1.77)	(2.05)	(1.98)	(1.37)	(1.61)	(1.51)
Complexity×UMPM		−11.88***	−9.87***		−7.43***	−5.57**
		(3.53)	(3.61)		(2.77)	(2.76)
UMPM		32.69***	27.12***		20.55***	15.35**
		(9.64)	(9.89)		(7.57)	(7.56)
Bond Return			0.57			0.47*
			(0.35)			(0.27)
Rate Change Dummy			0.32			0.35**
			(0.20)			(0.15)
Δ Shadow Rate			0.26			0.25*
			(0.19)			(0.14)
Constant	11.54**	0.76	0.67	10.59***	3.88	3.89
	(4.84)	(5.59)	(5.41)	(3.76)	(4.39)	(4.14)
Observations	96	96	95	96	96	95
R^2	0.01	0.14	0.21	0.01	0.12	0.22

Note: Coefficients are estimated using an OLS regression. Standard errors are displayed in parentheses. ***, **, * indicate significance at the 1, 5, and 10 per cent level, respectively.

Smales and Apergis (2017a) and Smales and Apergis (2017b) for FOMC statements.

Second, Specification (2) and (5) document an event-differentiated correlation of linguistic complexity with trading activity in financial markets. While the coefficient of Complexity is positive (but insignificant), the coefficient of the interaction term $Complexity_t \times UMPM_t$ is negative and highly significant. The sum of the coefficients, i.e. $\beta_1 + \beta_2$, in Specification (2) is -9.2 (with a standard deviation of 2.47) and highly significant (p<.01).[38] This is not only consistent with You and Zhang (2009) and Miller (2010) who propose a negative relationship between information complexity and trading behaviour and our second hypothesis (H2), but also economically meaningful: Specifically, Specification (2) suggests that an increase in complexity by 1% is associated with a decrease in trading volume by

[38]For Specification (5) the coefficients add up to -5.98 with a standard deviation of 2.37, which is significant at the 5%-level.

up to 9%, or some 420 contracts per minute. Relatedly, a hike in the Flesch-Kincaid Grade Level index by one year beyond the average (i.e., from 15.4 to 16.4) is on average accompanied by a reduction in trading volumes by some 2,760 contracts per minute (about 75% of the standard deviation).

Third, Specification (3) and (6) document that these results remain intact, when we add our control variables. However, the coefficients of the interaction term decrease and thus the estimated correlation of linguistic complexity with trading activity in case of an UMPM event. Specifically, the sum of the coefficients β_1 and β_2 are -7.1 and -4.1 in Specification (3) and Specification (6), respectively. Finally, looking at the coefficients β_2 and β_3 of Specification (2), we find, consistent with Figure 3.2, that for an UMPM-event of average Complexity, which is -2.7, contemporaneous trading volume is about 23% (p<.10) higher than for a non-UMPM event with similar Complexity. In sum, the results from Table 3.2 are consistent with our argument that it is not only the linguistic complexity of the transcripts that matters, but also the complexity of the context and content that matters and a market that underreacts to complex central bank communication.

To assess our third hypothesis, we turn to the temporal distribution of trading activity. Table 3.3 reports results for two measures of temporal distribution of trading activity, $Volume_{Q\&A-to-Intro}$ and $Volume_{Q\&A-to-Conf}$, which are defined as the difference between $Volume_{Q\&A}$ and $Volume_{Intro}$ and $Volume_{Q\&A}$ and $Volume_{Conf}$, respectively. Again, for each of the measures, we estimate three specifications.

The results can be summarised as follows. First, Specification (1) and (4) reveal that linguistic complexity of introductory statements is positively (but insignificantly) correlated with delayed trading activity. While this is consistent with our third hypothesis (H3), the coefficients are far from significant. Second, again we find an event-differentiated correlation of linguistic complexity with trading activity in financial markets.[39] While in Specification (2) and (5) the coefficient of Complexity is negative (but insignificant), the coefficient of the interaction term $Complexity_t \times UMPM_t$ is positive and significant. The sum of the coefficients,

[39] We confirm the results presented here in unreported tests, where we also control for the complexity of Q&A statements and allow the complexity of Q&A statements to interact with UMPM.

Table 3.3: Analysis of the temporal distribution of trading volume

	Dependent variable:					
	$Volume_{Q\&A-to-Intro}$			$Volume_{Q\&A-to-Conf.}$		
	(1)	(2)	(3)	(4)	(5)	(6)
Complexity	0.62	−1.38	−1.44	0.36	−0.15	−0.18
	(1.25)	(1.51)	(1.56)	(0.34)	(0.41)	(0.43)
UMPM_dummy		−16.22**	−15.82**		−4.07**	−4.05*
		(7.14)	(7.80)		(1.94)	(2.12)
Complexity×UMPM		5.94**	5.80**		1.50**	1.49*
		(2.61)	(2.85)		(0.71)	(0.78)
Bond Return			−0.09			0.01
			(0.28)			(0.08)
Rate Change Dummy			0.02			−0.001
			(0.16)			(0.04)
Δ Shadow Rate			−0.01			−0.01
			(0.15)			(0.04)
Constant	−2.03	3.42	3.59	−1.08	0.30	0.37
	(3.42)	(4.13)	(4.27)	(0.93)	(1.12)	(1.16)
Observations	96	96	95	96	96	95
R^2	0.003	0.06	0.06	0.01	0.06	0.06

Note: Coefficients are estimated using an OLS regression. Standard errors are displayed in parentheses. ***, **, * indicate significance at the 1, 5, and 10 per cent level, respectively.

i.e. $\beta_1 + \beta_2$, in Specification (2) is 4.6 (with a standard deviation of 1.8) and significant (p<.05).[40] Third, Specification (3) and (6) document that these results remain intact, when we add our control variables. However, the coefficients of the interaction term are slightly lower. Finally, looking at the coefficients β_2 and β_3 of Specification (2), we find no significant difference between UMPM-events and non-UMPM events for average Complexity (p>.30).

In sum, the results from Table 3.3 again are consistent with our argument that markets underreact to complex central bank communication and delay trading. Moreover, the results suggest that trading is not uniformly delayed, but – for UMPM events – gains momentum with the beginning of the Q&A session, which supports our third hypothesis (H3).

[40]For Specification (5) the coefficients add up to 1.35 with a standard deviation of 0.52, which again is significant at the 5%-level.

3.4.3 Additional Analysis

In this section, we aim to shed some light on the difference between UMPM-events and non-UMPM events. Therefore, we investigate whether UMPM-announcements contain more novel information than standard announcements, i.e. whether they are more 'unconventional'. Specifically, we employ Amaya and Filbien's (2015) similarity index to assess the degree of homogeneity between different statements.

To calculate the index, we (i) remove all numbers, dates, and stop words, and (ii) construct word bi-grams (two-word combinations) in order to capture combined expressions, for example, 'quantitative easing'. We calculate the cosine similarity of two subsequent introductory statements for all events in our sample, as follows:

$$Similarity_t = \frac{\sum_{b=1}^{B} fr_{b,t} \times fr_{b,t-1}}{\sqrt{\sum_{b=1}^{B} fr_{b,t}^2}\sqrt{\sum_{b=1}^{B} fr_{b,t-1}^2}}, \tag{3.3}$$

where B represents the total number of unique bi-grams in all press releases and $fr_{b,t}$ and $fr_{b,t-1}$ are the frequencies of bi-gram b in press releases t and $t-1$, respectively.

To illustrate the idea of our similarity measure, consider the following two sentences: *'The Governing Council expects the euro area economy to grow at a moderate pace in 2010'* and *'We expect price stability to be maintained over the medium term, thereby supporting the purchasing power of euro area households'* from two introductory statements from 2010. They contain one shared bigram (euro_area) and 34 unique bigrams (the_governing, governing_council, council_expects, ...). The similarity index value of those two sentences is 1/35=0.03. Comparing longer texts tends to increase the value of the index, as the probability of recurring bigrams rises. In our sample, the similarity index has an average score of 0.44, indicating that 44% of all bigrams in an introductory statement occurred in the previous one too. To assess whether UMPM announcements differ from non-UMPM announcements, we run the following regression:

$$Similarity_{(t\ and\ t-1)} = \alpha + \beta \times UMPM_t + \gamma \times Similarity_{(t\ and\ t-2)} + \epsilon_t, \tag{3.4}$$

Note that Amaya and Filbien (2015) find that ECB introductory statements become more similar over time. To capture this development, we include a delayed similarity index as a regressor. It is based on comparing the content of the current statement with the text of the statement in t-2. The Durbin-Watson Test supports our choice of the delayed similarity term.

Table 3.4 reports the estimation results. The UMPM-Dummy is statistically significant and economically relevant. Statements with UMPM announcements are 3% less similar to the previous period statements than statements without UMPM announcements. Given an average of 1420 bigrams, this increases the number of unique bigrams by 45. Since UMPM announcements do not significantly differ in length from other announcements, they appear to contain more 'novel' information. Arguably, it is this new information that drives the previous results. That is, through the deviation from 'standard' announcements, complexity increases traders' cognitive costs, which causes them to postpone their trading decisions to the 'easier' Q&A session. These findings do not change when controlling for the previous event type (i.e. using a lagged UMPM-Dummy and interaction terms).[41]

Table 3.4: Similarity analysis

	Dependent variable:
	$\text{Similarity}_{(t\ and\ t-1)}$
UMPM-Dummy	-0.03**
	(0.01)
$\text{Similarity}_{(t\ and\ t-2)}$	0.87***
	(0.07)
Constant	0.14**
	(0.02)
Observations	96
R^2	0.66

Note: Coefficients are estimated using an OLS regression. Standard errors are displayed in parentheses. ***, **, * indicate significance at the 1, 5, and 10 per cent level, respectively.

In a next step, we analyse the similarity between introductory statement and Q&A session. First, there is a remarkably strong similarity between the two

[41] All omitted results here and elsewhere in the paper are available on request.

institutionalised forms of communication. Given that Figure 3.1 suggests considerable differences in linguistic structure, an average similarity index of about 0.2 provides empirical evidence that the substance of the Q&A session is close to the preceding introductory statement. Second, the degree of similarity between introductory statement and Q&A session does not differ during UMPM-events, which suggests that traders can generally rely on Q&A questions to clarify the more complex content of the introductory statement.

These findings complement our previous results, namely (1) linguistic complexity of Q&A sessions is lower than that of introductory statements, (2) for UMPM-events with high linguistic complexity trading is delayed to Q&A sessions, and (3) the similarity between subsequent press conferences is lower for UMPM-events. Thus, we discover empirical evidence supporting the following transmission process from statement complexity to financial market trading behaviour: Traders know that they can reliably depend on the context of the introductory statements being elaborated on in the subsequent Q&A. They realise that introductory statements referring to UMPM's are complex and contain relatively more novel information. While this causes them to underreact to the new information, the discussion and clarification of the cognitively costly content during the subsequent Q&A session mitigates this effect. An outcome of this process is that parts of the trading shifts from the statement phase to the Q&A phase of the ECB's press conference when UMPM's are discussed.

3.5 Robustness of Results

As robustness tests, we (1) increase the time horizon, (2) address the concept of vagueness in our complexity metric, (3) consider alternative measures for the latent variable of complexity, and (4) determine complexity via factor analyses based on multiple complexity measures as well as further communication-related measures.

3.5.1 Time Horizon

To incorporate events prior to the period of the effective lower bound, we increase the observation period to January 2003 until December 2017. This extension

Table 3.5: Robustness check - Time horizon

	Dependent variable:			
	Volume_{Intro}	Volume_{Conf}	$\text{Volume}_{Q\&A-to-Intro}$	$\text{Volume}_{Q\&A-to-Conf.}$
	H1		H2	
	(1)	(2)	(3)	(4)
Complexity	4.39***	3.16***	−1.72**	−0.49**
	(1.11)	(0.92)	(0.76)	(0.21)
Complexity×UMPM	−13.35***	−9.17***	5.99**	1.81***
	(3.58)	(2.96)	(2.44)	(0.68)
UMPM	37.14***	25.67***	−16.44**	−4.96***
	(9.79)	(8.09)	(6.67)	(1.86)
Bond Return	0.53	0.35	−0.21	−0.03
	(0.34)	(0.28)	(0.23)	(0.06)
Rate Change	0.21	0.26*	0.06	0.02
	(0.16)	(0.14)	(0.11)	(0.03)
ΔShadow Rate	−0.14	−0.13	0.01	0.005
	(0.19)	(0.15)	(0.13)	(0.04)
Constant	−4.32	−1.16	4.41**	1.25**
	(3.02)	(2.50)	(2.06)	(0.57)
Observations	163	163	163	163
R^2	0.24	0.24	0.07	0.06

Note: Coefficients are estimated using an OLS regression. Standard errors are displayed in parentheses. ***, **, * indicate significance at the 1, 5, and 10 per cent level, respectively.

roughly doubles the number of observations to around 163 press conferences. The estimation results for extending the sample are presented in Table 3.5 and demonstrate that our previous findings are robust. In addition to the earlier results and similar to the results of Smales and Apergis (2017a) and Smales and Apergis (2017b) for the FOMC, for non-UMPM-events the relationship between complexity and trading volume is now statistically significant.

3.5.2 Complexity or Vague Talk

Next, we examine the possibility that the Flesch-Kincaid Grade Level metric may better be interpreted as an indicator for vagueness rather than complexity. We argue that complexity is a proxy for the cognitive cost of comprehending the content. However, vagueness also generates information that is difficult to follow, but originates from a lack of clarity. The Flesch-Kincaid Grade Level index consists of two components, the average length of a sentence (WS) and the average word length (SW):

$$FK_i = 0.39 \underbrace{\frac{total\ words_i}{total\ sentences_i}}_{WS} + 11.8 \underbrace{\frac{total\ syllables_i}{total\ words_i}}_{SW} - 15.59 \tag{3.5}$$

Longer sentences, i.e. higher WS, may be associated with both, more complexity and more vagueness, whereas the use of longer words, i.e. higher SW, should only affect comprehensibility. In other words, SW measures complexity but not vagueness, whereas WS is a representation of complexity and vagueness.

Table 3.6: Robustness check - Vague talk

	Dependent variable:							
	Volume_{Intro}		Volume_{Conf}		$\text{Volume}_{Q\&A,Intro}$		$\text{Volume}_{Q\&A,Conf.}$	
	H2				H3			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
SW	1.76		1.49		−0.26		0.003	
	(1.39)		(1.04)		(1.09)		(0.30)	
WS		−3.48		−6.07		−2.59		0.001
		(6.81)		(5.08)		(5.31)		(1.44)
SW×UMPM	−5.71**		−3.90**		2.44		0.62	
	(2.59)		(1.95)		(2.04)		(0.56)	
WS×UMPM		−15.08		−7.51		11.30		3.74
		(14.47)		(10.80)		(11.29)		(3.06)
UMPM	18.47**	9.01	12.64**	4.51	−7.80	−6.68	−1.96	−2.18
	(8.34)	(8.61)	(6.28)	(6.42)	(6.57)	(6.72)	(1.79)	(1.82)
Δ Shadow Rate	0.29	0.30	0.26*	0.23	−0.06	−0.08	−0.03	−0.02
	(0.19)	(0.20)	(0.14)	(0.15)	(0.15)	(0.16)	(0.04)	(0.04)
Bond Return	0.71**	0.65*	0.55**	0.52*	−0.18	−0.13	−0.02	0.004
	(0.35)	(0.37)	(0.26)	(0.28)	(0.28)	(0.29)	(0.08)	(0.08)
Rate Change Dummy	0.34*	0.31	0.36**	0.35**	0.01	0.06	−0.005	0.01
	(0.20)	(0.21)	(0.15)	(0.16)	(0.16)	(0.16)	(0.04)	(0.04)
Constant	2.52	10.22**	3.12	11.53***	0.48	1.20	−0.12	−0.11
	(4.42)	(4.11)	(3.33)	(3.06)	(3.49)	(3.20)	(0.95)	(0.87)
Observations	95	95	95	95	95	95	95	95
R^2	0.18	0.16	0.22	0.21	0.03	0.02	0.04	0.04

Note: Coefficients are estimated using an OLS regression. Standard errors are displayed in parentheses. ***, **, * indicate significance at the 1, 5, and 10 per cent level, respectively.

Disaggregating the Flesch-Kincaid Grade Level into these two components, we discover that the correlation between SW and the complete index is almost 90%, compared to around 10% for WS. Furthermore, if we include SW and WS in our regression model (see Table 3.6), we find that SW (i.e. 'complexity') appears to drive our results rather than WS ('vagueness'). Qualitatively, this conclusion holds for both hypotheses but only the estimates for H2 are statistically significant.

3.5.3 Alternative Measures of Complexity

Table 3.7: Definitions of Alternative Complexity Measures

Complexity measures	

Flesch Reading Ease (inverted)	$1/(206.835 - 1.015 \times WS - 84.6 \times SW)$
	WS = #words divided by #sentences;
	SW = #syllables divided by #words

Gunning Fog Index	$0.4 \times WS + 40 \times CWW$
	WS = #words divided by #sentences;
	CWW = #complex words divided by #words

SMOG Index	$1.0430 \times \sqrt{PS \times (30/S)} + 3.1291$ PS = #polysyllables (3 or more syllables); S = #sentences

Coleman-Liau Index	$5.88 * AL + (0.296 \times Nst/Nw) - 15.8$
	AL = Average #letters per 100 words;
	AS = Average #sentences per 100 words

Automated Readability Index	$4.71 \times (C/W) + 0.5 \times (W/S) - 21.43$ C = #characters; W = #words; S = #sentences

Note: We use the inverse of the Flesch Reading Ease, so as to ensure that for all indicators larger values represent a higher degree of complexity.

Our operationalisation of the latent complexity variable in the form of the Flesch-Kincaid Grade Level follows Smales and Apergis (2017a) and Smales and Apergis (2017b). To demonstrate that our results do not depend on this choice, we employ a variety of alternatives. The most common measures for linguistic complexity are the Flesch Reading Ease (Flesch, 1948), the Gunning Fog Index (Gunning, 1952), the SMOG Index (McLaughlin, 1969), the Coleman-Liau Index (Coleman and Liau, 1975), and the Automated Readability Index (Senter and Smith, 1967). Table 3.7 sets out the respective definitions. Table 3.8 reports the estimated coefficients for Equation (3.1) for the various complexity measures indicators. We include all control variables, but only report the coefficients for the interaction term between UMPM and the respective complexity measurement.

Regardless of the underlying complexity definition, the coefficients have the expected sign and most of them are significant at the 10% level or below. Thus, we conclude that our results are generally robust with regard to the definition of complexity.

Table 3.8: Coefficients for alternative measures of complexity

	Dependent variable:	
	Volume$_{Intro}$ H2	Volume$_{Q\&A-to-Intro}$ H3
Flesch Kincaid Grade Level	-12.09***	5.42*
	(3.91)	(3.06)
Flesch Reading Ease	-5.83***	3.00*
	(2.04)	(1.59)
Gunning Fog Index	-12.69***	6.32*
	(4.23)	(3.29)
SMOG Index	-15.67***	7.73*
	(5.21)	(4.06)
Coleman-Liau Index	-7.65	1.37
	(5.66)	(4.33)
Automated Readability Index	-10.36***	3.69
	(3.23)	(2.54)

Note: Coefficients are estimated using an OLS regression. Standard errors are displayed in parentheses. ***, **, * indicate significance at the 1, 5, and 10 per cent level, respectively.

3.5.4 Complexity Approximated via Factor Analyses

Since all indicators in the previous section are supposed to measure the same latent variable, it is appropriate to approximate complexity using factor analysis. We employ two sets of underlying variables. In a first step, we conduct a factor analysis using the six complexity measures discussed above and find two common factors (Eigenvalue > 1). In a second step, we conduct another factor analysis containing an additional set of variables quantifying communication. An overview of these additional variables is provided in Table 3.9. Combining the six

Table 3.9: Additional Variables for Quantifying Communication - Related Aspects

Communication measures	
Future-Orientation	% future adjectives
Uncertainty	% uncertainty verbs
Active/Passive	(% active verbs - % passive verbs)+1
Overstated/Understated	(% overstated verbs - % understated verbs)+1
Positive/Negative	(% positive verbs - % negative verbs)+1
Positive/Negative	(% positive verbs - % negative verbs)+1 [Loughran-McDonald definition]
Strong/Weak	(% strong verbs - % weak verbs)+1

complexity indicators with the seven additional variables, we find three common factors (Eigenvalues > 1). The six complexity indicators primarily load on the first factor. The communication indicators Future, Positive/Negative and Active/Passive mainly load on the second factor, while the remaining ones tend to load on the last factor. We use both factors from the first factor analysis and the three factors from the second factor analysis to re-estimate Equation (3.1), with V_t defined as $Volume_{Intro}$ and $Volume_{Q\&A-to-Conf}$. The results are reported in Table 3.10.

Table 3.10: Coefficients for complexity measures based on factor analysis

	Dependent variable:	
Panel A: Factor Analysis (Complexity Measures)	Volume_{Intro} H2	$\text{Volume}_{Q\&A-to-Intro}$ H3
Factor 1×UMPM-Dummy	-0.42***	0.06*
	(0.15)	(0.03)
Factor 2×UMPM-Dummy	-0.26	0.05
	(0.19)	(0.04)
Observations	91	91
R^2	0.29	0.11
Panel B: Factor Analysis (Complexity + Add. Measures)	H2	H3
Factor 1×UMPM-Dummy	-0.48***	0.06*
	(0.16)	(0.03)
Factor 2×UMPM-Dummy	0.24	-0.03
	(0.15)	(0.03)
Factor 3×UMPM-Dummy	-0.05	0.03
	(0.16)	(0.04)
Observations	91	91
R^2	0.29	0.11

Note: Coefficients are estimated using an OLS regression. Standard errors are displayed in parentheses. ***, **, * indicate significance at the 1, 5, and 10 per cent level, respectively.

Consistent with our previous results in Table 3.2 and Table 3.3, the factor capturing complexity has a significantly negative coefficient for H2 and a significantly positive one for H3. In contrast, the two other factors reflecting the additional communication-related aspects are insignificant. We therefore conclude that our findings are also robust to complexity approximated via a factor analysis as well as with regard to other aspects of communication.

3.6 Conclusion

In this paper, we assess the effects of central bank communication complexity on trading behaviour of financial market participants. Our analysis covers the official ECB press conference following regular GCMs between January 2009 and December 2017, during which unprecedented UMPM substantially increased communication complexity. Examining the transcripts of the introductory statements and using high-frequency data on European stock index futures, we investigate whether complexity of ECB communication affects contemporaneous trading in financial markets.

Our findings can be summarised as follows. First, differentiating between UMPM-events and non-UMPM-events, we do not find evidence for any difference in the linguistic complexity of introductory statements. Second, we discover a negative relationship between ECB communication complexity and contemporaneous trading volume during events where unconventional monetary policy is discussed. This event-differentiated underreaction of the market suggests, that when the ECB shares information with financial markets, it is not only the linguistic complexity of the communication that matters, but also the complexity of the context and content. To support this view, we demonstrate that more 'novel' information is transmitted during UMPM-related press conferences than during other press conferences. These findings, which are in line with results reported by You and Zhang (2009) and Miller (2010) and consistent with the argument that investors underreact to cognitively costly/complex information (Hirshleifer, 2001; Hong and Stein, 1999; McEwen and Hunton, 1999), extent the findings of Smales and Apergis (2017a) and Smales and Apergis (2017b) for the Federal Reserve.

Finally, we shed some light on the question of whether the ECB's Q&A-sessions may help to mitigate the underreaction of the market. Consistent with that view, we find a positive relationship between the complexity of ECB communication in UMPM-events and a shift of trading activity from introductory statement to Q&A session. Going forward, promising future research could focus on the question of what drives complexity of central bank communication and whether, in case of the ECB, a shift of trading activity to Q&A sessions can be explained by

Q&A sessions effectively mitigating complexity issues. In addition, it would be interesting examine whether our findings apply to other central banks and other forms of central bank communication. This could help to identify best practices for central banks' communication strategies vis-a-vis financial markets.

4 Whatever it takes to understand a central banker - Embedding their words using neural networks

Martin Baumgärtner[f] and Johannes Zahner[b]

Abstract

Dictionary approaches are at the forefront of current techniques for quantifying central bank communication. This paper proposes embeddings – a language model trained using machine learning techniques – to locate words and documents in a multidimensional vector space. To accomplish this, we gather a text corpus that is unparalleled in size and diversity in the central bank communication literature, as well as introduce a novel approach to text quantification from computational linguistics. Utilizing this novel text corpus of over 23,000 documents from over 130 central banks we are able to provide high quality text-representations –embeddings– for central bank communication. Finally, we demonstrate the applicability of embeddings in this paper by several examples in the fields of monetary policy surprises, financial uncertainty, and gender bias.

Keywords: Word Embedding, Neural Network, Central Bank Communication, Natural Language Processing, Transfer Learning

JEL classification: C45, C53, E52, Z13

[f]THM Business School, JLU Gießen, Germany

[b]Corresponding author. School of Business and Economics, Institutional Economics Research Group, Philipps-Universität Marburg, Germany.

4.1 Introduction

Over the last few decades, there has been an increase in the analysis and interpretation of central bank communication (Blinder, Ehrmann, et al., 2008). This development was accelerated by the emergence of the zero lower bound and the emergence of forward guidance, wherein central bankers recognized the possibility to complement actions with well-placed language to steer market participants towards a desired equilibrium path. As a result, central banks increased their communication substantially. The FOMC, for example, started publishing press conferences since 2011, and the ECB began disclosing monetary policy meeting minutes in 2015.

Consequently, a substantial body of literature has emerged that developed methods for quantifying different aspects of such communication. The literature is still primarily driven by dictionary approaches, in which pre-defined dictionaries, such as Loughran and McDonald (2011), Apel and Grimaldi (2014), and Picault and Renault (2017), are used to count terms (for example, positive and negative ones) to extract a single dimension (for example, sentiment) from a text corpus. However, these methods fall short of incorporating the language's richness, its multidimensionality, and its context-dependence. Moreover, dictionary approaches are inherently subjective, as discussed in Gentzkow, Kelly, et al. (2019, p. 554) survey on text mining in economics, where the authors emphasize that *"dictionary-based methods heavily weigh prior information."*

To address these weaknesses, we turn to linguistic and computer science, using machine learning tools to develop a novel *language model*. Such a model can be estimated from a set of texts – the *corpus* –, and an *algorithm* that locates words a multidimensional vector space. In this vector space, conceptually similar terms are mapped in close proximity, reflecting meaningful relationships. Following recent advances in computational linguistics, we propose (word) embeddings as a novel language model for quantifying central bank communication. Therefore, the main objective of this paper is to bring computational linguistics research into the economic sphere. By developing a language model trained explicitly for monetary policy, our focus is essentially twofold. On the one hand, we sharpen the previously broad focus of embeddings, while, on the other hand, we enhance content extraction compared to the simplicity of dictionary approaches. We see this paper as an essential step in the endeavor of modern text quantification, initialized by Gentzkow, Kelly, et al. (2019, p. 553) who state that *"approaches [...] which use embeddings as the basis for mathematical analyses of text, can play a role in the next generation of text-as-data applications in social science"*.

This paper contributes to the current literature on several fronts. First, we collect a novel text-corpus unparalleled in size and diversity. The corpus, which contains approximately 23.000 speeches by 130 central banks, is considerably larger than any previously used in the central bank communication literature. Second, this paper introduces novel machine learning algorithms for text quantification. We compare a multitude of different algorithms according to objective criteria. Third, by training the novel algorithm on the novel text corpus, we introduce a language model previously unseen in monetary policy (and likely economics). We demonstrate how this language model can be used in various applications throughout this paper, such as examining the effect of central bank speeches during the Euro Area crisis, predicting monetary policy surprises, comparing central bank objectives, and measuring gender bias. Finally, by making the language model publicly available[42], this paper's most important contribution is to make this new string of research accessible to other researchers, allowing them to incorporate embeddings into their own research.

The remainder of this paper is structured as follows. Section 4.2 provides a literature overview of the current state of Natural Language Processing (NLP) in monetary economics. In Section 4.3 we introduce both the text corpus and the algorithms, combining both elements into language models. We then evaluate the quality of the resulting embeddings in the central bank context in Section 4.4 before applying the best performing language model in Section 4.5. The final section concludes this paper.

[42] https://sites.google.com/view/whatever-it-takes-bz2021

4.2 Related literature

NLP has established itself in the central banking literature with an abundance of high-quality research. There are several methods available to researchers for quantifying qualitative information; Gentzkow, Kelly, et al. (2019) provides an excellent survey on the use of text data with a focus on economics.

Rather than the explicit analysis of text, tracking market reactions during periods when a text is published is a frequent dimensionality reduction method. This strand of literature disregards the qualitative data provided and instead entirely focuses on the market's interpretation of the text. Among successful implementations are Gürkaynak, Sack, et al. (2005), Brand, Buncic, et al. (2010), Jarociński and Karadi (2020), and Swanson (2021) who utilize intraday data around the reading of press-conference statements to measure the effect of monetary policy decisions.

When working with text data, a different approach is to manually classify them, whereby humans categorize sentences, paragraphs or even sections and thus quantify the qualitative information themselves. Although the process is labour-intensive and prone to misclassification, it allows the researcher to capture highly specific patterns. Ehrmann and Fratzscher (2007) use manual classification to compare different types of communication between central banks, Hayo and Neuenkirch (2013) measure the home biases of central bankers, and Tillmann (2020) classifies answers during the ECB press conference's Q&A to estimate a disagreement index.

However, most applications today concentrate on rule-based classification utilizing computers. The majority of NLP in economics focuses on so-called dictionary methods, whereby a predefined dictionary classifies certain words, thereby quantifying the qualitative information into few dimensions. Famous examples in economics include the calculation of an uncertainty and recession index by counting respective terms in news-articles (e.g. Baker, Bloom, et al., 2016; Ferrari and Le Mezo, 2021), stock market predictions using a psychosocial dictionary on a Wall Street Journal column (Tetlock, 2007), or measuring media slant in American news-outlets from phrase frequencies in Congressional Records (Gentzkow and

Shapiro, 2010). There are also numerous applications utilizing dictionaries in the context of central bank communication. In fact, dictionaries have been explicitly designed for the use in financial and central bank context (e.g. Loughran and McDonald, 2011; Apel and Grimaldi, 2014; Picault and Renault, 2017; Correa, Garud, et al., 2021). The peculiarity of the terminology spoken in the central bank context necessitates the usage of such central bank-specific dictionaries. These dictionaries have been applied in numerous ways, for example, to measure implied inflation targets (Shapiro and Wilson, 2019; Zahner, 2020), or financial stability objectives (Peek, Rosengren, et al., 2016; Wischnewsky, Jansen, et al., 2021).

The benefit of dictionary-based methods is their ease of understanding and evaluation through their straightforward and transparent quantification of an underlying corpus. However, at the same time, they lack objectivity and omit relevant information. By definition, dictionaries are subjective, as researchers define a subset of a language's vocabulary based on their own assessment of the underlying true meaning of the respective word. Furthermore, due to their low dimensionality, dictionaries are incapable of capturing nuance as well as interactions between terms. For example, the phrase *great recession* is classified as neutral in Loughran and McDonald's (2011) sentiment dictionary, even though the term *great* is not meant to be positive in this context. Finally, a substantial portion of text is omitted when relying on a dictionary, an argument made before by Harris (1954, p. 156), who state that *"language is not merely a bag of words but a tool with particular properties which have been fashioned in the course of its use"*.

Recent research recognizes and highlights the dictionary approach's disregarding element, suggesting either augmenting such an index or combining different dictionaries to improve predictive power. Tadle (2021), for instance, uses the former approach utilizing two dictionaries (one for hawkish/dovish and the other for positive/negative), rejecting a sentence's classification as hawkish or dovish if it contained more negative than positive terms. The author shows how this augmented sentiment index helps explain movements in high-frequency variables during the FOMC press conference. Another famous example is the interaction of topic-modelling and sentiment analysis by Hansen and McMahon (2016) and

Fraccaroli, Giovannini, et al. (2020). A different approach is applied by Azqueta-Gavaldon, Hirschbühl, et al. (2019), Kalamara, Turrell, et al. (2020), Shapiro, Sudhof, et al. (2020), and Gorodnichenko, Pham, et al. (2021), who combine different sentiment indices in a regression model at the same time. They find that different dictionaries capture various aspects of an underlying corpus and can thus complement each other.

In addition to these augmentations, alternatives to dictionary approaches are becoming more popular. One example is the concept of *similarity*, which is operationalized using the distance between two documents' vocabulary. This metric gained popularity through Acosta and Meade (2015), Amaya and Filbien (2015), and Ehrmann and Talmi (2020), who find that introductory statements became more similar over time. Another example is the measurement of verbal complexity, which is commonly approximated with the Flesch-Kincaid grade level by Kincaid, Fishburne Jr., et al. (1975). Smales and Apergis (2017a) and Hayo, Henseler, et al. (2020) illustrate that markets react strongly concerning the complexity of the information communicated in press statements. As helpful as these new approaches are, some of the corpus' relevant underlying information remains neglected. For example, exchanging the term *inflation* with *deflation* does not change the level of complexity but substantially alters the message.

In the last years, embeddings have entered the realm of monetary policy, following a trend predicted by Gentzkow, Kelly, et al.'s (2019) quote in the introduction. Word embeddings are multidimensional word representations that are used to measure similarity in Twitter tweets (Masciandaro, Romelli, et al., 2020), in the development of a real-time economic sentiment index (Aguilar, Ghirelli, et al., 2021), for the improvement of the Euro Area uncertainty index (Azqueta-Gavaldon, Hirschbühl, et al., 2019), for the decomposition of central bank vague talk (Hu and Sun, 2021), and to measure central banker disagreement (Apel, Blix Grimaldi, et al., 2019). Generally, economic research relies on general language models trained on a general text corpus such as Wikipedia. Shapiro, Sudhof, et al. (2020), for example, use such embeddings in their analysis of news articles. The authors are unconvinced by the results and resort to the modified dictionary approach mentioned earlier. However, the lack of predictive power is most likely

the result of the limited sample size and may possibly be due to the absence of specificity in the training corpus. For example, some general language models lack relevant monetary policy specific terms, such as *hicp*. One notable exception, and thus methodologically the closest research to our paper, is Apel, Blix Grimaldi, et al. (2019), who employ a recurrent neural network to develop their disagreement metric, thereby training word embeddings as a byproduct. However, the authors neither disclose information about their embeddings, nor use them outside this specific context.
To the best of our knowledge, we are the first to train embeddings on a specific text corpus and apply the language model to a variety of applications. Thereby, this paper touches two different literature strings. On the one hand, in the development of novel text-representation (Apel, Blix Grimaldi, et al., 2019), and on the other hand, in the need to fine-tune these representations for their respective use (Loughran and McDonald, 2011).

4.3 Methodology

> *"The meaning of words lies in their use"*
>
> — Wittgenstein (1958, p. 80)

A language model maps a text corpus into an n-dimensional space, whereby the model itself can be arbitrarily simple. Take, for instance, dictionary approaches in sentiment analysis that classify terms as positive, negative and neutral, thereby mapping a corpus' vocabulary into a single dimension. This paper's proposed language model is a multidimensional representation called embedding, received by training an algorithm on a text corpus. A stylised overview of the procedure - and an overview of the structure of this section - is provided in Figure 4.1.

4.3.1 Text Corpus

Our text corpus reflects our paper's primary focus on monetary policy. To make the corpus as broad as possible, we acquire all English central bank speeches published by the Bank of International Settlement (BIS).[43] We complement the

[43] We determine the language of the individual text using Google's Compact Language Detector 3.

Figure 4.1: How to get a language model

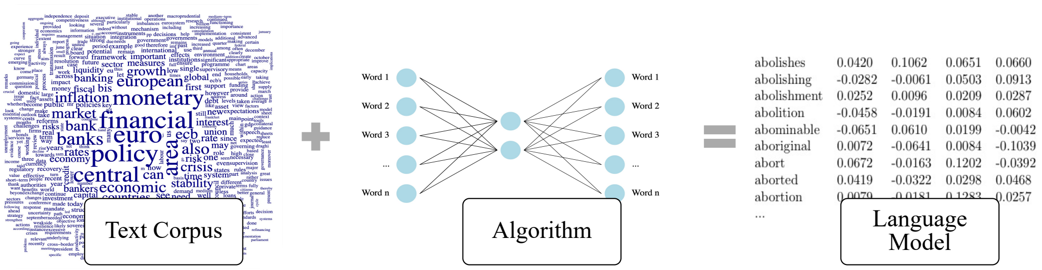


Table 4.1: Corpus Summary

Source	Type	n
BIS	Speech	16,627
FED	Minute, Press Conference, Transcript, Agenda, Blue-, Green-, Teal-, Beige- and Red-Book	2,238
BOJ	Minute, Economic Report, Release, Outlook Report	2,187
ECB	Minute, Press Conference, Economic Outlook, Blog	343
Riksbank	Minute, Economic Review, Monetary Policy Report	330
Australia	Minute	159
Poland	Minute	156
Iceland	Minute	101

Note: The table summarizes the number of documents (n) by sources in the our text corpus.

corpus with as much meta-information as possible, collecting title, speaker, role of speaker, event at which the speech was delivered, and further information. In the next step, we enrich the corpus with documents gathered from central bank websites. Among them are reports, minutes, forecasts, press conferences and economic reviews. To keep our corpus as homogeneous as possible, we exclude all presentations and scientific papers. The former usually contain little coherent text; the latter are primarily oriented towards the academic audience in their jargon and are thus not official central bank communication. The use of information on the respective central bank allows us to create features for the country, the currency area and each central banker. We provide a set of descriptive illustrations in the appendix.

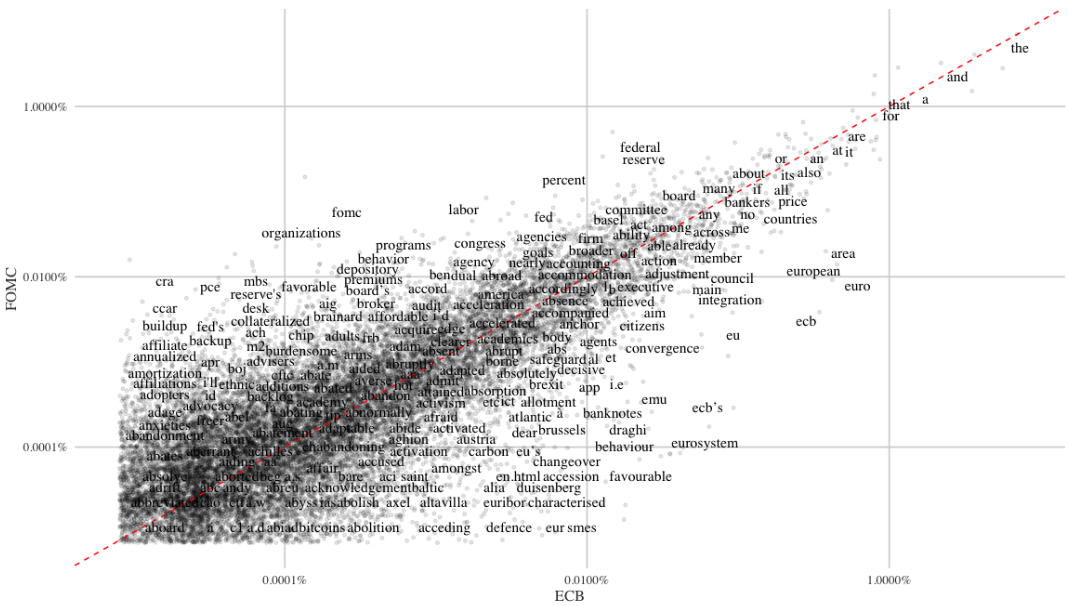
Compared to the previous NLP application in monetary policy (e.g. Amaya and Filbien, 2015; Hansen and McMahon, 2016; Ehrmann and Talmi, 2020), we apply a minimum of pre-processing on the text corpus. This is generally done in the embeddings literature (e.g. Mikolov, Yih, et al., 2013) since similar words should be close in the vector space, which eliminates the need for standardisation through stemming, lemmatisation or removal of stopwords. As a result, we limit the pre-processing to improve the expressiveness of the word tokens. First, we identify so-called collocations, that is, words with specific meaning when used together. It is important to notice the distinctive features of collocation and context were already highlighted by Firth (1957, p. 11), whereas *"collocation is not to be interpreted as context, by which the whole conceptual meaning is implied"* but as *"mere word accompaniment"*. One example is the word pair of *federal* and *reserve*, which have one specific meaning when used together. Another example is the word *quantitative*, which in itself means expressible in terms of quantity. In contrast, *quantitative easing* represents a specific measure of central banks that cannot be concluded from its individual parts. To map these relationships in the embeddings, it is advantageous to identify related words and combine them as a token, for example, *federal_reserve* or *quantitative_easing*. To do this efficiently in our large corpus, we use the algorithm introduced by Blaheta and Johnson (2001) to obtain a basic set of collocations. Furthermore, we form collocations from all speakers of the BIS corpus. For example, *ben* and *bernanke* becomes *ben_bernanke*.

Second, to keep the embeddings as uniform as possible, we replace several unique entities with placeholder tokens. Therefore, all email addresses are encoded as [email], URLs by [url], Unicode tokens by [unicode] and decimal numbers by [decimal]. Furthermore, we remove all apostrophes and quotation marks. In a final step, we convert the entire text to lower case.

Our final corpus includes over 23.000 documents, more than 100 million individual word-tokens, more than 130 central banks worldwide and over 1,000 individual speakers. As a result, on the one hand, we have a text corpus that is unprecedented in quantity and diversity in the monetary policy literature, and on the other, containing highly specific central bank vocabulary.

The corpus' homogeneity is what we address next. To compare the central bank's jargon, we estimate and compare the relative word frequency for the seven most frequent central banks in our sample. An illustration for the ECB and the FOMC is provided in Figure 4.2. Formally testing the homogeneity, we discover that neither of the six central banks has a correlation below 98 percent in their relative word use when compared to the ECB, implying that jargon is very homogeneous across central banks.[44] We conclude from this that the institutions do not differ in any relevant way concerning their jargon. We conduct the same test for the underlying years and derive qualitatively and quantitatively the same result.

Figure 4.2: Illustration of frequency of used terms between FOMC and ECB



Notes: This graph depicts the wording of the FOMC and its European counterpart. The relative frequency of each word is measured for both central banks and presented in this jitter plot. To make it easier to read, numbers and terms that appeared only in the texts of one central bank (mainly names) are removed. In addition both we scaled both axes by the logarithm and added noise, so the correlation is even stronger than shown here.

4.3.2 Algorithm

Our emphasis in this subsubsection is on the introduction of algorithms that provide numerical representations for text documents. Due to the evident shortcomings of traditional dictionary approaches mentioned in the previous section, we turn to linguistics and computer science for our language model.

Modern language models follow the proposition of leading linguistic academic Zellig S. Harris in their pursuit of superior text representation. According to Harris

[44] The precise values are: FOMC: 98.7% (t = 884.67), Riksbank: 98.0% (t = 585), BoE: 98.9% (t = 966), Bank of Japan (BoJ): 98.4% (t = 668), Bundesbank: 99.4% (t = 1257), and Central Bank of India: 98.5% (t = 783). The results are also illustrated in the Appendix in Figure D.2.

(1954, p.151), *"meaning is not a unique property of language, but a general characteristic of human activity"*, implying that the distinction between meaning and the quantifiable properties of language is not always evident. His distributional theory builds on this observation and approximates the meaning of words using the distribution over the environments (context) a word occurs. If a word (for example, *outlook*) can be found repeatedly in the same environments as another word (for example, *forecast*), these words represent a similar concept, whereas the difference in environments corresponds to the difference in meaning.

In the following, we will introduce four algorithms building on the distributional hypothesis that we will subsequently apply to obtain embeddings. These algorithms can be broadly divided into two categories: prediction-based methods and count-based methods. The former use surrounding words to make predictions, whereas the latter uses corpus-wide statistical properties such as word co-occurrence – that is, how often words appear together. The following section introduces both categories and their most prevalent techniques.

Prediction-based algorithms Prior to formally introducing the algorithms, we provide a simplistic example to facilitate comprehension between the concepts of terms, target words and context. Following Harris (1954)'s distributional theory, a word's meaning is based on the environment in which it appears. The context of a word, the set of its surrounding words, operationalizes this environment. Given a context window of one, the context of the word *brighter* (called the target word) in the following sentence would be *this* and *outlook*:

> *"[...] this* **brighter** *outlook remains subject to considerable uncertainty, also regarding the path of the pandemic [...]"*
>
> — Christine Lagarde, IMF Spring Meetings, 8 April 2021

The prediction-based algorithms are generally tasked to predict the target word given the context words, i.e. *p(brighter | this, outlook)*. They then proceed with the next target word, i.e. to predict *p(outlook | brighter, remains)*, then *p(remains*

| *outlook, subject)* and so on.[45] By optimizing some objective function, the algorithm improves its ability to predict target words based on their context. Note how the approach directly incorporates the previously stated linguistic premise by Harris (1954) whereas similar words occur in the same context. It also becomes evident why the context is key. Assume the model is given the (slightly larger) context *"this brighter outlook remains subject to considerable..."* from the preceding sentence and is tasked with predicting the next word. To perform well on this task on average, it must not only assign a high probability to the word *uncertainty*, but also to semantically similar words that frequently occur in the same context, such as *risk*. As a consequence of the prediction task, the algorithm places these words close to each other in the word-embedding space, ultimately capturing the semantic meaning as a byproduct.

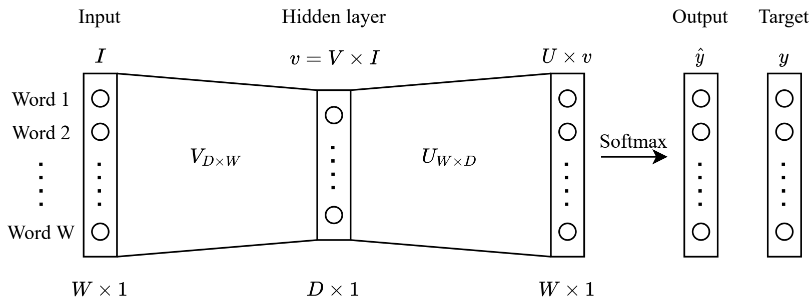
Word2Vec The Word2Vec model of Mikolov, Yih, et al. (2013), Mikolov, Chen, et al. (2013), and Mikolov, Sutskever, et al. (2013) is based on the above principle. Building on the work of Bengio, Ducharme, et al. (2003), Collobert and Weston (2008), and Turian, Ratinov, et al. (2010), the authors propose a neural network capable of predicting words from their context. In doing so, the algorithm is both accurate and efficient.

Mathematically, Word2Vec, and similar prediction-based models, are single-layer log-linear models based on the inner product between two word vectors. The hidden layer's size determines the dimensionality of the word-embedding's representation. An illustration of such a model is provided in Figure 4.3. Formally, the target of the neural network underlying the Word2Vec approach is to predict a single word w_t – the target word – based on its surrounding words w_c – its context – for a vocabulary size W. The objective of the network is to maximize the log-likelihood:

$$L = \frac{1}{T}\sum_{t=1}^{T} \log p(w_t|w_c). \tag{4.1}$$

[45]The demonstrated example is called *continuous bag of words* model. In addition, there is a reverse approach, i.e. the algorithm is tasked to predict the context from the target word. This method is called *skip-gram*.

Figure 4.3: Graphical illustration of Mikolov, Yih, et al.'s (2013) Word2Vec model.



Notes: This figure illustrates the model architecture of a feed-forward neural network with three layers. The first layer is called the input layer, the second hidden layer, and the third output layer. The connections between the layer (particularly the nodes) are called weights and adjusted during the training process. The ensuing word-embedding matrix is, therefore, the projection of the input layer into the hidden layer. A second weight matrix maps the hidden layer into the output layer.

The probability of word w_t, given the words w_c is estimated using the following softmax function:

$$p(w_t|w_c) = \frac{exp(v_{w_t}^T v_{w_c})}{\sum_{w=1}^{W} exp(v_w^T v_{w_c})} \tag{4.2}$$

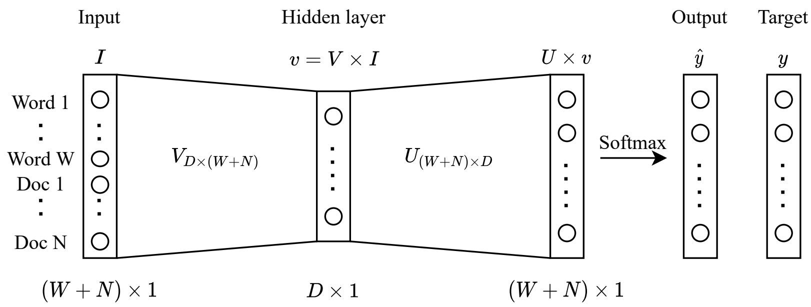
where v_i is the embedding vector. In other words, the models' functional structure represents a single linear hidden layer linked to a softmax output layer, where the exponential function prevents negative numbers and could be omitted without loss of generality. The objective is maximized using an iterative optimization algorithm (stochastic gradient descent, see, e.g. Chakraborty and Joseph, 2017; Athey, 2019) to identify a local – in best case global – maximum. Ultimately, we are only interested in the vector representations for the target words $\hat{V}$, as those are the corresponding embeddings.

There are several interesting points to note from this approach. First, the hidden layer's size is equivalent to the dimensionality D of the embeddings by design. This size has traditionally been set to 300 (e.g. Mikolov, Yih, et al., 2013), but different sized representations are entirely feasible. Second, it is apparent that the window size m (the context) significantly impacts the embedding. Since each word in the context has equal weight on the target prediction, a broad word context may not capture important semantic meaning. In contrast, a very narrow context may miss relevant details. The initial calibrations of Word2Vec and Doc2Vec (the algorithm described next) used single-digit window sizes, namely five (Mikolov, Sutskever, et al., 2013) and eight (Le and Mikolov, 2014).

Third, due to the unsupervised nature of this machine learning model, there is no necessity to provide labelled data. In other words, no manual input is required to obtain the desired word embeddings, which is a substantial advantage since training such models necessitates a large training corpus. Furthermore, if the underlying text is sufficiently homogeneous, researchers can use a much larger text-corpus during the training phase of the language model compared to its final application. We utilise this advantage by training the central bank specific language model on texts from numerous central banks.

Doc2Vec There are several extensions to the original Word2Vec model. The Doc2Vec approach by Le and Mikolov (2014), which proposes the inclusion of document specific information in the input layer, is one notable example. In its simplest form, Doc2Vec incorporates an ID for each document into the neural network's input layer, resulting in an embedding vector for each document. This representation is referred to as document embedding in the remainder of this paper. An illustration of the Doc2Vec model is provided in Figure 4.4. This approach is intuitively similar to controlling for specific characteristics in traditional economic regressions, such as country-dummies in a panel regression. The main advantage of Doc2Vec over Word2Vec is that the document embedding can be

Figure 4.4: Graphical illustration of Le and Mikolov (2014)'s Doc2Vec model.



Notes: This figure is intended to provide an illustration of the Doc2Vec model architecture. It is inspired by Le and Mikolov (2014)'s depiction. The only difference to Figure 4.3 is the additional document ID being fed into the neural network. The ensuing word-embedding and document-embedding is the projection of the input layer into the hidden layer.

used as a summary of the document in subsequent regressions. In Section 4.4 and Section 4.5, we will demonstrate how similarity in document embeddings may be used in a regression model. However, it should be noted that, unlike word embeddings, document embeddings cannot be easily transferred to new corpora.

Count-based algorithm An alternative to obtaining embeddings through neural networks is leveraging corpus-wide statistics to obtain word representations. Our analysis focuses on two approaches: one designed for topic modelling and the other developed explicitly as a substitute for the previously introduced prediction-based algorithms.

LDA The most famous example of a count-based model in economics is unquestionably the Latent Dirichelet Allocation (LDA) algorithm. Since its introduction by Blei, Ng, et al. (2003), it has been used in monetary policy numerous times (e.g. Tobback, Nardelli, et al., 2017; Hansen, McMahon, and Tong, 2019; Wischnewsky, Jansen, et al., 2021). We will not formally introduce the concept of LDA here owing to its popularity in economics and central banking. Interested readers are directed to Bholat, Hansen, et al. (2015) for an introduction to LDA in monetary policy text-mining applications. The premise of LDA is that documents contain a combination of latent topics, which themselves are based on a distribution over words in the underlying corpus. The generative probabilistic model is used in most economic applications to uncover latent topics in a corpus. As a byproduct, LDA generates topic distributions over the vocabulary as well, a concept closely related to the embedding matrices of prediction-based approaches, which is why we incorporate LDA into our analysis.
However, there are several distinctions between our LDA approach and previous ones in economics. First, to the best of our knowledge, these "topic"-embeddings have never been used in an economic context. Second, the number of topics – an important hyperparameter in LDA– varies widely across applications, ranging from two (Schmeling and Wagner, 2019) to 70 (Hansen, McMahon, and Prat, 2018). As our objective is to maximise predictive power and to keep LDA comparable to others algorithms, we cover a much larger number of topics, namely 300.

Finally, in economic applications, the identification and analysis of latent topics are generally the main priority. We refrain from interpreting (or even selecting) topics in the same fashion as we do for all other algorithms.

GloVe The most famous count-based algorithm in NLP is the global factorization method, called GloVe. Following the success of Word2Vec, Pennington, Socher, et al. (2014) propose GloVe, which trains a language model on global word co-occurrences. The approach is based on the notion that the global relative probability of terms, co-occurring in the same context, captures the relevant semantic information. Formally, the following least squared regression model is proposed:

$$L = \sum_{t,c=1}^{W} f(X_{t,c})(w_t^T w_c + b_c + b_t - \log X_{t,c})^2. \tag{4.3}$$

In Equation (4.3) w_t is the word-embedding vector for word t, $f(.)$ is a concave weighing function, b_c and b_t are bias expressions, and $X_{t,c}$ the co-occurrence counts for the context and target word within a defined window. Equation (4.3) is then iteratively optimized given the scale of the regression. The authors find substantial improvements over Word2Vec using the same corpus, vocabulary, and window size.

General corpus models As mentioned in the introduction, no attempts have been made to train embeddings specifically for the economic context to the best of our knowledge. This may be due to the computational burden, the necessary amount of text, or other factors. An alternative to training embeddings from scratch is the use of pre-trained general language models called *transfer learning* (e.g. Binette and Tchebotarev, 2019; Doh, Song, et al., 2020; Istrefi, Odendahl, et al., 2020; Shapiro, Sudhof, et al., 2020; Hu and Sun, 2021). These are open-source language models that have been trained on large general corpora. Since pre-trained language models are methodology-independent, one can find both pre-trained GloVe models and pre-trained Word2Vec models. We compare our results to two such general models as a benchmark: GloVe6B and Word2Vec

Table 4.2: Model Overview

Model	Word embedding	Document embedding	Corpus
Word2Vec	x		CB corpus
Word2Vec GoogleNews	x		Google News
GloVe	x		CB corpus
GloVe6B	x		Wikipedia/Gigaword
Doc2Vec	x	x	CB corpus
LDA	x	x	CB corpus

Note: The columns 'Word embedding' and 'Document embedding' refer to the model language model's ability to generate the respective embeddings. 'CB' is used as an abbreviation for 'Central Bank'. Word2Vec GoogleNews refers to the Le and Mikolov (2014) language model and GloVe6B refers to Pennington, Socher, et al. (2014).

GoogleNews.[46]

In Table 4.2, we provide an overview of all algorithms and corpora applied in this paper to train the language models. Since many algorithms can be computed in different configurations, we test different specifications here. The hyperparameters we used for each model can be found in the Appendix D.2.

4.4 Evaluation of language models

In this section, we apply the algorithms introduced in the previous section to our corpus and evaluate their performance. In this way, we expect to answer the question of which algorithm summarizes the content of the text corpus best and thus provides the most convincing language model. Due to the algorithm's heterogeneity – Doc2Vec and LDA estimate document embeddings in addition to word embeddings – we proceed by estimating a word representation and a document representation whenever possible.

Since there exists no benchmark for evaluating language models in economics yet, we turn to the fields of traditional linguistics. There, evaluation tasks can be broadly distinguished as intrinsic or extrinsic. Intrinsic procedures examine whether the embeddings reflect an assumed relationship between words. One typical task would be to determine whether the vectors indicate associations similar to humans' perceptions. Another task would be the ability to find word

[46] GloVe6B (Pennington, Socher, et al., 2014) is trained on 6 billion tokens from Wikipedia text and News articles with a vocabulary of 0.4 million tokens. Word2Vec GoogleNews (Le and Mikolov, 2014) results from the original paper and is trained on Google News articles.

analogies that resemble real analogies. We present several intrinsic evaluations at the second part of this section.

4.4.1 Extrinsic evaluation

Extrinsic tasks involve evaluating the vectors against other, externally known contexts, i.e., assessing the embeddings' ability to solve specific tasks. Typical methods would be classification tasks or named-entity recognition. However, the datasets, on which these tasks generally rely on, are designed to evaluate embeddings in a broad context, when we are interested in the opposite, their domain specificity. Due to a lack of external evaluation methods, we follow Le and Mikolov (2014) and evaluate each model's predictive performance in a classification task.

Our evaluation task concerns the current interest rate level of the ECB and FOMC, which we forecast using the respective central bank's speeches.[47] Since we are primarily concerned with the correct level, we divide the corresponding 3-month interbank rates into quintiles to derive our evaluation target.[48] Finally, as we are interested in the best possible performance, we employ a neural network to predict the respective interest rate levels with our embeddings.[49] This algorithm allows for complex non-linear relationships between the individual dimensions, which may be relevant. Each language model is trained on 75% of our data (the training sample), with the remaining observations serving as the test set for out-of-sample evaluation. Table 4.3 summarizes the accuracy of the evaluation results split by Document- and Word Embedding and task. Since there exists several variants in the Word2Vec and Doc2Vec algorithms and we

[47] Note that we evaluate our language model on a sub-sample of the available embeddings. However, Section D.3.1 demonstrates that the presented results are robust on a more general task.

[48] It is not uncommon in machine learning and monetary policy to convert a regression analysis into a classification one. The previously discussed Apel, Blix Grimaldi, et al. (2019) are one noteworthy example.

[49] With a few exceptions, the network structure closely resembles the representation in the previous section. We employ a single hidden layer neural network with 64 units, dropout regularization, and a Relu activation. Softmax activation is used once more in the output layer. The Adam optimizer is used to train the model on a categorical cross-entropy loss function. We tested various specifications, but the performance does not change substantially. The exact parameterization are available upon request.

Table 4.3: Evaluation results of algorithms.

Algorithm	3-month Euribor	3-month FFR
Document Embeddings		
Doc2Vec Bow Pre	0.74	0.61
Doc2Vec Bow	0.75	0.59
Doc2Vec PVDM	0.70	0.48
Doc2Vec PVDM Pre	0.67	0.52
LDA	0.55	0.42
Word Embeddings		
Doc2Vec PVDM Pre	0.41	0.35
Doc2Vec Bow Pre	0.40	0.28
Doc2Vec PVDM	0.44	0.22
GloVe	0.38	0.22
Word2Vec GoogleNews	0.36	0.31
GloVe 6B	0.34	0.19
LDA	0.25	0.22
Doc2Vec Bow	0.21	0.25
Word2Vec Bow	0.20	0.21
Word2Vec Skipgram	0.19	0.21

Note: The table shows the evaluation results across the different algorithms introduced in the previous section. The accuracy was evaluated on a classification task with five categories + one outside option if the model was unsure. Therefore the expected performance would be $1/6 \approx 0.17$. With regards to the specifications: Bow = (Distributed) Bag Of Words; PVDM = Paragraph Vector Distributed Memory; Pre = pre-trained embeddings were used as more efficient starting points.

aim for a broad comparison, we estimate all variants. The name in column one starts with the algorithm followed by the variant's abbreviations.[50] Our evaluation yields some interesting results. First, the federal funds rate level appears to be more challenging to predict across models. Second, we find a consistent difference in the level of accuracy between document embeddings and word embeddings. While the former are consistently above 40% accurate, only a few word embedding models achieve this level. Finally, the Doc2Vec algorithm appears to be most suitable for our context, outperforming the others on both the document and word levels.

As a result, we decide to concentrate on Doc2Vec as our primary algorithm. The

[50]For further details, see Mikolov, Sutskever, et al. (2013) and Le and Mikolov (2014).

Bag-of-words variant with pre-trained word-embeddings is explicitly chosen because of its high performance on the document level (being close to par for the ECB task and best at the FOMC task) and consistently good performance on the word embeddings task.[51]

4.4.2 Intrinsic evaluation

Following the extrinsic evaluation, we turn to an intrinsic assessment of our Doc2Vec model. As stated at the outset of this section, these assessments are inherently subjective and should therefore be viewed cautiously. The presented intrinsic evaluations are based on the cosine distance in the embeddings space, which is a measure of similarity between two-word vectors a and b of length n, and defined as follows:

$$similarity_{a,b} = \frac{a \cdot b}{||a|| \times ||b||} = \frac{\sum_{i=1}^{n} a_i \times b_i}{\sqrt{\sum_{i=1}^{n} a_i^2} \times \sqrt{\sum_{i=1}^{n} b_i^2}} \tag{4.4}$$

In the first evaluation, we compare the embeddings based on their assessment of which terms are most similar to a given word. We then assess the embeddings' ability to handle homonyms. Finally, we determine the central bank's similarity score and evaluate whether the relationship shows meaningful results.

We present the first intrinsic evaluation of our embedding in Table 4.4, which lists the most similar words based on the cosine distance to the words *inflation*, *unemployment*, and *output*. For example, the words *unemployment* and *joblessness* are relatively close to each other in our embedding space. It is evident that our language model is capable of grouping words with semantically similar meanings. For example, it is reassuring that through the training process, several terms containing the word *inflation*, such as *core_inflation* and *inflation_expectations*, are grouped together. The same is true for the terms *unemployment* and *output*. Furthermore, it appears that the language model captures the relationships between economic concepts such as *unemployment* and *labor market*.

Next, we turn to an evaluation of homonyms. Some homonyms arise because

[51] Please note that the upcoming results are robust for the individual Doc2Vec variants. Results are available upon request. To ease readability, we will refer in the following to the language model "Doc2Vec Bow Pre" only as "Doc2Vec".

Table 4.4: Intrinsic Evaluation: Similarity in selected word embeddings.

inflation	unemployment	output
core_inflation	unemployment_rate	nonfarm_business
inflation_expectations	natural_rate	sector
economic_slack	joblessness	per_hour
underlying_inflation	jobless	output_growth
inflation_outlook	labor_force	producers
price_inflation	unemployed	manufacturing_output
actual_inflation	labor_market	factory
disinflationary	economic_slack	hourly_compensation
inflation_rate	unemployment_rates	business_equipment
disinflation	participation_rate	labor_costs

Note: The table shows the most similar terms to the words *inflation*, *unemployment* and *output* according to the cosine distance of the underlying word embeddings as defined by Equation (4.4). The underscore is used to highlight collocations as described in Section 4.3.1.

their meanings differ in different contexts. Since our language model is very context-specific, the issue with certain homonyms should be less prevalent than in language models trained on a more general context. In the following, we illustrate this by identifying terms with the closest similarity to the term *basel* and comparing our results to the general language model GloVe6b and GoogleNews. The results can be found in Table 4.5, where we can see that *basel* is associated with the city in GloVe6b and some abbreviations in Word2Vec GoogleNews, but it is only associated with banking regulation vocabulary in our language model. Remarkably, it even correctly matches abbreviations such as the Basel Committee on Banking Supervision (BCBS).[52] Finally, we turn to an intrinsic evaluation of the document embeddings. Here, we measure the similarity between central banks, assuming that central banks in western countries are more akin to one another based on similar objectives. We operationalise this idea by averaging the document embeddings for each central bank and estimating their similarity towards the ECB. The result is depicted in Figure 4.5 with darker colors indicating greater similarity. It appears that central banks in Europe and North America are closest to the ECB, which is consistent with our assumption.[53] This

[52] In the Appendix, we provide additional examples for the interested reader.
[53] Note that we find the same results when using word embeddings.

Table 4.5: Intrinsic Evaluation: Homonym across language models.

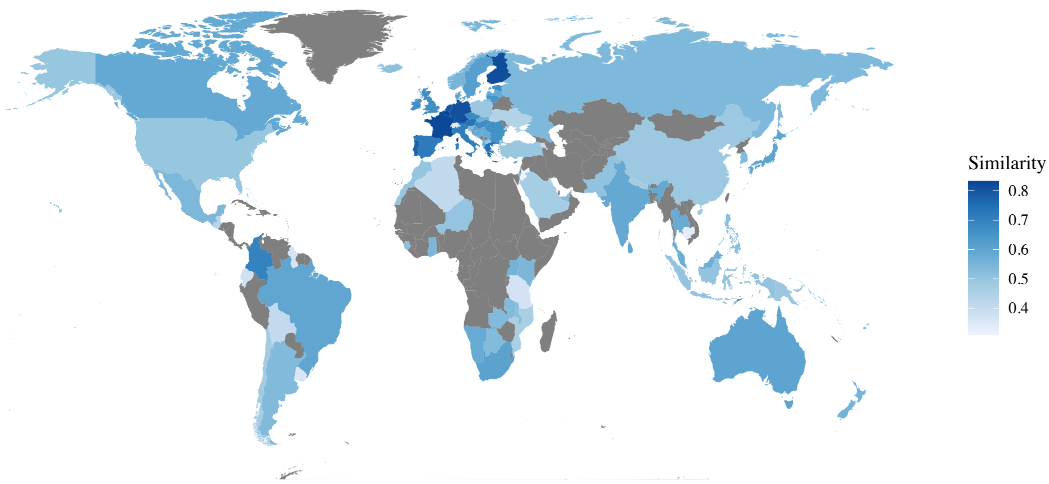
doc2vec	GloVe6B	Word2Vec GoogleNews
basel_committee	zurich	abbr
basle	basle	Tst
capital_accord	zürich	iva
basel_accord	bern	tHe
bcbs	switzerland	Neurol
basle_committee	stuttgart	BASLE
basel_ii	hamburg	PARAGRAPH
basel_iii	cologne	tellus
consultative	lausanne	Def.
minimum_capital	schaffhausen	Complementarity

Note: The table shows for the Doc2Vec and the two genereal corpus models the ten most similar words to the word *basel* according to the cosine distance of the underlying word embeddings as defined by Equation (4.4). The underscore is used to highlight collocations as described in Section 4.3.1.

observation is investigated further in our first application in Section 4.5.

To summarize, we used the previously introduced algorithms for quantifying words and documents in this section. We evaluated all methods using out-of-sample prediction and chose the one with the highest overall predictive power. Subsequently, we used three intrinsic assessments to determine whether previously assumed relationships are embedded in our model. We conclude that the embeddings contain meaningful information at both the word and document level.

Figure 4.5: Central banks' similarity



Notes: This graph illustrates the cosine distance between the average ECB document embedding and all average central bank document embeddings in our dataset. Darker colors depict a lower distance, i.e. a higher similarity. The cosine distance is defined in Equation (4.4).

4.5 Applications

Genberg and Karagedikli (2021) suggest in their survey on machine learning in central banking that empirical approaches in the realm of monetary policy serve one of four purposes: data description, forecasting, structural analysis, and decision communication. The previous section concentrated on the description component, whereas now we apply embeddings in forecasting exercises and structural analysis. In particular we present four potential applications in monetary policy, finance and sociology, using our previously chosen language model.

The first application assesses whether central banks' objectives drive the differences in similarity we found in the previous section. Indeed, we find that text issued by inflation targeting central banks are more similar. Next, we use Mario Draghi's *whatever it takes* speech to create an indicator of the ECB's commitment to act as a lender of last resort. We find that in times of crisis ECB communication can calm financial markets. In our third application, we investigate prejudice and biases in the technical language of central bankers. The final application is in forecasting, where we put our embeddings to the test as a predictor for monetary policy surprises.

The applications are intended to provide case studies for the use of embeddings via transfer learning. Please note that the source code for all applications can be found online.[54] This is done for two reasons: First, we want other researchers to be able to comprehend and replicate our findings. Second, and most importantly, it should demonstrate how conveniently embeddings can be incorporated into one's own research.

4.5.1 Inflation targeting

The first application investigates factors that influence central bank similarity, using the document similarity index introduced in the previous section as a dependent variable. We are particularly interested whether these results are influenced by similar institutional preferences. As a result, we investigate whether the relative similarity to the ECB can be explained by the adoption of inflation targeting, since this is among the most prevalent and observable institutional settings.

In a first step, we label all central banks as "inflation targeting" after their official announcement as an institution aiming for a specific inflation rate, resulting in 44 central banks being classified as such. Next, we determine the average embedding of all central banks. If an inflation target was announced between 1999 and 2020, the institution is divided based on the date of the respective announcement. As outlined in Section 4.4, we calculate the similarity of those average embeddings to the one representing the ECB. Finally, using these similarities as dependent variable and the inflation-objective dummy as independent variable, we run an OLS regression. The results are displayed in column one of Table 4.6.

According to the regression results, adopting an inflation target appears to significantly increase the similarity of texts to the ECB. However, since this effect may be explained by factors such as common currency or membership in the EU, we control for both in columns two and three. While the magnitude of the effect decreases, it remains positive and statistically significant. Both controls enter the regression with positive and significant coefficients.

As a robustness test, we run the same regression using the similarity between

[54] https://sites.google.com/view/whatever-it-takes-bz2021

Table 4.6: Inflation target regression

	Dependent variable:					
	Similarity to ECB					
	Document embedding			Word embedding		
	(1)	(2)	(3)	(4)	(5)	(6)
Inflation target	0.124***	0.081***	0.077***	0.003***	0.003***	0.003***
	(0.018)	(0.020)	(0.022)	(0.001)	(0.001)	(0.001)
Euro Area		0.120***			−0.001	
		(0.028)			(0.001)	
ECB member			0.091***			−0.001
			(0.026)			(0.001)
Constant	0.490***	0.489***	0.488***	0.995***	0.995***	0.995***
	(0.011)	(0.010)	(0.010)	(0.0003)	(0.0003)	(0.0003)
Observations	142	142	142	139	139	139
R^2	0.300	0.482	0.469	0.122	0.130	0.128

Note: Coefficients are estimated using an OLS regression. Standard errors are displayed in parentheses. ***, **, * indicate significance at the 1, 5, and 10 per cent level, respectively.

word embeddings.[55] We find that the adoption of an inflation target remains a highly significant variable. This result makes us confident that one of the factors driving central bank similarity is the adoption of a mutual objective.

4.5.2 Whatever it takes

The second application focuses on the effect of central bank communication in times of heightened uncertainty, utilizing the document space. Although there is literature on this topic using word embeddings (Azqueta-Gavaldon, Hirschbühl, et al., 2019), its focus is on measuring uncertainty using news articles.

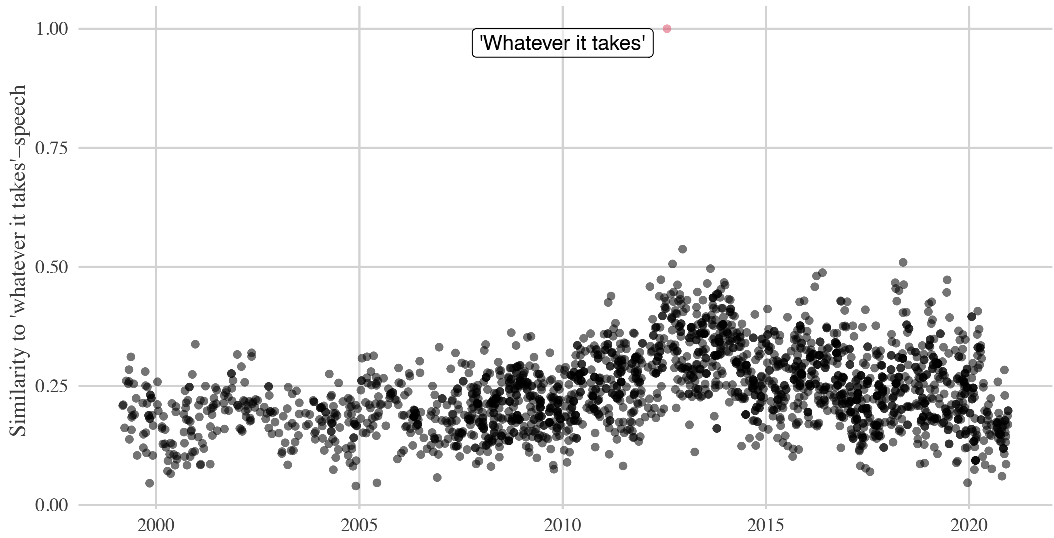
We showcase a novel approach utilizing the cosine distance between the central bank document representations. The focal point is the famous speech by Mario Draghi in London on 26 July 2012, containing the iconic quote: *"Within our mandate, the ECB is ready to do whatever it takes to preserve the euro. And believe me, it will be enough."* This is widely interpreted as the ECB signaling its willingness to act as a lender of last resort if necessary.

Exploiting the particularity of this speech, we calculate the cosine distance be-

[55] Recall that the jargon used by central banks is very similar as highlighted in Section 4.3.1.

tween the ECB's remaining speeches to this event's embedding, thereby creating a time-series of an index, indicating the central bank's willingness to act as a lender of last resort. Figure 4.6 illustrates that, particularly during the Euro Area crisis, the embeddings of central bank speeches appear more similar to the *whatever it takes*-speech. To investigate whether the similarity to that speech can

Figure 4.6: Similarity of all European Central Bank speeches to the "Whatever it takes" speech.



Notes: This graph illustrates the cosine distance between a speech and the *whatever it takes* speech. The cosine distance is defined in Equation (4.4).

calm financial markets in times of heightened uncertainty, we run the following regression:

$$\Delta spread_{10y,t} = wit_{simil,t} + Unc_t + wit_{simil,t} \times Unc_t + X_t + \epsilon_t \tag{4.5}$$

where $\Delta spread_{10y}$ is the change in the spread between Greek 10-year and German 10-year government bonds and wit_{simil} is the cosine similarity of each speech to the *whatever it takes* (wit) speech.[56] We use three different specifications as uncertainty measures Unc: First, the implied volatility of the STOXX50 on the day before the speech ($VSTOXX$), second a decomposition of the VSTOXX into un-

[56]Note that, due to the irregularity of speeches, we use the difference in bond prices between the day before a speech and the closing price of the day after a speech.

certainty (UC) and risk aversion (RA) based on Bekaert, Hoerova, et al. (2021),[57] and finally the ECB's daily CISS index (Hollo, Kremer, et al., 2012). X represents a set of control variables, among them a dummy for the *wit* speech, Moodys agency ratings for Greek bonds, European and U.S. stock prices, monetary policy surprises based on Altavilla, Brugnolini, et al. (2019), and a dummy for the ECB's different central bank presidents. Since considerable risk of autocorrelation, we integrate the first lag of the bond spreads.

Table 4.7: Regression results: Whatever it takes

	Dependent variable:		
	$\Delta spread_{10y}$		
$Unc_t =$	$VSTOXX_{pd,t}$	$CISS_{pd}$	UC_{pd}
wit_{simil}	1.416***	0.353**	0.485***
	(0.482)	(0.161)	(0.179)
$wit_{simil} \times Unc_t$	−0.070***	−2.911**	−0.020***
	(0.026)	(1.262)	(0.007)
Unc_t	0.016***	0.675**	0.005***
	(0.006)	(0.287)	(0.002)
RA_{pd}			−0.0001
			(0.001)
wit_{dummy}	−1.303***	−1.140***	−1.424***
	(0.317)	(0.406)	(0.278)
L(spread10y_d, 1)	0.248**	0.249**	0.249**
	(0.115)	(0.115)	(0.115)
Constant	−0.318	−0.125	−0.123
	(0.283)	(0.235)	(0.267)
Moodys Rating	Yes	Yes	Yes
MP shocks	Yes	Yes	Yes
Stock prices	Yes	Yes	Yes
President Dummy	Yes	Yes	Yes
Observations	2,028	2,028	2,028
R^2	0.116	0.113	0.116

Note: Coefficients are estimated using an OLS regression. Standard errors are displayed in parentheses. ***, **, * indicate significance at the 1, 5, and 10 per cent level, respectively. The test statistics are calculated with heteroscedasticity and autocorrelation robust (HAC) standard errors.

The results can be found in Table 4.7. Starting with the first specification, we

[57] We thank Marie Hoerova for providing the data series.

find a positive and highly significant relationship between $VSTOXX$ and bond spreads, which is consistent with finance theory. Furthermore, there is a clear effect due to the actual speech of Mario Draghi that had a significant negative impact on the spread. Due to the interaction term the effect direction of wit_{simil} depends on the level of uncertainty and changes with increasing uncertainty. At low uncertainty ($VSTOXX < 20$), the coefficient is positive and then becomes negative. A possible explanation for the initial positive effect would be that a *whatever it takes* speech has exactly the opposite effect at low uncertainty. When financial markets are calm, such a speech could be interpreted as a signal of impending troubles. In this situation, the speech would become a self-fulfilling prophecy, triggering spreads to rise. We find no major differences in the other specifications. The sign of the similarity variable remains positive and significant in both cases, but it reverses as the level of uncertainty rises. Only the configuration with the $CISS$ shows a generally lower level of significance. To control for possible other effects, we add additional variables to our model.[58] None of these variables cause the coefficients of interest to change substantially. Overall, we conclude that both Mario Draghi's speech and similar speeches can lower the spread between government bonds when tensions are high and may thus be part of a targeted forward guidance strategy.

4.5.3 Gender Bias

The next application is in an area of monetary policy that is rarely studied: the analysis of biases in central bankers' language. Biases have been found in ordinary language on numerous occasions. However, it may be informative if the very technical language of central bankers contains the same prejudices. Gender bias was chosen as an example of potential partiality in the embeddings primarily because of its contemporary relevance and to showcase how even central bankers' technical jargon might be biased.

Our analysis builds on a fast growing literature that identifies biases in publicly available embeddings (e.g. Caliskan, Bryson, et al., 2017; Garg, Schiebinger, et al., 2018; Manzini, Lim, et al., 2019; Sweeney and Najafian, 2019; Badilla, Bravo-

[58] The full table can be found in Appendix D.4.1

Marquez, et al., 2020), including those used as general models in the previous section. Inherent in those approaches is the idea that language reflects the latent biases of the underlying institutions. Therefore any language model derived from a biased text corpus must inherent these biases as well.

We are following Garg, Schiebinger, et al. (2018), who proposed the *Relative Norm Distance (RND)* to represent the latent variable of a bias, a metric that measures a group's association with a neutral word. When two groups are compared, the latent bias of either group can be estimated by their distance towards the neutral term. In practice, the authors recommend gathering two lists of terms (i.e., male and female pronouns) and then averaging their embeddings. The distance between these averages and a neutral word (i.e., childcare) can then be used to calculate the prejudice torwards this neutral term. For instance, if the distance for the female average embedding is smaller than the distance for the male average embedding, the term is more closely associated with women, and vice versa. Formally for the average word list v_a and v_b with n dimensions each and a neutral word w, the RND can be calculated by:

$$RND_{a,b} = \sqrt{\sum_{i=1}^{n}(w - v_{a,n})^2} - \sqrt{\sum_{i=1}^{n}(w - v_{b,n})^2} \tag{4.6}$$

To test for underlying biases in our embedding, we collect study programs and their respective gender ratios in Bachelor programs across Europe.[59] Next, we estimate the RND for each study program with respect to a set of male and female pronouns as suggested by Garg, Schiebinger, et al. (2018).[60] The most feminine and masculine programs according to our language model can be found in Table 4.8. With a few exceptions, male pronouns are most closely associated with STEM fields, whereas female pronouns are most closely associated with care-

[59] We use data from Eurostat on students enrolled in Bachelors Programs by sex. The dataset can be found here: https://ec.europa.eu/eurostat/databrowser/view/EDUC_UOE_ENRA03___custom_1091992/default/table

[60] Specifically, the following pronouns are used: Female pronouns (v_a): she, daughter, hers, her, mother, woman, girl, herself, female, sister, daughters, mothers, women, girls, females, sisters, aunt, aunts. Male pronouns (v_b): he, son, his, him, father, man, boy, himself, male, brother, sons, fathers, men, boys, males, brothers, uncle, uncles, nephew. A complete list of the academic fields is available upon request.

Table 4.8: Academic profession association by gender

Female pronouns	Male pronouns
childcare	fashion
wildlife	physics
nursing	architecture
pre-school	mechanics
welfare	computer
education	automation

Note: The table replicates the findings of the RND measure as introduced in Garg, Schiebinger, et al. (2018). It illustrates the subset of occupations most associated with gender pronouns.

taking and education.

To formally test whether this bias in association may be driven by the dominance (or lack thereof) of any gender in the respective academic profession, we run a simple OLS regression with the former as explanatory variable. The result can be found in Table 4.9. The regression indicates that female participation is much higher in fields closer associated with female pronouns. The effect is statistically significant and economically relevant.

Importantly, these findings do not imply that any specific central banker or institution is communicating a gender bias on purpose. Rather, we believe that general social patterns, such as occupational gender distribution, are likely to be reflected in central bank texts as well. We hope to emphasize that any text (and thus its embeddings) are not without prejudice and should therefore be used with caution.

Table 4.9: Regression results - Gender Bias

	Dependent variable:
	Relative norm distance
Fraction of female students	0.039***
	(0.013)
Constant	−0.030***
	(0.008)
Observations	67
R^2	0.113

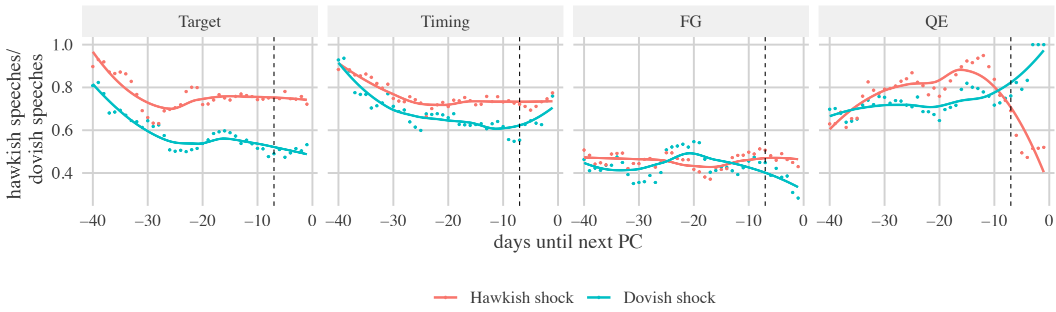
Note: The RND measure is used as defined in Equation (4.6). Higher values indicate closer association to female pronouns and lower values closer association with make pronouns. The respective pronouns can be found in Footnote 60. Coefficients are estimated using an OLS regression. Standard errors are displayed in parentheses. ***, **, * indicate significance at the 1, 5, and 10 per cent level, respectively.

4.5.4 Predicting monetary policy surprises

In our final application, we turn towards the prediction of financial variables, specifically whether ECB speeches can accurately predict the central banks' monetary policy. To investigate whether there is predictive power in the embeddings, we turn to the monetary policy surprises by Altavilla, Brugnolini, et al. (2019). The authors construct four surprises based on different parts of the term structure using high-frequency financial data around the ECB press-conference. The researchers use a rotated factor model to calculate the change in Overnight Index Swap (OIS) rates from one month to 10 years on four latent variables. They call the relevant factors target, timing, forward guidance (FG) and quantitative easing (QE) according to their effect horizon.[61] Since these policy surprises are expected to be unpredictable (otherwise, markets would price in the change), this provides an interesting evaluation with respect to the wealth of information provided by the embeddings.

[61] For the US Gürkaynak, Sack, et al. (2005) identified a target and path factor. Due to the unique institutional setting of the ECB, the path factor can be further separated into timing, FG and QE (e.g. Brand, Buncic, et al., 2010; Swanson, 2021). The target surprise loads most on one-month OIS rates, timing on 6-month rates, FG on 2-year rates, and QE on 10-year rates. In general, a positive surprise corresponds to an increase in OIS rates and thus to restrictive monetary policy and vice versa.

Figure 4.7: Relative frequency of hawkish and dovish speeches preceding an ECB press-conference.



Notes: This graph depicts the relative frequency of hawkish to dovish speeches, as measured by the relative distance between a speech and hawkish to dovish press-conferences separated by Altavilla, Brugnolini, et al.'s (2019) policy surprises and measured in days. The upper red line represents the frequency of hawkish speeches preceding a hawkish press-conference. In contrast, the lower blue line represents the frequency of hawkish speeches preceding a dovish press-conference. The frequencies were measured with a sliding window of +-3 days.

We begin by categorizing each press-conference and surprise as *hawkish* if its surprise is positive and *dovish* if it is negative. Using the RND introduced in the previous application, we measure the relative euclidean distance for each speech with respect to all hawkish and dovish press-conferences and classify speeches as hawkish if they are relatively closer to the hawkish press-conferences and dovish otherwise. First anecdotal evidence of these speeches' potential predictive power can be found in Figure 4.7, where we find a difference in the relative frequency of hawkish speeches preceding a hawkish press-conference and vice versa. While the difference increases closer to the relevant press-conference for target and timing surprises, the pattern is less clear for FG and QE. For the former, there is no clear difference, and for the latter, the gap fluctuates.[62]

We run a Probit model regression with the relative frequency as independent variable to formally test this relationship. The results can be found in Table 4.10. Several macroeconomic variables, such as the EONIA rate, current unemployment, and inflation rate, are also included. We find the following: First, the frequency is statistically significant for surprises with a short horizon. Both target and timing surprises show a significant positive correlation between the relative

[62]Note that QE surprises are not available before 1 October 2014, as QE surprises are not expected in the Euro Area until that date. Accordingly, the number of observations varies between surprises.

Table 4.10: Regression results: Altavilla et. al. (2019) shocks

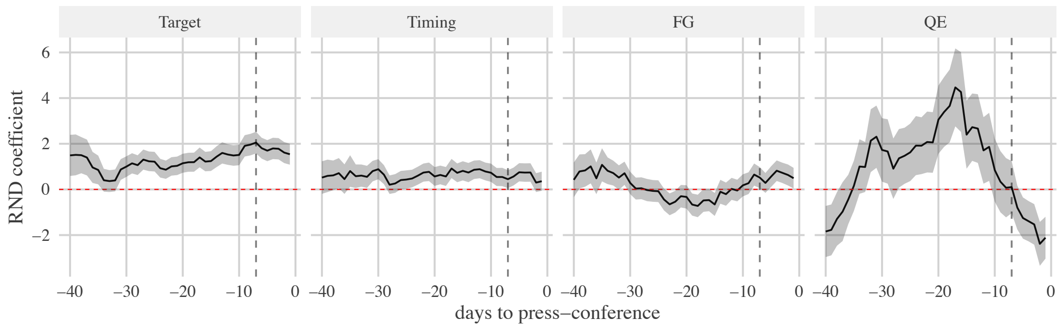
	Dependent variable:							
	Target		Timing		FG		QE	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Frequency	1.46***	1.47**	0.83**	1.03**	−0.05	0.06	0.48	0.03
	(0.43)	(0.65)	(0.42)	(0.51)	(0.61)	(0.63)	(1.48)	(2.00)
BBD Uncertainty		0.01*		0.002		0.001		−0.004
		(0.003)		(0.003)		(0.003)		(0.01)
EONIA		0.30		0.03		−0.04		−1.36
		(0.19)		(0.18)		(0.16)		(4.75)
GDP growth		−0.14		0.05		−0.03		0.10
		(0.11)		(0.08)		(0.08)		(0.12)
Unemployment rate		0.13		0.001		−0.01		−0.01
		(0.14)		(0.13)		(0.13)		(0.56)
Inflation		−0.16		0.21		−0.03		−0.36
		(0.21)		(0.20)		(0.19)		(0.67)
Constant	−1.57***	−3.65**	−0.31	−1.08	0.15	0.17	−0.51	0.49
	(0.32)	(1.62)	(0.31)	(1.62)	(0.31)	(1.51)	(1.16)	(7.06)
Observations	195	195	195	195	194	194	52	52
Akaike Inf. Crit.	239.64	241.92	267.65	275.26	272.19	281.40	75.68	83.27

Note: The dependent variables are the monetary policy surprises by Altavilla, Brugnolini, et al. (2019). Frequency depicts the relative frequency of hawkish to dovish speeches prior an ECB press-conference. Each speech is categorized by the relative distance to the average hawkish and dovish press conference, for each policy surprises respectively. BBD Uncertainty represents the current Baker, Bloom, et al. (2016) uncertainty index. Coefficients are estimated using an OLS regression. Standard errors are displayed in parentheses. ***, **, * indicate significance at the 1, 5, and 10 per cent level, respectively.

number of hawkish speeches and the direction of the surprise. With an average of 10 speeches per press-conference, each additional hawkish speech increases the probability of observing such a hawkish target or timing policy surprise by 10-15%, depending on the shock. Second, this relationship remains even when incorporating the different control variables, so we assume that the speeches and embeddings contain more recent information.

Although we cannot provide empirical evidence as to why pre-decision speeches have predictive power for monetary policy surprises, it seems helpful to point out a possible theoretical channel from the literature. Bauer and Swanson (2020) find that Fed surprises are correlated with macroeconomic news. This news could also be reflected in the central bank's speeches and thus have predictive potential. We find these results particular fascinating since we do not filter speeches at this point, i.e. all speeches are equally weighted, whether they occur ten days

Figure 4.8: Regression results of rolling window approach.



Notes: This graph depicts the hawk-frequency coefficient from the regression results of table 4.10 re-estimated using a rolling window of +-3 days. The y-axis depicts the days to the next ECB press-conference. The grey area is the standard deviation for the respective coefficient.

before a press-conference or 40 days before. However, it seems unlikely that all speeches carry the same weight since the executive board has a quiet period prior to press-conferences and since it seems unlikely that future monetary policy can be communicated this effectively months in advance. To investigate whether the results may be affected by either the short end (through the quiet period phase) or the long end (though monetary policy uncertainty), we run the same regression using a rolling window. Since we cannot effectively control for macroeconomic variables, we use only the frequency of hawkish to dovish speeches as a dependent variable. The resulting coefficients as well its standard error, are illustrated in Figure 4.8.

The findings are not uniform across surprises but can be summarized as follows. First, about one month before the press-conference, the predictions become reliable. Although the results vary between horizons, the general pattern remains the same: the coefficient stabilizes or rises around 20-30 days before the press-conference. Second, we observe the quiet period's expected effect. The grey dotted line depicts the seven days leading up to the press-conference. There, the coefficients become insignificant and thus are no longer a reliable predictor of the monetary policy stance. We find qualitatively and quantitatively the same results for the first three surprises when we use the narrower window for our regression, but a substantial improvement in significance for the QE surprise (2.95**).

Finally, we test the embeddings' out-of-sample performance to evaluate whether the language model has actual predictive power. We use an expanding horizon approach to estimate all regressions in order to formally test forecasting performance and avoid look-ahead bias (Chakraborty and Joseph, 2017). The model is parameterized based on observations prior to 2017, and the predicted policy surprise are compared to their true values between 2017 and 2020 (33 observations, i.e. ≈ 20% of the sample). We specifically choose this period since we are interested in the predictive power of our model during different time periods.[63] The results can be found in Table 4.11. Across the different surprises, the accuracy of the predictions is remarkable, all predict higher than 50% correctly.[64] The accuracy appears to decrease with increasing horizon, which is consistent with our earlier findings. This result may provide first evidence that speeches target expectations on the shorter side.[65]

Table 4.11: Out of sample accuracy Altavilla surprises.

Policy surprise	Accuracy
Target	70 %
Timing	61 %
Quantitative easing	58 %
Forward guidance	52 %

Note: This table summarizes the out-of-sample prediction performance across different surprises. The models are estimated on ECB speeches before 2017-01-01 and evaluated on speeches after 2017-01-01. The accuracy displayed is constitutes the fraction of correct predictions by all predictions.

This result makes us confident that i) speeches have predictive power beyond previous findings and ii) that the embeddings can capture some of it. The find-

[63] Since we anticipate a shift in jargon as the COVID-19 pandemic hits the global economy in 2020, this event provides an interesting basis for evaluating the language model outside its training environment. However, we tried other time periods as well and came to the same conclusions.

[64] It is worth noting that we also tested word embeddings in this prediction task. Although they did not outperform the document embeddings, they did provide surprisingly good prediction as well: Target: 64%, Timing: 58%; FG: 58% and QE: 55%. The results are available upon request.

[65] It is important to note that, due to the small number of remaining observations for the training phase, we are not particularly confident in our QE results. We selected multiple time periods, both longer and shorter, and found that QE performed fairly consistently above 50%. However, we would caution against over-interpreting this result based on only 22 observations.

ings provide many potential future research questions regarding the most relevant dimensions in the embedding space and factors affecting those. Furthermore, we employ a simple linear model, whereas recent contributions such as Kalamara, Turrell, et al. (2020) and Hinterlang (2020) demonstrate how machine learning (and, in particular, neural networks) could be applied to such prediction tasks.

4.6 Conclusion

Understanding the communication of central banks has developed to be a substantial entity in monetary policy, with dictionary approaches at the forefront of current techniques to quantify their speeches, press-conference and reports. In this paper, we expanded this literature in three ways: the compilation of a novel text-corpus, the introduction of algorithms stemming from computational linguistic to extract embeddings – a language model – and the provision of central bank specific embeddings.

First, we collect a text-corpus that is unparalleled in size and diversity within this literature, as both is necessary to train such a language model sufficiently. Then, we introduce embeddings, a novel approach from computational linguistics to quantify texts. These language models are trained using machine learning techniques that locate words and documents in a multidimensional vector space. It has been demonstrated that these embeddings can capture meaningful real-world relationships. Finally, we are able to provide high quality text-representations for central bank communication by training and evaluating different algorithms using an objective criteria. The algorithm with the highest predictive power is able to generate both multidimensional word and document representations.

Within this paper we highlighted the broad applicability of embeddings by illustrating four prominent examples in the fields of central bank institutions, financial uncertainty, gender bias, and monetary policy shock prediction. For example, we illustrate that our language model is able to extract relevant information to forecast future monetary policy shocks from public speeches. Throughout our applications, we emphasize several techniques for extracting the abundance of information contained within embeddings. In our work with embeddings, we found that similarities – euclidean and cosine – are a suitable metric for integrating textual information into economic models or investigating them as dependent and independent variables themselves. Furthermore, we highlight how the use of embeddings in neural networks is a field to be further explored in future research. Our approach has important implications for policymakers and central bankers, allowing for more nuanced ex-ante and ex-post evaluations of communication strategies, such as obtaining preliminary assessments of future communication. We believe this paper to be just a first step toward answering many exciting questions, for example extracting superior measures for concepts such as sentiment, or uncertainty, modelling institutional differences, and improving real-time predictions. We hope that by making our language models publicly available, we will be able to assist in this process.

A Appendix: What is on the ECB's mind?

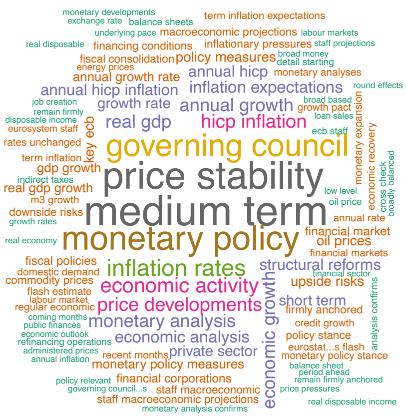
A.1 Taylor rule estimate

Table A.1: Regression results: Taylor rule

	Dependent variable	
	EONIA	
	(1)	(2)
Inflation	0.975***	
	(0.102)	
GDP	0.287*	
	(0.170)	
Inflation (expected)		2.862***
		(0.254)
GDP (expected)		0.774***
		(0.080)
Constant	−0.010	−4.248***
	(0.202)	(0.427)
Observations	224	221
R^2	0.339	0.560
Adjusted R^2	0.333	0.556
Note:	*p<0.1; **p<0.05; ***p<0.01	

A.2 Textual analysis: Most mentioned bigrams in ECB communication

Figure A.1: Wordcloud: ECB communication



Source: ECB press conference introductory statements.

A.3 Summary of variables

Table A.2: Overview of used data and transformations.

Name	Source	Transformation
Dependent Variables		
EONIA Rate	Datastream	-
Euro Area Shadow Rate (Wu & Xia (2016))	Quandl.com	-
Business Cycle		
HICP Inflation Rate	ECB (RTD)	Annual growth
Exp. Inflation Rate	ECB (SPF)	Annual growth
Output Gap	ECB (RTD)	HP transformed real GDP
Exp. Output Gap	ECB (SPF)	HP transformed exp. GDP
Unemployment	ECB (RTD)	-
Exp. Unemployment	ECB (SPF)	-
Core Inflation Rate	ECB (RTD)	Annual growth
Commodity Prices	IMF	Monthly growth
Financial Markets		
Effective Euro Exchange Rate	ECB (RTD)	Monthly growth
Stock Prices (Euro Stoxx 50)	Datastream	Monthly growth
Stock Market Volatility (VSTOXX)	Datastream	Monthly growth
Credit Volume	ECB	Monthly growth
Euro Area Government Bond Yield	FRED	Monthly growth
CISS	ECB	Weekly + monthly growth
Money Supply (M3)	ECB (RTD)	Annual growth
Further Indicators		
Uncertainty Index	policyuncertainty.com	Monthly growth
Trade Deficit	ECB (RTD)	Quarterly growth

A.4 Ten top models (pre- and post-crisis)

Figure A.2: Top models: Pre crisis.

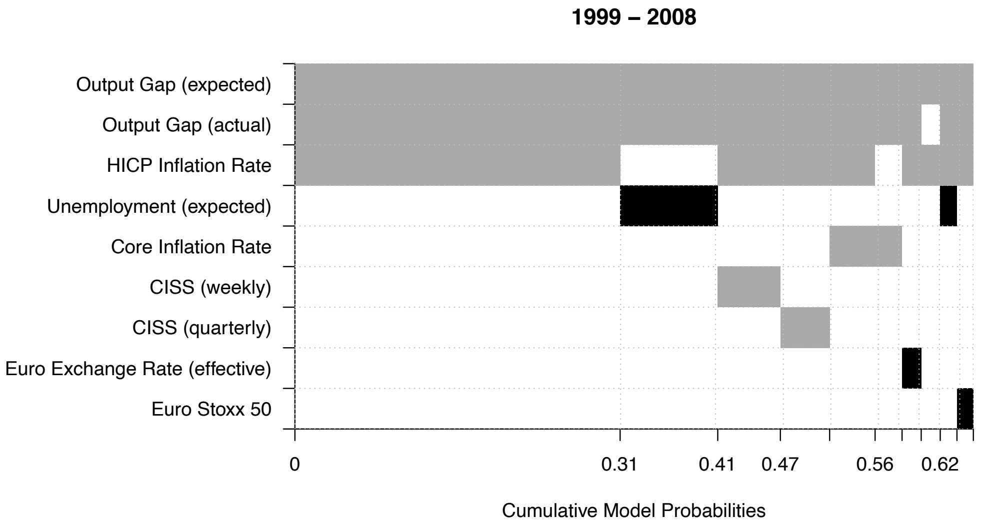
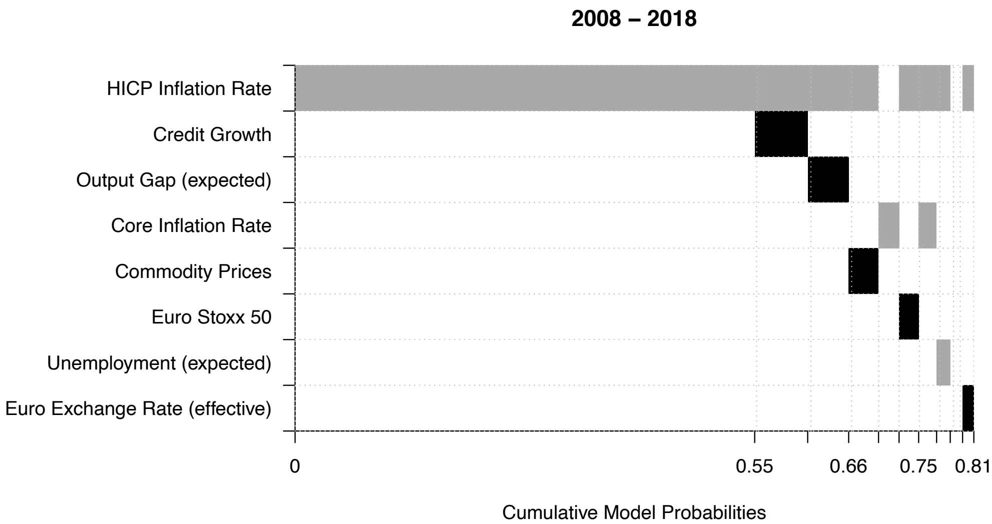


Figure A.3: Top models: Post crisis.



A.5 Robustness checks

Table A.3: Robustness: Prior on the model space

	1999 – 2018		*1999 – 2008*		*2008 – 2018*	
	PIP	Post Mean	PIP	Post Mean	PIP	Post Mean
HICP Inflation Rate	1.000	1.100	0.719	0.313	0.931	0.770
		(0.169)		(0.221)		(0.281)
Unemployment (expected)	1.000	-0.666	0.228	-0.069		
		(0.110)		(0.140)		
Output Gap (expected)			0.980	1.712		
				(0.456)		
Output Gap (actual)			0.952	0.490		
				(0.190)		
Observations		216		114		102

Note: Only robust variables with a PIP ≥ 0.15 are presented.

Table A.4: Robustness: Prior on the parameter space

	1999 – 2018		*1999 – 2008*		*2008 – 2018*	
	PIP	Post Mean	PIP	Post Mean	PIP	Post Mean
HICP Inflation Rate	1.000	1.094	0.748	0.310	0.948	0.778
		(0.169)		(0.211)		(0.267)
Unemployment (expected)	1.000	-0.666	0.254	-0.068		
		(0.111)		(0.113)		
Output Gap (expected)			0.989	1.700		
				(0.427)		
Output Gap (actual)			0.977	0.502		
				(0.172)		
Observations		216		114		102

Note: Only robust variables with a PIP ≥ 0.15 are presented.

Table A.5: Robustness: Starting date of the crisis

	1999 – 2018		*1999 – 2007*		*2007 – 2018*	
	PIP	Post Mean	PIP	Post Mean	PIP	Post Mean
HICP Inflation Rate	1.000	1.095 (0.169)	0.224	0.091 (0.189)	1.000	1.216 (0.219)
Unemployment (expected)	1.000	-0.666 (0.111)				
Output Gap (expected)			0.971	1.890 (0.568)		
Output Gap (actual)			0.874	0.474 (0.349)		
Observations		216		101		115

Note: Only robust variables with a PIP ≥ 0.15 are presented.

Table A.6: Robustness: EONIA

	1999-2018		*1999-2008*		*2008-2018*	
	PIP	Post Mean	PIP	Post Mean	PIP	Post Mean
Unemployment (expected)	1.000	-0.694 (0.059)	0.238	-0.067 (0.135)	0.990	-0.191 (0.048)
HICP Inflation Rate	0.998	0.440 (0.094)	0.736	0.313 (0.216)	0.856	0.200 (0.101)
Output Gap (expected)			0.985	1.708 (0.438)		
Output Gap (actual)			0.967	0.498 (0.179)	0.609	-0.143 0.130
Commodity Prices					0.999	-0.061 (0.010)
Core Inflation Rate					0.299	-0.164 (0.281)
Observations		216		114		102

Note: Only robust variables with a PIP ≥ 0.15 are presented.

B Appendix: Above, but close to two percent.

B.1 Sample text

The two exemplary texts here provide an illustration the 'bag-of-words' approach and Loughran and McDonald's (2011) dictionary. **Positive** terms are highlighted in bold, *negative* terms in italic, and negations are underlined.

> "[...] There is no *doubt* in my mind that the ultimate goal of economic policy, which is to **improve** the standard of living and quality of life of all members of the community, can only be **achieved** by an **effective** interplay between policies with a view to **achieving** simultaneously **efficiency**, **stability** and equity. [...] "
> - Eugenio Solans (2003)[66]

> "[...] *Unfortunately*, many euro area countries entered the financial *crisis* and the economic *downturn* with *unnecessarily weak* fiscal balances, having *missed* the **opportunity** presented by past years' revenue windfalls to consolidate their budgets. [...] "
> - Jose Manuel Gonzalez-Paramo (2009)[67]

These exemplary text contains 1 (5) negative and 4 (1) positive negative terms. However, note that the negative term 'doubt' is preceded by a negation and therefore does not affect the sentiment index. The respective sentiment index for this sentences would therefore be: $S_{Solans,2003} = \frac{5-0}{5} = 1$ and $S_{Gonzalez-Paramo,2009} = \frac{1-5}{1+5} = -2/3$.

[66] https://www.ecb.europa.eu/press/key/date/2003/html/sp030307.en.html (accessed 2020-05-01)

[67] https://www.ecb.europa.eu/press/key/date/2009/html/sp090206.en.html (accessed 2020-05-01)

B.2 Similarity index

The method of estimating the similarity between documents is borrowed from Amaya and Filbien (2015). Amaya and Filbien compare two preceding ECB press conferences by counting the occurrence of bigrams. Bigrams are all combinations in one sentence of two succeeding words. For example, the sentence *"Inflation expectations are higher than expected"* has the bigrams "inflation-expectations", "expectations-are", "are-higher" and so on. By extracting all possible bigrams in the speech, they measure the similarity between two press conferences that take place in month t as the number of bigrams that occur in both press conferences, divided by all bigrams that occur in the two conferences:

$$sim_{t+1} = \frac{bigrams_t \cap \ bigrams_{t+1}}{bigrams_t \cup \ bigrams_{t+1}} \tag{B.1}$$

As this paper compares speeches at a much higher frequency in this paper, I have adjusted the similarity index by Amaya and Filbien (2015) by aggregating the bigrams from all individual i speeches within one month and then comparing them with the bigrams of the following month:

$$sim_{t+1} = \frac{\sum_i bigrams_{i,t} \cap \ \sum_i bigrams_{i,t+1}}{\sum_i bigrams_{i,t} \cup \ \sum_i bigrams_{i,t+1}} \tag{B.2}$$

To detect whether speeches within a month become more similar with each other, I calculated a similarity index within a month (sim_t^{wm}) as follows:

$$sim_t^{wm} = \frac{bigrams_{i,t} \cap \ \sum_i bigrams_{j,t}}{bigrams_{i,t} \cup \ \sum_i bigrams_{j,t}} \tag{B.3}$$

with $i \neq j$. The results are presented in the paper in Section 2.4.1 and can be provided upon request.

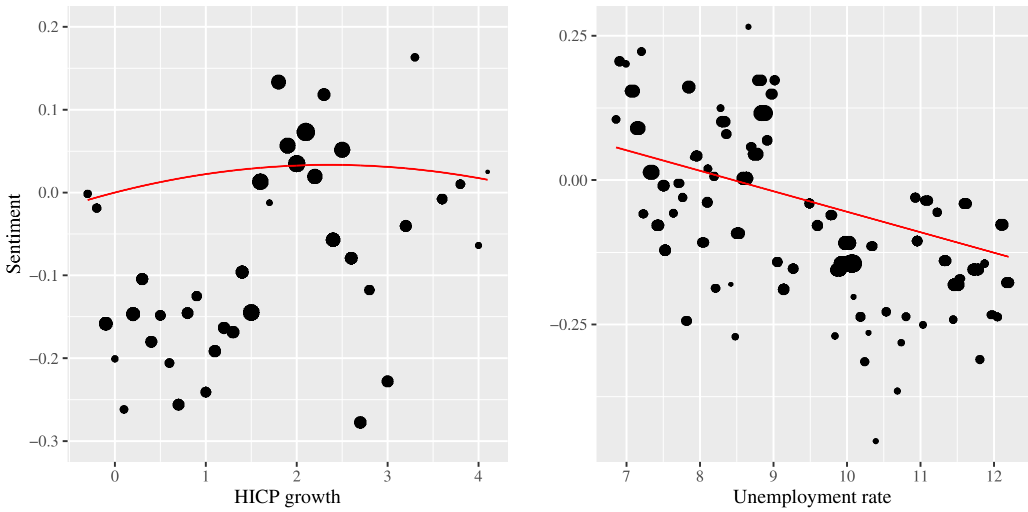
B.3 Descriptive speaker summary

Table B.1: Summary statistics per speaker

	Speakers	Speeches	Mean # of sentences	Mean sentiment
1	Jean-Claude Trichet	318	89	0.090
2	Mario Draghi	183	81	−0.050
3	Benoit Cure	179	100	−0.190
4	Yves Mersch	149	69	−0.040
5	Gertrude Tumpel-Gugerell	148	74	0.270
6	Vitor Constancio	121	121	−0.220
7	Jose Manuel Gonzalez-Paramo	117	96	−0.230
8	Peter Praet	117	94	−0.240
9	Lorenzo Bini Smaghi	107	122	−0.290
10	Lucas Papademos	87	112	0.020
11	Juergen Stark	77	93	−0.060
12	Sabine Lautenschlaeger	76	70	−0.120
13	Otmar Issing	47	106	−0.050
14	Juerg Asmussen	45	60	−0.210
15	Luis de Guindos	40	71	−0.080
16	Willem F. Duisenberg	25	59	0.190
17	Eugenio Domingo Solans	20	93	0.250
18	Tommaso Padoa-Schioppa	16	104	−0.050
19	Philip R. Lane	8	184	−0.310
20	Christine Lagarde	6	49	0.110
21	Sirkka Haemaelaeinen	6	102	0.170
		1,892	91	−0.059

B.4 Unemployment target illustration

Figure B.1: Bin-Scatter Plot of Unemployment Regression (6)



Notes: The illustration of a 'bin-scatter-plot' represents a practical alternative to the more conventional scatter plot. The data points are grouped into bins, and each bin is averaged. In addition, the size of each point is proportional to the number of data points within the respective bin. The objective (red continuous line) is illustrated as of regression results (3). The author's calculations are available upon request.

B.5 Robustness Check Regressions

Table B.2: Regression Results Robustness Checks

	Sentiment Index $S_{i,t}$						
	Squared Economic Activity		*Output-Gap*	*Monthly AR1*		*Resorts*	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
π	0.086***	0.028	0.057**	0.072***	0.025	0.079***	0.028
	(0.024)	(0.027)	(0.026)	(0.027)	(0.027)	(0.027)	(0.023)
π^2	−0.016**	−0.006	−0.008	−0.013*	−0.005	−0.014*	−0.005
	(0.007)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.007)
y_t	0.066***			0.039***		0.039***	
	(0.012)			(0.007)		(0.010)	
y_t^2	0.011***						
	(0.004)						
u_t		−0.031			−0.031*		−0.039***
		(0.143)			(0.018)		(0.011)
u_t^2		−0.0002					
		(0.008)					
x_t			−0.013				
			(0.011)				
Sentiment$_{t-1}$	0.042	0.044	0.057			0.041	0.038
	(0.028)	(0.027)	(0.052)			(0.028)	(0.027)
Sentiment$_{month}$				0.119***	0.115***		
				(0.039)	(0.026)		
Year	−0.014**	−0.015***	−0.006	−0.010**	−0.013*	−0.011	−0.017**
	(0.006)	(0.005)	(0.007)	(0.005)	(0.008)	(0.009)	(0.008)
EPU_{Europe}	−0.071***	−0.080***	−0.119***	−0.083***	−0.075***	−0.080***	−0.077***
	(0.021)	(0.022)	(0.022)	(0.015)	(0.024)	(0.022)	(0.018)
Inflation target	2.64	2.38	2.82	2.75	3.39	2.82	2.93
Observations	1,748	1,748	1,748	1,748	1,748	1,748	1,748
Adjusted R^2	0.227	0.223	0.215	0.225	0.224	0.232	0.232

Note: Coefficients are estimated using an OLS regression. The variable y_t denotes log GDP. The output gap (x_t) is estimated using an HP-Filter with $\lambda = 1600$, no drift and measured in percent. Sentiment$_{month}$ is the delayed previous month's average sentiment. The inflation target is estimated as defined in Equation (2.4). Financial market controls, as well as a time trend and speaker controls, are included in all specifications as defined in Section 2.4.2. EPU_{Europe} is the uncertainty index by Baker, Bloom, et al. (2016). Individual ECB Executive Board members' resorts in specifications (6) and (7) were manually collected from the ECB website and are available upon request. Robust standard errors (clustered by speaker) are displayed in parentheses. ***, **, * indicate significance at the 1, 5, and 10 per cent level, respectively.

C Appendix: Complexity of ECB Communication.

C.1 UMPM Press conferences

Table C.1: ECB's GCM press conferences included in the sample

No.	Date	UMPM disclosure (predominant)
1	07 May 2009	Asset Purchase Programme
2	04 Jun 2009	Asset Purchase Programme
3	06 Aug 2009	Swap Agreement
4	03 Dec 2009	Forward Guidance
5	04 Mar 2010	Allotment Policy
6	10 Jun 2010	Allotment Policy
7	02 Sep 2010	Allotment Policy
8	02 Dec 2010	Allotment Policy
9	03 Mar 2011	Allotment Policy
10	09 Jun 2011	Allotment Policy
11	04 Aug 2011	Allotment Policy
12	06 Oct 2011	Asset Purchase Programme
13	03 Nov 2011	Asset Purchase Programme
14	06 Jun 2012	Allotment Policy
15	02 Aug 2012	Asset Purchase Programme
16	06 Sep 2012	Asset Purchase Programme
17	06 Dec 2012	Allotment Policy
18	02 May 2013	Allotment Policy
19	05 Jun 2014	Allotment Policy
20	03 Jul 2014	Allotment Policy
21	04 Sep 2014	Asset Purchase Programme
22	02 Oct 2014	Asset Purchase Programme
23	22 Jan 2015	Asset Purchase Programme
24	10 Mar 2016	Asset Purchase Programme
25	21 Apr 2016	Asset Purchase Programme
26	02 Jun 2016	Asset Purchase Programme
27	21 Jul 2016	Forward Guidance
28	08 Sep 2016	Forward Guidance
29	20 Oct 2016	Forward Guidance
30	08 Dec 2016	Asset Purchase Programme
31	19 Jan 2017	Asset Purchase Programme
32	09 Mar 2017	Forward Guidance
33	27 Apr 2017	Forward Guidance
34	08 Jun 2017	Forward Guidance

Note: ECB's GCM press conferences sampled following the 2008 financial crisis, when the ECB started conducting UMPM on a recurring basis, apparent by the announcement of ECB's first covered bond purchase programme on 07 May 2009, and covering the period until June 2017. Limitation to press conferences where details on UMPM are disclosed, i.e., details on Asset Purchase Programmes, Swap Agreements, Allotment Policy, and/or Forward Guidance.

C.2 Discriptive summary

Table C.2: Descriptive summary

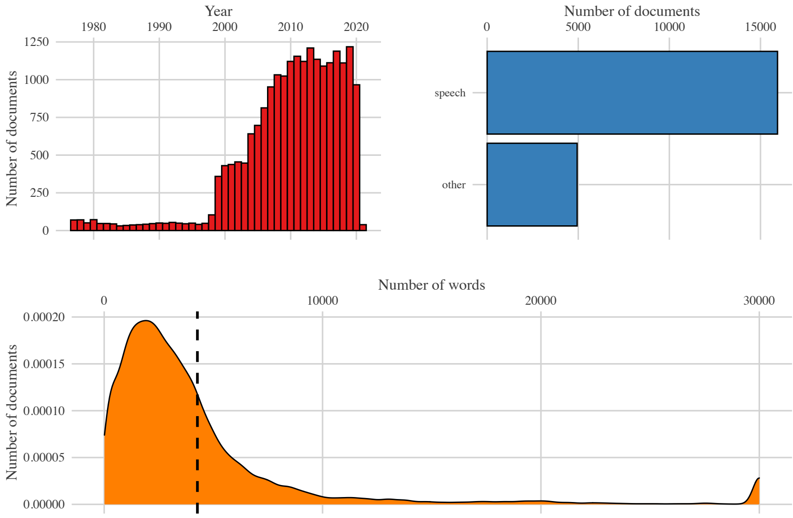
	UMPM (n=34)				Non-UMPM (n=61)			
	Mean	SD	Min	Max	Mean	SD	Min	Max
Complexity Measures								
Flesch Kincaid	15.5	0.55	14.6	16.5	15.4	0.56	14.2	16.5
Flesch	28.7	1.94	25.0	32.6	27.6	1.96	21.7	31.7
FOG	19.6	0.62	18.2	20.8	19.6	0.70	18.1	21.1
SMOG	17.1	0.44	16.2	17.9	17.1	0.50	16.8	18.2
Coleman Liau	13.8	0.35	13.3	14.7	14.1	0.39	13.2	14.8
ARI	15.8	0.66	14.6	17.0	15.6	0.73	13.9	17.1
Trading Volume								
Volume_{Intro}	8.3	0.67	7.19	9.9	8.1	0.72	6.6	9.9
Volume_{Conf}	8.1	0.56	7.14	9.2	7.9	0.49	6.6	9.1
$\text{Volume}_{Q\&A-to-Intro}$	-0.33	0.43	-1.4	0.47	-0.31	0.55	-1.3	1.2
$\text{Volume}_{Q\&A-to-Conf.}$	-0.08	0.13	-0.43	0.12	-0.1	0.14	-0.41	0.17
Control Variables:								
Bond Return	-0.04	0.20	-0.54	0.46	0.95	0.01	0.19	0.50
Shadow Rate	0.19	0.43	-0.99	0.92	-0.17	0.31	-0.78	0.50
Rate Change Dummy	0.12	0.33	0	1	0.08	0.28	0	1
Similarity	0.41	0.09	0.28	0.61	0.45	0.09	0.28	0.65
Robustness Check								
WS	25.0	1.34	22.9	27.2	24.2	1.49	20.80	27.50
SW	1.81	0.02	1.77	1.87	1.83	0.022	1.79	1.89
Future-Orientation	1.18	0.32	0.53	1.87	1.15	0.31	0.49	1.70
Uncertainty	0.68	0.26	0.15	1.19	0.75	0.22	0.30	1.24
Active/Passive	1.14	0.02	1.11	1.19	1.14	0.02	1.11	1.19
Overstated/Understated	1.05	0.01	1.03	1.07	1.05	0.01	1.03	1.08
Positive/Negative	1.03	0.01	1.01	1.06	1.03	0.01	1.02	1.06
Positive/Negative LM	1.00	0.01	0.98	1.01	1.00	0.01	0.99	1.01
Strong/Weak	1.11	0.01	1.09	1.14	1.11	0.02	1.08	1.15

Note: Descriptive statistics of our variables based on a total of 95 observations.

D Appendix: Whatever it takes to understand a central banker.

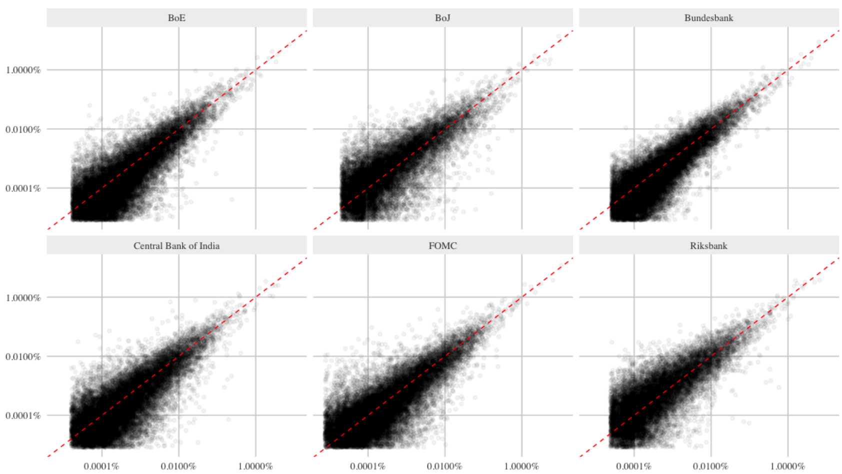
D.1 Graphical illustrations of text corpus

Figure D.1: Descriptive summary of the corpus



Notes: This figure shows the basic properties of our central bank corpus, broken down by year, type, and word length Documents with more than 30,000 words grouped in the *other* category.

Figure D.2: Illustration of frequency of used terms between ECB other central banks.



D.2 Language Model specifications

The following are the hyperparameters we use. For the Word2Vec model we refer to Mikolov, Yih, et al. (2013) and Rehurek and Sojka (2011) and for the GloVe model we use Pennington, Socher, et al.'s (2014) specification. The parameters of the Doc2Vec model are based on Lau and Baldwin (2016). For the LDA we use the findings of Blei and Lafferty (2009) as well as few modifications by Hornik and Grün (2011).[68] The hyperparameters are summarized in the following table:

Table D.1: Hyperparameter Settings for Evaluation

Method	Dim	Window Size	Sub-Sampling	Negative Sample	Itera-tions	learning-rate	alpha	delta
Doc2Vec-DBOW	300	15	0.0001	5	20	0.05	-	-
Doc2Vec-DM	300	5	0.0001	5	20	0.05	-	-
Word2Vec	300	5	0.0001	5	10	0.05	-	-
GloVe	300	-	-	10 20	0.1	0.75	-	
LDA	300	-	-	-	-	-	0.166	0.01

D.3 Additional evaluation

D.3.1 External evaluation

In addition to our economic evaluation task we test our whole embeddings in a more general setting. This should serve as a robustness test with a different task, different empirical methodologies, and far more central bank participation. We select classification tasks that are uninteresting in and of themselves to reduce the risk of spurious correlation between the embeddings and potential application outcome variables (Athey, 2019). In particular, the classification task used here is to predict each speech's central bank and publication year, assuming that higher performance implies a language model's relative superiority.

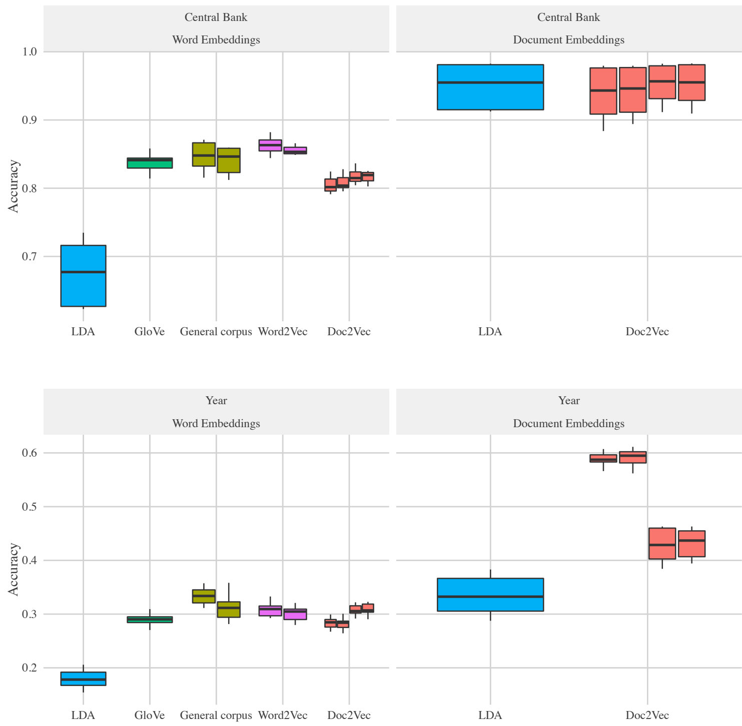
Following current research like Chakraborty and Joseph (2017), the assessment is carried out using out-of-sample testing via cross-validation. In particular, we use five-fold cross-validation, where each model is trained on four-fifths of the dataset

[68] For the Gibbs sampling draws we chose a burnin rate of 1000, sampled 2000 iterations and returned every fifth iteration.

and evaluated on the remaining fifth. This process is repeated five times, with the evaluation's accuracy estimated on each fold. We use the following two machine learning techniques for the classification task: K-Nearest-Neighbor (KNN) and random forest.[69]

The word embedding results are illustrated in Figure D.3, with one algorithm per row and one prediction task per column. The expected accuracy from guessing would be 0.25 for the central bank prediction and 0.06 for the year prediction.

Figure D.3: Evaluation of Embeddings



Notes: This graph depicts the evaluation of different algorithms as discussed in this chapter. The measurement on the y-axis is accuracy of the underlying task, which is measured as $(true\ positive + true\ negative)/(number\ of\ observations)$.

The result is similar to the results from the main text. Document embeddings seem to be better suited for summarizing text. For word embeddings, only minor differences are found between the algorithms. Thus, it seems that in these more general tasks, unlike in the economics-related tasks, our word embeddings do not have a clear corpus advantage over the general language models. However, they

[69] A great introduction into both non-parametric methods as well as the performance metric is provided by Chakraborty and Joseph (2017).

are not worse either. This again emphasizes the potential of our embeddings in the analysis of central banks. Interestingly, there appears no clear trend between KNN and Random Forest with regard to performance, which is – concerning the latter ones' complexity – remarkable. KNN appears to be better in predicting the central banks, whereas random forest is slightly superior in the year predictions.

D.3.2 Internal evaluation

Similar to our *basel* example, we find problems with potentially distorting contexts in general language models if we look at the term *greening*: While Word2Vec GoogleNews associates the colour with this term and Glove6B climate change, our language model associates this topic with terms from the area of climate policy regarding green finance.

Table D.2: Additional Intrinsic Evaluation: Homonym across language models.

doc2vec	GloVe6B	Word2Vec GoogleNews
ngfs	afforestation	greener
climate-related	forestation	sustainability
green_finance	beautification	greened
climate_change	reforestation	green
paris_agreement	canker	Greening
climate-	jagielka	greenest
greener	citrus	composting
frank_elderson	punxsutawney	revitalization
greenhouse	gartside	Greenest
climate_change	colonizing	Greener

Note: The table shows for the Doc2Vec and the two genereal corpus models the ten most similar words to the word *"greening"* according to the cosine distance of the underlying word embeddings as defined by Equation (4.4). The underscore is used to highlight collocations as described in Section 4.3.1.

D.4 Applications - Robustness checks

D.4.1 Application 2: Whatever it takes

Table D.3: Application 2: Whatever it takes - Full table

	Dependent variable:		
	Δspread$_{10y}$		
	(5)	(6)	(7)
$\text{wit}_{\text{simil}}$	1.416*** (0.482)	0.353** (0.161)	0.485*** (0.179)
$\text{wit}_{\text{simil}} \times \text{VSTOXX}_{\text{pd}}$	−0.070*** (0.026)		
$\text{wit}_{\text{simil}} \times \text{ciss}_{\text{pd}}$		−2.911** (1.262)	
$\text{wit}_{\text{simil}} \times \text{UC}_{\text{pd}}$			−0.020*** (0.007)
$\text{VSTOXX}_{\text{pd}}$	0.016*** (0.006)		
ciss_{pd}		0.675** (0.287)	
UC_{pd}			0.005*** (0.002)
RA_{pd}			−0.0001 (0.001)
$\text{wit}_{\text{dummy}}$	−1.303*** (0.317)	−1.140*** (0.406)	−1.424*** (0.278)
altavilla.Target	−0.034 (0.038)	−0.031 (0.038)	−0.034 (0.038)
altavilla.Timing	0.001 (0.008)	0.002 (0.008)	0.001 (0.008)
altavilla.FG	0.005 (0.007)	0.005 (0.007)	0.005 (0.007)
altavilla.QE	−0.024 (0.019)	−0.025 (0.018)	−0.024 (0.019)
lag_asset.sp500	−0.0001 (0.0001)	−0.0001 (0.0001)	−0.0001 (0.0001)
lag_asset.stoxx	0.0001* (0.00004)	0.0001 (0.00004)	0.0001* (0.00004)
MoodysA2	−0.049 (0.067)	−0.045 (0.067)	−0.046 (0.067)
MoodysA3	0.386** (0.168)	0.393** (0.170)	0.379** (0.166)
MoodysBa1	0.063 (0.042)	0.075* (0.044)	0.058 (0.041)
MoodysBa3	0.194 (0.120)	0.192 (0.121)	0.191 (0.117)
MoodysB1	0.154* (0.089)	0.148 (0.090)	0.146* (0.088)
MoodysB3	0.159* (0.089)	0.157* (0.089)	0.156* (0.088)
MoodysCaa1	0.106 (0.106)	0.109 (0.104)	0.102 (0.106)
MoodysCaa2	0.186* (0.108)	0.185* (0.108)	0.181* (0.107)
MoodysCaa3	0.083 (0.107)	0.090 (0.104)	0.080 (0.106)
MoodysCa	0.109 (0.207)	0.130 (0.206)	0.103 (0.205)
MoodysC	−0.060 (0.139)	−0.047 (0.131)	−0.060 (0.139)
lag(spread10y_d, 1)	0.248** (0.115)	0.249** (0.115)	0.249** (0.115)
presidentDuisenberg	−0.091 (0.207)	0.027 (0.195)	−0.073 (0.204)
presidentLagarde	0.087** (0.042)	0.074* (0.044)	0.084** (0.041)
presidentTrichet	−0.044 (0.197)	−0.016 (0.192)	−0.036 (0.196)
Constant	−0.318 (0.283)	−0.125 (0.235)	−0.123 (0.267)
Observations	2,028	2,028	2,028
R^2	0.116	0.113	0.116
F Statistic	10.529***	10.153***	10.101***

Note: Coefficients are estimated using an OLS regression. Standard errors are displayed in parentheses. ***, **, * indicate significance at the 1, 5, and 10 per cent level, respectively. The test statistics are calculated with heteroscedasticity and autocorrelation robust (HAC) standard errors.

D.4.2 Application 3: Gender bias

As a robustness test we replicate the job example of Garg et. al (2018) using female and male names. We use occupation data from Eurostat and match all descriptions with Garg et. al's (2018) pronouns. The following are the results:

Table D.4: Regression results - Gender Bias

	Dependent variable:
	Relative norm distance
Fraction of female students	0.0003*
	(0.0001)
Constant	−0.004
	(0.009)
Observations	32
R^2	0.092

Note: The RND measure is used as defined in Equation (4.6). Higher values indicate closer association to female pronouns and lower values closer association with make pronouns. The respective pronouns can be found in Footnote 60. Coefficients are estimated using an OLS regression. Standard errors are displayed in parentheses. ***, **, * indicate significance at the 1, 5, and 10 per cent level, respectively.

References

Aastveit, K., Natvik, G. J., & Sola, S. (2013). Economic uncertainty and the effectiveness of monetary policy. *Norges Bank Working Paper*, *17*.

Acosta, M., & Meade, E. E. (2015). Hanging on every word: Semantic analysis of the FOMC's postmeeting statement. *FEDS Notes*, (2015-09), 30.

Aguilar, P., Ghirelli, C., Pacce, M., & Urtasun, A. (2021). Can news help measure economic sentiment? an application in covid-19 times. *Economics Letters*, *199*, 109730.

Albulescu, C., Goyeau, D., & Pépin, D. (2013). Financial instability and ECB monetary policy. *Economics Bulletin*, *33*(1), 388–400.

Altavilla, C., Brugnolini, L., Gürkaynak, R. S., Motto, R., & Ragusa, G. (2019). Measuring euro area monetary policy. *Journal of Monetary Economics*, *108*, 162–179.

Amaya, D., & Filbien, J.-Y. (2015). The similarity of ECB's communication. *Finance Research Letters*, *13*, 234–242.

Andersson, M. (2010). Using intraday data to gauge financial market responses to federal reserve and ECB monetary policy decisions. *International Journal of Central Banking*.

Apel, M., Blix Grimaldi, M., & Hull, I. (2019). *How much information do monetary policy committees disclose? evidence from the fomc's minutes and transcripts* (tech. rep. No. 381). Sveriges Riksbank Working Paper Series.

Apel, M., & Grimaldi, M. B. (2014). How informative are central bank minutes? *Review of Economics*, *65*(1), 53–76.

Athey, S. (2019). 21. the impact of machine learning on economics. In *The economics of artificial intelligence* (pp. 507–552). University of Chicago Press.

Azqueta-Gavaldon, A., Hirschbühl, D., Onorante, L., & Saiz, L. (2019). Sources of economic policy uncertainty in the euro area: A machine learning approach, In *ECB economic bulletin*, European Central Bank.

Badilla, P., Bravo-Marquez, F., & Pérez, J. (2020). WEFE: The word embeddings fairness evaluation framework., In *IJCAI*.

Baker, S. R., Bloom, N., & Davis, S. J. (2016). Measuring economic policy uncertainty. *The quarterly journal of economics*, *131*(4), 1593–1636.

Ball, L. M. (2014). *The case for a long-run inflation target of four percent*. International Monetary Fund.

Baranowski, P., Bennani, H., & Doryń, W. (2021). Do the ECB's introductory statements help predict monetary policy? evidence from a tone analysis. *European Journal of Political Economy*, *66*, 101964.

Barro, R. J., & Gordon, D. B. (1983). Rules, discretion and reputation in a model of monetary policy. *Journal of Monetary Economics*, *12*(1), 101–121.

Bauer, M., & Swanson, E. T. (2020). *The Fed's response to economic news explains the "Fed information effect"* (Working Paper w27013).

Bekaert, G., Hoerova, M., & Xu, N. R. (2021). *Risk, monetary policy and asset prices in a global world* (SSRN Scholarly Paper No. 3599583). Social Science Research Network. Rochester, NY.

Belke, A., & Klose, J. (2010). (How) do the ECB and the Fed react to financial market uncertainty: The Taylor rule in times of crisis. *Discussion Papers of DIW Berlin*, *972*.

Bengio, Y., Ducharme, R., Vincent, P., & Janvin, C. (2003). A neural probabilistic language model. *The Journal of Machine Learning Research*, *3*, 1137–1155.

Bennani, H., Fanta, N., Gertler, P., & Horvath, R. (2020). Does central bank communication signal future monetary policy in a (post)-crisis era? The case of the ECB. *Journal of International Money and Finance*, *104*, 102167.

Bennani, H., & Neuenkirch, M. (2017). The (home) bias of european central bankers: New evidence based on speeches. *Applied Economics*, *49*(11), 1114–1131.

Benoit, K., Watanabe, K., Wang, H., Nulty, P., Obeng, A., Müller, S., & Matsuo, A. (2018). Quanteda: An R package for the quantitative analysis of textual data. *Journal of Open Source Software*, *3*(30), 774.

Bernanke, B. S., & Kuttner, K. N. (2005). What explains the stock market's reaction to federal reserve policy? *The Journal of Finance*, *60*(3), 1221–1257.

Bholat, D. M., Hansen, S., Santos, P. M., & Schonhardt-Bailey, C. (2015). *Text mining for central banks* (Handbook No. 33). Centre for Central Banking Studies. Bank of England.

Binette, A., & Tchebotarev, D. (2019). *Canada's monetary policy report: If text could speak, what would it say?* (Staff Analytical Notes No. 2019-5). Bank of Canada.

Black, J., Hashimzade, N., & Myles, G. (2017). *A dictionary of economics*. Oxford University Press.

Blaheta, D., & Johnson, M. (2001). Unsupervised learning of multi-word verbs, In *Association for computational linguistics workshop on collocation (2001)*. Association for Computational Linguistics (ACL).

Blanchard, O., Dell'Ariccia, G., & Mauro, P. (2010). Rethinking macroeconomic policy. *Journal of Money, Credit and Banking*, *42*, 199–215.

Blei, D. M., & Lafferty, J. D. (2009). Topic models. In *Text mining* (pp. 101–124). Chapman; Hall/CRC.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *The Journal of Machine Learning Research*, *3*, 993–1022.

Bleich, D., Fendel, R., & Rülke, J.-C. (2013). Monetary policy and stock market volatility. *Bundesbank Discussion Paper*, *45/2013*.

Blinder, A. S., Ehrmann, M., Fratzscher, M., De Haan, J., & Jansen, D.-J. (2008). Central bank communication and monetary policy: A survey of theory and evidence. *Journal of Economic Literature*, *46*(4), 910–45.

Blot, C., Creel, J., & Paul, H. (2019). Thoughts on a review of the ECB's monetary policy strategy: Monetary dialogue papers. *Monetary Dialogue Papers*, 5–32.

Boguth, O., Grégoire, V., & Martineau, C. (2019). Shaping expectations and coordinating attention: The unintended consequences of FOMC press conferences. *Journal of Financial and Quantitative Analysis*, *54*(6), 2327–2353.

Bomfim, A. N. (2003). Pre-announcement effects, news effects, and volatility: Monetary policy and the stock market. *Journal of Banking & Finance*, *27*(1), 133–151.

Bowdler, C., & Radia, A. (2012). Unconventional monetary policy: The assessment. *Oxford Review of Economic Policy*, *28*(4), 603–621.

Brand, C., Buncic, D., & Turunen, J. (2010). The impact of ECB monetary policy decisions and communication on the yield curve. *Journal of the European Economic Association*, *8*(6), 1266–1298.

Bulir, A., Cihak, M., & Jansen, D.-J. (2013). What drives clarity of central bank communication about inflation? *Open Economies Review*, *24*(1), 125–145.

Bulir, A., Cihak, M., & Smidkova, K. S. (2013). Writing clearly: The ECB's monetary policy communication. *German Economic Review*, *14*(1), 50–72.

Caliskan, A., Bryson, J. J., & Narayanan, A. (2017). Semantics derived automatically from language corpora contain human-like biases. *Science*, *356*(6334), 183–186.

Castro, V. (2011). Can central banks' monetary policy be described by a linear (augmented) Taylor rule or by a nonlinear rule? *Journal of Financial Stability*, *7*(4), 228–246.

Chakraborty, C., & Joseph, A. (2017). *Machine learning at central banks* (tech. rep.). Bank of England. Working Paper.

Coenen, G., Ehrmann, M., Gaballo, G., Hoffmann, P., Nakov, A., Nardelli, S., Persson, E., & Strasser, G. (2017). Communication of monetary policy in unconventional times. *ECB Working Paper Series*, *2080*.

Cœuré, B. (2018). Central banking in times of complexity, In *Panel remarks at a conference on the occasion of Sveriges Riksbank's 530th anniversary*, Stockholm.

Collobert, R., & Weston, J. (2008). A unified architecture for natural language processing: Deep neural networks with multitask learning, In *Proceedings of the 25th international conference on machine learning*, New York, Association for Computing Machinery.

Constancio, V. (2017). The future of monetary policy frameworks, In *Lecture at the instituto superior de economia e gestão*, Lisbon.

Correa, R., Garud, K., Londono, J. M., & Mislang, N. (2021). Sentiment in central banks's financial stability reports. *Review of Finance*, *25*(1), 85–120.

Cuaresma, J. C., & Slacik, T. (2009). On the determinants of currency crises: The role of model uncertainty. *Journal of Macroeconomics*, *31*(4), 621–632.

Cukierman, A., & Meltzer, A. H. (1986). A theory of ambiguity, credibility, and inflation under discretion and asymmetric information. *Econometrica*, *54*(5), 1099–1128.

de Haan, J., & Sturm, J.-E. (2019). Central bank communication. In *The Oxford handbook of the economics of central banking* (pp. 231–262). Oxford University Press.

de Jong, E., & van Esch, F. (2014). Culture matters: French-german conflicts on european central bank independence and crisis resolution, In *Frontiers of Political Economy*, Routledge.

Del Negro, M., Giannone, D., Giannoni, M. P., & Tambalotti, A. (2019). Global trends in interest rates. *Journal of International Economics*, *118*, 248–262.

Dell'Ariccia, G., Rabanal, P., & Sandri, D. (2018). Unconventional monetary policies in the euro area, Japan, and the United Kingdom. *Journal of Economic Perspectives*, *32*(4), 147–72.

Dickey, D. A., & Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, *74*(366a), 427–431.

Diron, M., & Mojon, B. (2005). Forecasting the central bank's inflation objective is a good rule of thumb. *ECB Working Paper Series*, *564*.

Doh, T., Song, D., & Yang, S.-K. (2020). Deciphering federal reserve communication via text analysis of alternative FOMC statements. *Kansas City Research Working Papers Series*, *20*(14).

Ehrmann, M., & Fratzscher, M. (2007). Communication by Central Bank Committee Members: Different Strategies, Same Effectiveness? *Journal of Money, Credit and Banking*, *39*(2-3), 509–541.

Ehrmann, M., & Talmi, J. (2020). Starting from a blank page? Semantic similarity in central bank communication and market volatility. *Journal of Monetary Economics*, *111*, 48–62.

European Central Bank. (2001). Issues related to monetary policy rules. *ECB Monthly Bulletin October*, 37–51.

European Central Bank. (2010). Output, demand and the labour market. *ECB Monthly Bulletin September*, 63–84.

European Central Bank. (2019a). *The definition of price stability*. Retrieved October 15, 2019, from https://www.ecb.europa.eu/mopo/strategy/pricestab/html/index.en.html

European Central Bank. (2019b). *Speeches dataset*. Retrieved February 5, 2020, from https://www.ecb.europa.eu/press/key/html/downloads.en.html

European Central Bank. (2021). *An overview of the ECB's monetary policy strategy.* Retrieved July 25, 2021, from https://www.ecb.europa.eu/home/search/review/pdf/ecb.strategyreview_monpol_strategy_overview.en.pdf

Falagiarda, M., & Reitz, S. (2015). Announcements of ECB unconventional programs: Implications for the sovereign spreads of stressed euro area countries. *Journal of International Money and Finance*, *53*, 276–295.

Fama, E. F. (1979). Efficient capital markets a review of theory and empirical work. *Journal of Finance*, *25*(2), 383–417.

Fendel, R. M., & Frenkel, M. R. (2006). Five years of single european monetary policy in practice: Is the ECB rule-based? *Contemporary Economic Policy*, *24*(1), 106–115.

Fernandez, C., Ley, E., & Steel, M. F. J. (2001a). Benchmark priors for bayesian model averaging. *Journal of Econometrics*, *100*(2), 381–427.

Fernandez, C., Ley, E., & Steel, M. F. (2001b). Model uncertainty in cross-country growth regressions. *Journal of Applied Econometrics*, *16*(5), 563–576.

Ferrari, M., & Le Mezo, H. (2021). Text-based recession probabilities. *ECB Working Paper Series*, *2516*.

Ferrero, A., Gertler, M., Svensson, L. E., & Pesenti, P. (2010). *4. current account dynamics and monetary policy.* University of Chicago Press.

Firth, J. R. (1957). A synopsis of linguistic theory, 1930-1955. *Studies in Linguistic Analysis*.

Foster, D. P., & George, E. I. (1994). The risk inflation criterion for multiple regression. *The Annals of Statistics*, 1947–1975.

Fraccaroli, N., Giovannini, A., & Jamet, J.-F. (2020). Central banks in parliaments: A text analysis of the parliamentary hearings of the bank of england, the european central bank and the federal reserve. *ECB Working Paper Series*, *2442*.

Friedman, M. (1963). *Inflation: Causes and consequences.* Mumbai, Asia Publishing House.

Funke, N., & Matsuda, A. (2006). Macroeconomic news and stock returns in the united states and germany. *German Economic Review*, *7*(2), 189–210.

Gali, J. (2015). *Monetary policy, inflation, and the business cycle.* Princeton University Press.

Garg, N., Schiebinger, L., Jurafsky, D., & Zou, J. (2018). Word embeddings quantify 100 years of gender and ethnic stereotypes. *Proceedings of the National Academy of Sciences*, *115*(16), E3635–E3644.

Genberg, H., & Karagedikli, Ö. (2021). *Machine learning and central banks: Ready for prime time?* (Tech. rep.). The South East Asian Central Banks Research and Training Centre.

Gentzkow, M., Kelly, B., & Taddy, M. (2019). Text as data. *Journal of Economic Literature*, *57*(3), 535–574.

Gentzkow, M., & Shapiro, J. M. (2010). What drives media slant? evidence from us daily newspapers. *Econometrica*, *78*(1), 35–71.

Gerdesmeier, D., & Roffia, B. (2003). Empirical estimates of reaction functions for the euro area. *ECB Working Paper Series*, *206*.

Gerlach, S. (2007). Interest rate setting by the ECB, 1999-2006: Words and deeds. *International Journal of Central Banking*, *3*(3), 1–46.

Gertler, P., & Horvath, R. (2018). Central Bank Communication and Financial Markets: New high-requency Evidence. *Journal of Financial Stability*, *36*, 336–345.

Gorodnichenko, Y., Pham, T., & Talavera, O. (2021). *The voice of monetary policy* (Working Paper No. 28592). National Bureau of Economic Research.

Gross, J., & Zahner, J. (2021). What is on the ECB's mind? monetary policy before and after the global financial crisis. *Journal of Macroeconomics*, *68*, 103292.

Grossman, S. J., & Stiglitz, J. E. (1976). Information and competitive price systems. *The American Economic Review*, *66*(2), 246–253.

Gürkaynak, R. S., Sack, B., & Swanson, E. T. (2005). Do Actions Speak Louder Than Words? The Response of Asset Prices to Monetary Policy Actions and Statements. *International Journal of Central Banking*, *1*(1), 39.

Hansen, S., & McMahon, M. (2016). Shocking language: Understanding the macroeconomic effects of central bank communication. *Journal of International Economics*, *99*, S114–S133.

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

Hansen, S., McMahon, M., & Tong, M. (2019). The long-run information effect of central bank communication. *Journal of Monetary Economics*, *108*, 185–202.

Hanson, S. G., & Stein, J. C. (2015). Monetary policy and long-term real rates. *Journal of Financial Economics*, *115*(3), 429–448.

Harris, M., & Raviv, A. (1993). Differences of opinion make a horse race. *The Review of Financial Studies*, *6*(3), 473–506.

Harris, Z. S. (1954). Distributional structure. *Word*, *10*(2-3), 146–162.

Hayo, B. (2018). On standard-error-decreasing complementarity: Why collinearity is not the whole story. *Journal of Quantitative Economics*, *16*(1), 289–307.

Hayo, B., Henseler, K., & Rapp, M. S. (2019). Estimating the monetary policy interest-rate-to-performance sensitivity of the european banking sector at the zero lower bound. *Finance Research Letters*, *31*.

Hayo, B., Henseler, K., Rapp, M. S., & Zahner, J. (2020). Complexity of ECB communication and financial market trading. *MAGKS Joint Discussion Paper Series in Economics*, *19-2019*.

Hayo, B., Kutan, A. M., & Neuenkirch, M. (2008). Communicating with many tongues: FOMC speeches and US financial market reaction. *MAGKS Joint Discussion Paper Series in Economics*, *08-2008*.

Hayo, B., & Neuenkirch, M. (2013). Do Federal Reserve Presidents communicate with a Regional bias? *Journal of Macroeconomics*, *35*, 62–72.

Hayo, B., & Neuenkirch, M. (2015a). Central bank communication in the financial crisis: Evidence from a survey of financial market participants. *Journal of International Money and Finance*, *59*, 166–181.

Hayo, B., & Neuenkirch, M. (2015b). Self-monitoring or reliance on media reporting: How do financial market participants process central bank news? *Journal of Banking & Finance*, *59*, 27–37.

Henseler, K., & Rapp, M. S. (2018). Stock market effects of ECB's asset purchase programmes: Firm-level evidence. *Economics Letters*, *169*, 7–10.

Hernández-Murillo, R., & Shell, H. (2014). The rising complexity of the FOMC statement. *Economic Synopses*, *23*.

Hinterlang, N. (2020). Predicting monetary policy using artificial neural networks. *Deutsche Bundesbank Discussion Paper*, *2020*(44).

Hirshleifer, D. (2001). Investor psychology and asset pricing. *The Journal of Finance*, *56*(4), 1533–1597.

Hollo, D., Kremer, M., & Lo Duca, M. (2012). CISS-a composite indicator of systemic stress in the financial system. *ECB Working Paper Series*, *1426*.

Hong, H., & Stein, J. C. (1999). A unified theory of underreaction, momentum trading, and overreaction in asset markets. *The Journal of finance*, *54*(6), 2143–2184.

Hornik, K., & Grün, B. (2011). Topicmodels: An R package for fitting topic models. *Journal of Statistical Software*, *40*(13), 1–30.

Hu, N., & Sun, Z. (2021). *Uncertain talking at central bank's press conference: News or noise?* (SSRN Scholarly Paper No. 4). Social Science Research Network. Elsevier BV.

Hussain, S. M. (2011). Simultaneous monetary policy announcements and international stock markets response: An intraday analysis. *Journal of Banking & Finance*, *35*(3), 752–764.

Istrefi, K., Odendahl, F., & Sestieri, G. (2020). Fed communication on financial stability concerns and monetary policy decisions: Revelations from speeches. *Banque de France Working Paper*, *779*.

Jarociński, M., & Karadi, P. (2020). Deconstructing Monetary Policy Surprises—The Role of Information Shocks. *American Economic Journal: Macroeconomics*, *12*(2), 1–43.

Kaefer, B. (2014). The Taylor rule and financial stability – a literature review with application for the eurozone. *Review of Economics*, *65*(2), 159–192.

Kalamara, E., Turrell, A., Redl, C., Kapetanios, G., & Kapadia, S. (2020). Making text count: Economic forecasting using newspaper text. *Bank of England Working Paper Series*, *865*.

Kandel, E., & Pearson, N. D. (1995). Differential interpretation of public signals and trade in speculative markets. *Journal of Political Economy*, *103*(4), 831–872.

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, *90*(430), 773–795.

Kincaid, J. P., Fishburne Jr., R. P., Rogers, R. L., & Chissom, B. S. (1975). *Derivation of new readability formulas (automated readability index, fog count and flesch reading ease formula) for navy enlisted personnel* (tech. rep.). Naval Technical Training Command Millington TN Research Branch.

Kohn, D. L., & Sack, B. P. (2003). *Central bank talk: Does it matter and why?* (Tech. rep.). Board of Governors of the Federal Reserve System (US).

Koop, G. M. (2003). *Bayesian econometrics* (1st ed.). John Wiley & Sons Inc.

Kuttner, K. N. (2001). Monetary policy surprises and interest rates: Evidence from the Fedfunds futures market. *Journal of monetary economics*, *47*(3), 523–544.

Kuttner, K. N. (2018). Outside the box: Unconventional monetary policy in the great recession and beyond. *Journal of Economic Perspectives*, *32*(4), 121–46.

Lane, P. (2019). Wir wollen positive Zinsen. *Handelsblatt: Düsseldorf*, *187*, 26–27.

Lau, J. H., & Baldwin, T. (2016). An empirical evaluation of doc2vec with practical insights into document embedding generation, In *Proceedings of the 1st Workshop on Representation Learning for NLP*.

Le, Q., & Mikolov, T. (2014). Distributed representations of sentences and documents, In *International conference on machine learning*. Proceedings of Machine Learning Research.

Lee, K., Morley, J., & Shields, K. (2015). The meta Taylor rule. *Journal of Money, Credit and Banking*, *47*(1), 73–98.

Lee, K., Olekalns, N., & Shields, K. (2013). Meta Taylor rules for the UK and Australia; accommodating regime uncertainty in monetary policy analysis using model averaging methods. *The Manchester School*, *81*, 28–53.

Ley, E., & Steel, M. F. J. (2009). On the effect of prior assumptions in bayesian model averaging with applications to growth regression. *Journal of Applied Econometrics*, *24*(4), 651–674.

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

Loughran, T., & McDonald, B. (2016). Textual Analysis in Accounting and Finance: A Survey. *Journal of Accounting Research*, *54*(4), 1187–1230.

Lucca, D. O., & Trebbi, F. (2009). *Measuring Central Bank Communication: An Automated Approach with Application to FOMC Statements* (Working Paper No. 15367). National Bureau of Economic Research.

Manzini, T., Lim, Y. C., Tsvetkov, Y., & Black, A. W. (2019). *Black is to criminal as caucasian is to police: Detecting and removing multiclass bias in word embeddings* (arXiV Preprint Working Papers No. 1904.04047). Cornell University.

Masciandaro, D., Romelli, D., & Rubera, G. (2020). *Tweeting on monetary policy and market sentiments: The central bank surprise index* (SSRN Scholarly Paper No. 3552699). Social Science Research Network. Elsevier BV.

McEwen, R. A., & Hunton, J. E. (1999). Is analyst forecast accuracy associated with accounting information use? *Accounting Horizons*, *13*(1), 1–16.

Meade, E. E. (2005). The FOMC: Preferences, voting, and consensus. *Federal Reserve Bank of St. Louis Review*, *87*(2).

Mikolov, T., Chen, K., Corrado, G. S., & Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiV Preprint Working Papers*, (1301.3781).

Mikolov, T., Sutskever, I., Chen, K., Corrado, G., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. *arXiV Preprint Working Papers*, (1310.4546).

Mikolov, T., Yih, W.-t., & Zweig, G. (2013). Linguistic regularities in continuous space word representations, In *Proceedings of the 2013 conference of the north american chapter of the association for computational linguistics: Human language technologies*.

Milani, F. (2004). *Monetary policy with a wider information set: A bayesian model averaging approach* (SSRN Scholarly Paper No. 490082). Social Science Research Network. Elsevier BV.

Milani, F. (2008). Monetary policy with a wider information set: A bayesian model averaging approach. *Scottish Journal of Political Economy*, *55*(1), 1–30.

Miles, D., Panizza, U., Reis, R., & Ubide, A. (2017). *And yet it moves: Inflation and the great Recession*. International Center for Monetary; Banking Studies/Centre for Economic.

Miller, B. P. (2010). The effects of reporting complexity on small and large investor trading. *The Accounting Review*, *85*(6), 2107–2143.

Molodtsova, T., & Papell, D. H. (2013). Taylor rule exchange rate forecasting during the financial crisis, In *NBER international seminar on macroeconomics*. University of Chicago.

Moral-Benito, E. (2015). Model averaging in economics. *Journal of Economic Surveys*, *29*(1), 46–75.

Nakamura, E., & Steinsson, J. (2018). High-frequency identification of monetary non-neutrality: The information effect. *The Quarterly Journal of Economics*, *133*(3), 1283–1330.

Obstfeld, M., & Rogoff, K. S. (2005). Global current account imbalances and exchange rate adjustments. *Brookings Papers on Economic Activity*, *2005*(1), 67–146.

Orphanides, A. (2001). Monetary policy rules based on real-time data. *American Economic Review*, *91*(4), 964–985.

Paloviita, M., Haavio, M., Jalasjoki, P., Kilponen, J., & Vänni, I. (2020). Reading between the lines - using text analysis to estimate the loss function of the ECB. *Bank of Finland Research Discussion Paper Series*, *12-2020*.

Peek, J., Rosengren, E. S., & Tootell, G. (2016). Does Fed policy reveal a ternary mandate? *Federal Reserve Bank of Boston Working Papers*, *16–11*.

Pennington, J., Socher, R., & Manning, C. D. (2014). Glove: Global vectors for word representation, In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*.

Picault, M., & Renault, T. (2017). Words are not all created equal: A new measure of ECB communication. *Journal of International Money and Finance*, *79*, 136–156.

Praet, P. (2014). Current issues and challenges for central bank communication, In *Conference on "the ECB and its watchers XV"*, Frankfurt am Main.

Praet, P. (2017). Communicating the complexity of unconventional monetary policy in emu, In *ECB central bank communications conference*, Frankfurt am Main.

Rehurek, R., & Sojka, P. (2011). Gensim–python framework for vector space modelling. *NLP Centre, Faculty of Informatics, Masaryk University, Brno, Czech Republic*, *3*(2).

Rosa, C. (2011a). The validity of the event-study approach: Evidence from the impact of the Fed's monetary policy on us and foreign asset prices. *Economica*, *78*(311), 429–439.

Rosa, C. (2011b). Words that shake traders: The stock market's reaction to central bank communication in real time. *Journal of Empirical Finance*, *18*(5), 915–934.

Rosa, C., & Verga, G. (2008). The Impact of Central Bank Announcements on Asset Prices in Real Time. *International Journal of Central Banking*, *4*(2), 175–217.

Roskelley, K. D. (2016). Augmenting the Taylor rule: Monetary policy and the bond market. *Economics Letters*, *144*, 64–67.

Sauer, S., & Sturm, J.-E. (2007). Using Taylor rules to understand European Central Bank monetary policy. *German Economic Review*, *8*(3), 375–398.

Schmeling, M., & Wagner, C. (2019). Does central bank tone move asset prices? *CEPR Discussion Papers*, *13490*.

Shapiro, A. H., Sudhof, M., & Wilson, D. J. (2020). Measuring news sentiment. *Journal of Econometrics*.

Shapiro, A. H., & Wilson, D. J. (2019). Taking the Fed at its word: A new approach to estimating central bank objectives using text analysis. *Federal Reserve Bank of San Francisco Working Papers Series*, *2019-02*.

Shiller, R. J. (2017). Narrative economics. *American Economic Review*, *107*(4), 967–1004.

Smales, L., & Apergis, N. (2017a). Does more complex language in FOMC decisions impact financial markets? *Journal of International Financial Markets, Institutions and Money*, *51*, 171–189.

Smales, L., & Apergis, N. (2017b). Understanding the impact of monetary policy announcements: The importance of language and surprises. *Journal of Banking & Finance*, *80*, 33–50.

Smith, M., & Taffler, R. (1992). Readability and understandability: Different measures of the textual complexity of accounting narrative. *Accounting, Auditing & Accountability Journal*, *5*(4), 84–98.

Stark, J. (2010). Geld- und Fiskalpolitik während und nach der Krise, In *Stuttgarter Steuerkongress*, Stuttgart.

Stigler, G. J. (1961). The economics of information. *Journal of political economy*, *69*(3), 213–225.

Sturm, J.-E., & Wollmershäuser, T. (2008). The stress of having a single monetary policy in europe. *KOF Working Papers*, *190*.

Svensson, L. E. (2003). What is wrong with Taylor rules? Using judgment in monetary policy through targeting rules. *Journal of Economic Literature*, *41*(2), 426–477.

Swanson, E. T. (2021). Measuring the effects of federal reserve forward guidance and asset purchases on financial markets. *Journal of Monetary Economics*, *118*, 32–53.

Sweeney, C., & Najafian, M. (2019). A transparent framework for evaluating unintended demographic bias in word embeddings, In *Proceedings of the 57th annual meeting of the association for computational linguistics*.

Tadle, R. C. (2021). FOMC minutes sentiments and their impact on financial markets. *Journal of Economics and Business*, 106021.

Taylor, J. B. (1993). Discretion versus policy rules in practice. *Carnegie-Rochester Conference Series on Public Policy*, *39*, 195–214.

Tetlock, P. C. (2007). Giving content to investor sentiment: The role of media in the stock market. *The Journal of Finance*, *62*(3), 1139–1168.

Tillmann, P. (2020). Financial Markets and Dissent in the ECB's Governing Council. *MAGKS Joint Discussion Paper Series in Economics*, *48-2020*.

Tillmann, P., & Walter, A. (2018). ECB vs Bundesbank: Diverging Tones and Policy Effectiveness. *MAGKS Joint Discussion Paper Series in Economics*, *20-2018*.

Tobback, E., Nardelli, S., & Martens, D. (2017). Between hawks and doves: Measuring central bank communication. *ECB Working Paper Series*, *2085*.

Turian, J., Ratinov, L., & Bengio, Y. (2010). Word representations: A simple and general method for semi-supervised learning, In *Proceedings of the 48th annual meeting of the association for computational linguistics*, Association for Computational Linguistics.

Walsh, C. E. (2017). *Monetary theory and policy*. MIT Press Ltd.

Wischnewsky, A., Jansen, D.-J., & Neuenkirch, M. (2021). Financial stability and the fed: Evidence from congressional hearings. *Economic Inquiry*, *59*(3), 1192–1214.

Wittgenstein, L. (1958). *Philosophische Untersuchungen (Zweite Auflage). English edition.*

Wu, J. C., & Xia, F. D. (2016). Measuring the macroeconomic impact of monetary policy at the zero lower bound. *Journal of Money, Credit and Banking*, *48*(2-3), 253–291.

You, H., & Zhang, X.-j. (2009). Financial reporting complexity and investor underreaction to 10-k information. *Review of Accounting studies*, *14*(4), 559–586.

Zahner, J. (2020). Above, but close to two percent. evidence on the ECB's inflation target using text mining. *MAGKS Joint Discussion Paper Series in Economics*.

Zellner, A. (1986). On assessing prior distributions and bayesian regression analysis with g-prior distributions. In P. Goel & A. Zellner (Eds.), *Bayesian inference and decision techniques: Essays in honor of Bruno de Finetti* (pp. 233–243). New York, Elsevier Science Publishers.

Zeugner, S., & Feldkircher, M. (2015). Bayesian model averaging employing fixed and flexible priors: The BMS package for R. *Journal of Statistical Software*, *68*(4), 1–37.

Eidesstattliche Erklärung

Hiermit versichere ich an Eides statt, dass ich die vorliegende schriftliche Arbeit selbstständig und ohne fremde Hilfe verfasst, eventuelle Beiträge von Koautoren dokumentiert, keine anderen als die von mir angegebenen Hilfsmittel benutzt und alle vollständig oder sinngemäß übernommenen Zitate als solche gekennzeichnet, sowie die Dissertation in der vorliegenden oder einer ähnlichen Form noch bei keiner anderen in- oder ausländischen Hochschule anlässlich eines Promotionsgesuchs oder zu anderen Prüfungszwecken eingereicht habe.

Johannes Zahner

Marburg, der 3. August 2021