Inaugural-Dissertation zur Erlangung der Doktorwürde der

Wirtschaftswissenschaften des Fachbereichs

Wirtschaftswissenschaften der Justus-Liebig-Universität Gießen

Essays on Uncertainty and Communication in Monetary Policy

vorgelegt von

Martin Baumgärtner, M.Sc.

Erstgutachter:	Prof. Dr. Jens Klose
Zweitgutachter:	Prof. Dr. Peter Tillmann
Prüfungsvorsitzende:	Prof. Dr. Corinna Ewelt-Knauer
Einreichungstermin:	18.07.2022

Acknowledgments

Working in recent years has shown me that science constantly questions the status quo. I have also benefited from the fact that countless people around me have discussed with me, improved me and motivated me. Without claiming to be exhaustive, I would like to mention the most important ones here.

First of all, I would like to thank my supervisor Jens Klose, who made it possible for me to dive into science. He was a mentor to me in every respect. He gave me the freedom to pursue my ideas, and I could seriously discuss even uncommon research proposals. On the one hand, he was always able to recognise the weaknesses of my research and pointed them out to me. On the other hand, he built me up again and pointed out solutions when I was stuck. Every time I talked to him, I learned something, and my work improved.

Peter Tillmann, who created an ideal creative space for science at his chair. His open-mindedness towards new ideas and approaches is a shining example for me. He motivated me to look into a wide variety of topics, which in retrospect, all seem to have a connection.

Clarissa Kees was an essential pillar for me throughout the entire period. Her positive attitude to life and her love has given me many wonderful memories.

Science is not isolated thinking, but rather cooperation of people. Johannes Zahner plays a vital role for me in this. Together with him, I tackled the research project I am most proud of. Short phone calls turn into long conversations because his thirst for knowledge continues to infect me.

Finally, I would like to thank Timo Aldenhoff, Angelina Feustel, Matthias Klein, Martin Palzer und Annette, Franz and Simon Baumgärtner and my whole family for the many joys in my life.

Summary

If you tell a human being what will happen the next day, they will prepare for it. This simple idea has now become a cornerstone of economic policy. However, whether these preparations are crowned with success depends on many factors.

First it is necessary that we understand what is going to happen. For this, good communication is inevitable. But what constitutes effective communication is much more challenging to answer. It depends on the sender and receiver of the message and the circumstances and forms it is conveyed. Second, the message that is sent out must be credible. Announcing an unrealistic target or not delivering on a commitment can undermine this credibility. This then effectively leads to a loss of communication capability in the future.

In monetary policy, a central bank faces exactly this problem. Due to the financial crisis, many central banks have shifted their monetary policy to influencing the expectations of financial market participants and thus of society through forward guidance and purchase programs. Complex measures must be prepared appropriately for the target group and communicated via various channels. At the same time, it is of utmost importance that the announced measures are also implemented.

However, because these measures have only been in existence for 15 years, we are still lacking a detailed understanding of the impact channels in some cases. We know that the announcement is at least as necessary as the actual implementation of the measures. As soon as an announcement is made about the future, society tries to prepare for it and thus anticipates the effect. However, the literature shows that there are still many unanswered questions, as sometimes the central bank's intention differs from the markets' reaction.

This dissertation examines different questions based on four essays on the intersection of expectations, communication and how both can be measured empirically. I use established and new methodologies to empirically test the effects of monetary policy. More specifically, in the first essay, I use high-frequency data to measure the reaction of financial markets to European Central Bank (ECB) decisions and to quantify the macroeconomic effects. The second paper takes this

idea forward and shows that these effects depend on the uncertainty at the time of the announcement. The third paper focuses on how central bank texts can be measured empirically. The final paper demonstrates the existence and explores the effects of a previously unknown monetary policy surprise in the euro area, triggered by a break in central bank communication style.

My first paper examines the effect of ECB measures. Jens Klose and I use a VAR model to show the difference between conventional interest rate policy and communicative measures. We estimate monetary surprises for conventional, communication and asset purchase measures from the change in financial market variables during the announcement of ECB measures. Using these, we estimate a vector autoregressive model identified by external instruments. It shows that expansionary conventional and asset purchase measures lower the interest rate. This does not hold for communication measures: Contrary to expectations, expansionary measures do not increase inflation. By subdividing the surprises, we can show that the effect can be explained by the pattern of information shocks known from the literature. The comparison between the measures shows that the phenomenon does not occur with conventional measures.

My second paper returns to the topic of information shocks. The literature speaks of an information shock when stock prices fall (rise) together with an expansionary (restrictive) shock. While it is evident that this pattern exists, the theoretical explanation is disputed in the literature. I incorporate uncertainty into the analysis of monetary policy surprises and can thus show that uncertainty can explain the observed pattern at the time of the decision. When uncertainty is high, *information shocks* occur more frequently. Furthermore, I integrate uncertainty into a VAR model and can thus show that identification by uncertainty yields the impulse responses known from the literature. These findings provide possible alternative explanations for why information shocks occur and illustrate that uncertainty is essential for the effectiveness of specific monetary policy measures.

In the third paper, Johannes Zahner and I address how text data can be used in the quantitative analysis of central banks. Due to the focus of central banks on communication, the analysis of texts is becoming more and more important

in research. At the same time, there are advances in computational linguistics that make it possible to use language for analysis in a more nuanced way than before. The work brings both strands together: We collect the largest text dataset on central bank communication and compute and evaluate a language model adapted to central banks. In this way, we enable researchers to capture language in detail. Furthermore, we show in three economic applications how the model can be used in a classical economic context despite its many dimensions. First, we demonstrate that the goals of central banks are reflected in their texts. Next, we investigate the effect of communication similar to Mario Draghi's 'whatever it takes' speech and show that credit spreads can be lowered in periods of high uncertainty. As a final application, we show that central bank speech is not free of social stereotypes. We find a gender bias, which is, however, declining due to a change in central bank communication.

In my last paper, I focus on identifying monetary shocks in the European context. In 2016, the European Central Bank changed its communication structure and integrated the asset purchase programmes into its press release. I can show that due to this modification, two Quantitative Easing (QE) surprises occur from 2016: One during the press release and one during the press conference. Adapting the methodology established in the literature allows me to identify and compare the shocks. Interestingly, a difference emerges between the shocks. The short, pre-formulated message has a significantly more potent effect on stock prices than the shock during the press conference, where the communication is more spontaneous and detailed.

Contents

List of Figures

List of Tables

List of Acronyms

BIS Bank of International Settlement

BoJ Bank of Japan

BoE Bank of England

CISS Composite Indicator of Systemic Stress

DSGE Dynamic Stochastic General Equilibrium

EA-MPD Euro Area Monetary Policy Event-Study Database

ECB European Central Bank

FED Federal Reserve

FG forward guidance

FOMC Federal Open Market Committee

KNN K-Nearest-Neighbor

LDA Latent Dirichelet Allocation

LSAP large-scale asset purchases

NLP natural language processing

OIS overnight interest rate swap

QE Quantitative Easing

RA risk aversion component

RBNZ Reserve Bank of New Zealand

RND relative norm distance

UC uncertainty component

VAR vector autoregression

ZLB zero lower bound

1 Why central banks announcing liquidity injections is more effective than forward guidance.[*]

Martin Baumgärtner and Jens Klose[b]

Abstract

We distinguish the announcement effects of conventional and unconventional monetary policy measures on macroeconomic variables using a high-frequency data set that measures the impact of the European Central Bank's monetary policy decisions. For the period 2002 to 2019, we show that conventional and unconventional monetary policy measures differ considerably in their impact on inflation. While conventional measures show the expected response, that is, an interest rate cut increases inflation, unconventional measures appear to generally have no significant influence. However, this does not hold for quantitative easing, which is found to have a similar influence on inflation as the conventional interest rate changes.

Keywords: Unconventional monetary policy, high-frequency data, information shock, European Central Bank

JEL classification: E52, E58, C36

[*]This essay is published as Baumgärtner, M., & Klose, J. (2021). Why central banks announcing liquidity injections is more effective than forward guidance. International Finance, 24(2), 236-256.

[b]THM Business School, Germany.

1.1 Introduction

The financial crisis starting in 2008/2009 changed traditional monetary policy. The interest rate channel's effectiveness reached its limit, at the zero lower bound (ZLB), and central banks therefore broadened their range of instruments. However, these new measures have raised new challenges for researchers. It is not easy to map all policies through a single model and analyze their effects simultaneously. Central bank measures can be roughly broken down into conventional and unconventional measures. While the effects of conventional measures on important macroeconomic variables have been extensively investigated, the empirical effects of forward guidance and QE are far less investigated and still a controversial matter in the literature. Forward guidance is difficult to capture, because there is no clear indicator that makes the effect objectively observable.

We fill this gap in the literature by estimating the announcement effect of all monetary policy measures in the euro area with a single model.[1] This allows for a detailed comparison of the macroeconomic impacts of the various measures. More specifically, we use the findings of Altavilla et al. (2019) and the Euro Area Monetary Policy Event-Study Database (EA-MPD), published by the authors, to estimate the monetary policy surprises around ECB meeting dates. These estimations, in turn, are employed to determine the effects of individual measures on the key macroeconomic variables in an external instruments vector autoregression (VAR) approach. Using monthly data for the period between 2002:01 and 2019:06, we indeed find significant differences between conventional and unconventional measures, but also between forward guidance and QE.

We can further show that Delphic shocks can be found through the EA-MPD data. These shocks are events that cannot be explained by basic economic theory. Contrary to expectations, the central bank's expansionary shock does not lead to rising inflation but falling inflation. One explanation is that the markets are reacting to the central bank's negative expectations for the future. The term Delphic refers to Greek mythology, where the Oracle of Delphi makes predic-

[1] Besides announcement effects, application effects of monetary policy can also arise. Although these are not addressed in our high-frequency model, they have already been extensively investigated by Haitsma et al. (2016), Jäger and Grigoriadis (2017), and Borrallo Egea and Hierro (2019), and Dominguez-Torres and Hierro (2020).

tions that need to be interpreted by the individuals, and thus trigger actions. By subdividing the individual measures more precisely, we can further narrow down the effect and increase our knowledge about Delphic shocks. Especially in the case of short-term expectation-forming timing measures, a clear difference in macroeconomic effects can be seen. Information effects seem to play a vital role here. In contrast, the difference is not clear with conventional policies. Based on this result, we can empirically validate the assumption that Delphic shocks are particularly important in forward guidance and less so in conventional policy.
The remainder of the paper proceeds as follows. Section 1.2 reviews the literature, including an overview of the different approaches to distinguish monetary policy measures. Section 1.3 describes the methodology, that is, the econometric framework, the construction and justification of the instruments, and the data used in this study. Section 1.4 presents our estimation results, showing, first, differences between conventional and unconventional measures; second, differences between the various forms of unconventional measures; and, third, differences between market reactions toward unconventional qualitative announcements such as forward guidance. To test if our results are influenced by a structural break around the financial crisis, we conduct a robustness test in Section 1.5. In Section 1.6, we present a possible explanation for our results by splitting the effects of forward guidance by the different market reactions. Section 1.7 concludes the paper and draws policy conclusions.

1.2 Related literature

The financial crisis demonstrated that the existing transmission channels of monetary policy can be affected by uncertainty. At the same time, empirical evidence shows a negative trend in inflation developments, that could even be beyond the reach of central banks (Bonam et al., 2019). Traditional empirical approaches to identify monetary policy shocks reach their limits because of the common use of a short-term interest rate when the ZLB becomes binding. The ECB, as other central banks in industrialized countries, therefore switched its policy to include additional unconventional measures. Therefore, other ways must be found to model these kinds of shocks.[2]

The simplest and most straightforward way is to switch to longer-term interest rates as a policy variable, to avoid the problem of variables that equal zero. However, this approach is also influenced by the ZLB, that is, long-term interest rates can approach zero if the zero-interest period lasts too long (Swanson and Williams, 2014). Moreover, when relying on longer-term interest rates, the risk of factors besides monetary policy (e.g., changing market expectations) biasing this variable increases.

A second approach besides classical interest rates involves artificial (shadow) rates that include unconventional measures (Krippner, 2013; Wu and Xia, 2016). Recent studies urge caution, since the estimates are very sensitive (Krippner, 2020).

Another method of identification in unconventional times is a combination of sign and zero restrictions (Arias et al., 2018). A large strand of the literature combines this method with central bank assets (Gambacorta et al., 2014; Boeckx, Dossche, and Peersman, 2017; Burriel and Galesi, 2018). Whether this combination identifies unconventional shocks is currently being discussed (Elbourne and Ji, 2019; Boeckx, Dossche, Galesi, et al., 2019; Elbourne, 2019).

Since Kuttner (2001), there has been a growing literature using high-frequency data sets. The author has shown that financial variables react to changes in US Federal Reserve policy. Building on these insights, Gürkaynak et al. (2005) identify different monetary shocks, namely, a target factor and a path factor. Brand et al. (2010) develop this method further concerning the ECB, not only by considering the differences before and after the decision, but also by separating the effect of the press release and subsequent press conference. To the best of our knowledge, Gertler and Karadi (2015) are the first to use these high-frequency monetary shocks in an external instrument VAR. The assumption made in these kinds of estimations is that no other shocks distort the results if the time window is small enough. The authors find different effects of conventional and high-frequency identification in VAR models. Swanson (2021) expands the previous identification of shocks. The author shows that it is possible to extract the effects of large-scale asset purchases (LSAP) for the period from 2009 to 2015 in

[2] For a detailed overview of these approaches, see Rossi (2020).

the United States.

An even more accurate approach to identifying shocks from high-frequency data is made by Andrade and Ferroni (2021). They combine principle component analysis and sign restrictions to distinguish between Delphic and standard forward-guidance shocks. These shocks were first established by Campbell et al. (2012). In their theory, a Delphic shock lowers interest rates, but has a dampening effect on stock prices due to new, worse information from the central bank. Jarociński and Karadi (2020) and Kerssenfischer (2019) show that Delphic shocks from central banks play an essential role in both the United States and the euro area.

Altavilla et al. (2019) build on the previous findings and address the reality of Delphic and monetary shocks. They provide a high-frequency data set for the euro area and extract various orthogonal shocks for the press release and press conference. The authors provide a first insight into how the shocks affect individual financial variables, but they do not address the macroeconomic effects.

1.3 Methodology

The publication of the EA-MPD by Altavilla et al. (2019) provided the opportunity to examine the influence of monetary policy measures on the European economy. It is possible to distinguish between individual orthogonal measures, such as interest rate policy, forward guidance, or QE, and examine the different effects. In the following, we will first describe the detailed formulation of our econometric model and then construct our instruments. In a third step, we combine both with data and show that the instruments we have chosen are permissible in our model and produce reliable results.

1.3.1 Econometric model

In our model, we want to estimate the reactions of economic variables to different monetary policy shocks ϵ_t^p. However, since most of the variables are affected by various shocks simultaneously, we use an approach with exogenous instruments developed by Stock and Watson (2012) and Mertens and Ravn (2013) and also applied by Gertler and Karadi (2015). This approach allows us to isolate the individual shocks that simultaneously affect our policy variables.

Let Y_t be an $(N \times 1)$ matrix of N economic variables in T periods. Consider a VAR model in general structural form:

$$AY_t = C + \sum_{j=1}^{J} C_j Y_{t-j} + \epsilon_t \tag{1.1}$$

where C represents a constant, while A and C_j form the coefficient matrices, including J lags. Inverting A leads to

$$Y_t = A^{-1}C + \sum_{j=1}^{J} A^{-1} C_j Y_{t-j} + v_t \tag{1.2}$$

with v_t denoting the reduced-form residuals. They are connected to the structural shocks ϵ_t by

$$v_t = A^{-1}\epsilon_t \tag{1.3}$$

Replacing $S = A^{-1}$ and (1.3) in (1.2) yields the following model:

$$Y_t = SC + \sum_{j=1}^{J} SC_j Y_{t-j} + S\epsilon_t \tag{1.4}$$

We are especially interested in estimating one column of S. The column s^p indicates how the reduced-form residuals v_t change in response to a unit increase in the structural shock ϵ_t^p. We follow Gertler and Karadi (2015) and focus our analysis on column $s^{mp} = S_{\cdot,mp}$, which reflects the reaction of our variables to a monetary policy shock. All the other columns are represented by $s^q = S_{\cdot,q}$. Together with (1.3), we obtain the following equations:

$$v_t^{mp} = s^{mp}\epsilon_t^{mp} \tag{1.5a}$$

$$v_t^{q} = s^{q}\epsilon_t^{mp} \tag{1.5b}$$

These can be solved for v_t^q with

$$v_t^q = \frac{s^q}{s^{mp}} * v_t^{mp} \tag{1.6}$$

The fraction corresponds to a unit effect normalization. A unit shock in ϵ_t^{mp} increases v_t^{mp} by the same amount. All the other effects on the variables are expressed proportionally. If we want to solve this equation, we come across an endogeneity problem. To resolve this issue, we use a two-stage approach with an instrument Z. A good instrument must, according to Stock and Watson (2018), have the following characteristics to obtain consistent estimates:

$$E[\epsilon_t^{mp} Z'] = \alpha \neq 0 \qquad \text{(relevance)} \tag{1.7a}$$

$$E[\epsilon_t^{q} Z'] = 0 \qquad \text{(exogeneity with respect to other current shocks)} \tag{1.7b}$$

Therefore, an instrument is needed that is highly correlated with the monetary policy shock ϵ_t^{mp}, but not correlated with any other shock ϵ_t^q at the same time. With a feasible instrument and the reduced-form variance-covariance matrix Σ, we obtain a consistent estimation of s by using a two-stage approach. In the first stage, we regress v_t^{mp} on Z_t to estimate the fitted value $\hat{v}_t^{mp}$. We thus obtain the part of the variation in v_t^{mp} that is only due to a structural shock ϵ_t^{mp}. If we insert this in (1.6), we obtain

$$v_t^q = \frac{s^q}{s^{mp}} * \hat{v}_t^{mp} + \xi_t \tag{1.8}$$

The second-stage regression (1.8) yields a consistent estimation of $\frac{s^q}{s^{mp}}$. With Σ, we can then determine all the components of s^{mp}, which, in turn, allows us to estimate impulse responses from our partially identified structural VAR model: [3]

$$Y_t = SC + \sum_{j=1}^{J} SC_j Y_{t-j} + s\epsilon_t^{mp} \tag{1.9}$$

[3]For a detailed derivation, see Gertler and Karadi (2015).

1.3.2 Instrument choice

Two points must be considered when choosing the instrument: the instrument must be exogenous (1.7b) and relevant (1.7a). While we will empirically prove the validity of condition (1.7a) in the following Section, Section 1.3.3, the validity of (1.7b) follows from theoretical considerations, which are described below.

We will apply the EA-MPD to extract monetary surprises measured by high-frequency deviations of financial variables around the ECB press release and press conference. The advantage is that these high-frequency deviations are most likely to be driven only by the ECB's decision. According to Kuttner (2001), there will probably be no effects in this period, and certainly no systematically distorting, ones.[4] Therefore, condition (2.11b) should be fulfilled, since we use a very narrow time window around the press release and press conference.

In the following, we briefly describe the replication of the four different monetary policy surprises, based on the work of Gürkaynak et al. (2005), Brand et al. (2010), and Swanson (2021), and Altavilla et al. (2019). Since Kuttner (2001), we know that the central bank's measures influence high-frequency data during central bank announcements. However, it is not easy to attribute changes to specific policies; the effects will overlap and influence each other, so that they cannot be observed directly. Therefore, latent factor models are used to separate the underlying unobservable influences and to determine how many factors are sufficient to describe our high-frequency data accurately.

The factor model has the equation

$$X^w = F^w \Lambda^w + \epsilon^w \tag{1.10}$$

with w in $\{$*press release, press conference*$\}$

where X^w is the change in the overnight interest rate swap (OIS) with maturities from one month to 10 years, F is an $(N \times T)$ matrix of latent factors, Λ comprises

[4] Furthermore, Brand et al. (2010) and Altavilla et al. (2019) control for a possible effect in this time window, the publication of US labor market figures. They find no evidence of any impact on European financial market variables during this time window.

the factor loadings, and ϵ is the idiosyncratic variation. We can estimate the latent factors (1.10) by using principal components on X^w. The matrix rank test of Cragg and Donald (1997) is used to determine the number of underlying factors in each subset. We find one latent factor for the press release window and two factors for the conference window in the pre-QE period and three for the full sample.[5] The factors alone are difficult to interpret in terms of content, since each factor is usually correlated with all OIS futures. This issue can be resolved by introducing restrictions through rotation of the factor matrix to the factor loadings:

$$F^{w,*} = F^w U \tag{1.11}$$

with $UU' = I$.

We use the restrictions established by Gürkaynak et al. (2005) to determine the first three factors and that established by Swanson (2021) for the fourth factor. The rotation is performed so that the second and third factors are not correlated with the monthly OIS rate, and, simultaneously, the variance of the third factor is minimal for the pre-crisis period. Since $UU' = I$, the factors are orthogonal to each other.

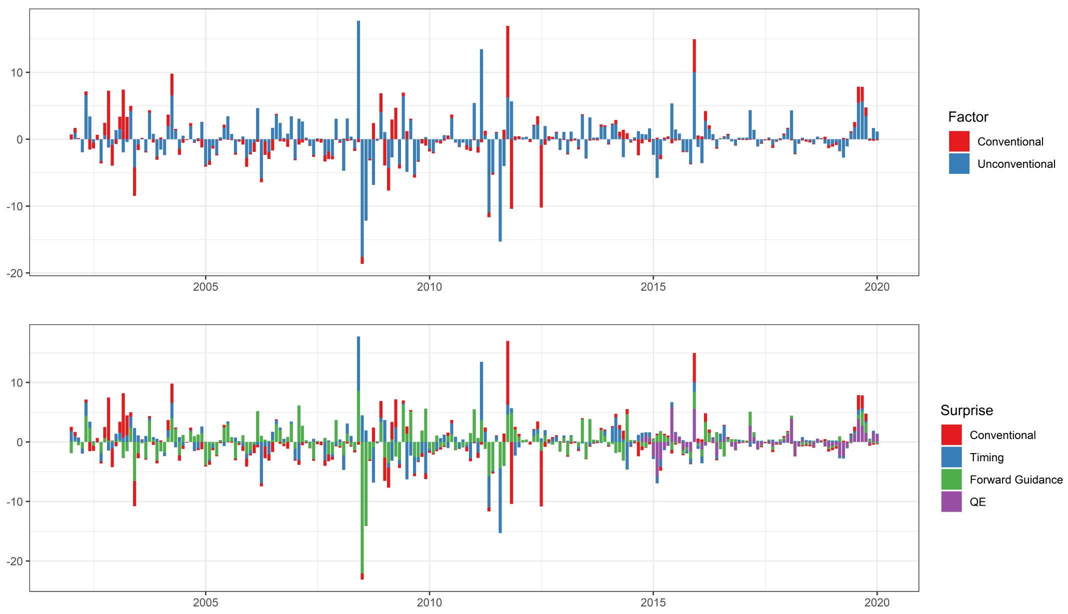
The press release window's resulting factor is called conventional surprise, because it loads strongly on the one-month OIS rates. This is the theoretical mechanism of a conventional interest rate policy.[6] These results are consistent with the expected functioning: conventional surprises are based on market reactions to the ECB press release directly after the ECB governing council meeting. Till the end of 2014, it contains only the pure policy rate decision. In 2016, there was an announcement that other measures would be intimated. After 2016:03, a short note about the concrete implementation of new measures was attached. Over the vast majority of the observation period, these surprises reflect surprises in interest rates, which are, by definition, conventional monetary policy.

[5] See Altavilla et al. (2019) for detailed results, which we can reconstruct.

[6] Altavilla et al. (2019) call it the conventional shock target. We use the first expression for a more intuitive understanding.

The first factor loads on the OIS rates under one year in the press conference window, but not on the one-month OIS rate. This factor is therefore forward guidance with a short time horizon, and we name it timing. In contrast, the second factor, by design, does not influence the one-month OIS rates, but loads most strongly on the medium term, that is, two- to five-year OIS rates. We therefore call it forward guidance. The third factor has the greatest influence on 10-year OIS rates. Besides, it has been rotated so that its influence before the financial crisis is minimal. This result corresponds to the theoretical functioning of QE. The combination of the three press conference surprises, moreover, sums up to unconventional surprises. In addition, we construct total surprises, which include all factors simultaneously. The influence of the conventional and unconventional measures at different time points is shown in Figure 1.1.

Figure 1.1: Estimated Factors



Notes: Sample period: 2002:01- 2019:06, accumulated factors in basis points.

By construction, QE started in 2014:10, with the introduction of the Covered Bond Purchase Programme 3, which later became part of the Asset Purchase Programme, which was introduced in 2015:01 and started officially in 2015:03.[7]

[7]Note that, for this reason, all models that only contain QE shocks are estimated with data starting 2014:10. The approach of Gertler and Karadi (2015) of estimating the different stages for different time spans to increase efficiency is impossible. The problem is not that high-frequency data are not available, but that there was no QE before 2014:10.

Note that expectations already played a role before the financial crisis, even though they were not an official policy tool of the ECB. This can be explained by the influence of central bank communication on market expectations. Before forward guidance was explicitly introduced, ECB press conferences were used to asking about the central bank's expectations regarding its future policies. Even though these questions were answered very restrictively, the answers seem to have affected medium-term OIS rates. However, compared to the period after 2008, when active forward guidance was applied, the shocks were substantially lower in the pre-crisis period.

Since these surprises are estimated with other macroeconomic variables, the shocks must be transformed into monthly data. Following Gertler and Karadi (2015), we use monthly average surprises. The shock values of the 31 elapsed days are added up and, in the next step, the arithmetic mean of all the accumulated values in each month is formed. This procedure accounts for the effect of variable meeting dates within a month. Shocks at the beginning of a month are given a higher weight, whereas shocks at the end of the month are more relevant to the next period.

1.3.3 Data

The endogenous variables Y_t in our model consist of *Output* (ECB industrial production),[8] *Prices* (ECB Harmonised Index of Consumer Prices), *Commodities* (International Monetary Fund Primary Commodity Price index), *Stock prices* (Euro Stoxx 50), *Uncertainty* (ECB Composite Indicator of Systemic Stress (CISS)), and two-year German government bonds ($DE2Y$).[9]

We use German government bonds, since the risk component in interest rates should be minimal here and not distorted by speculation. At the same time, this is potentially not the case for other euro area countries. Jarociński and Karadi (2020) also use German government bonds for this reason. The variable $DE2Y$ shows the best suitability, since the correlation between all surprises and the resid-

[8] To check for the influence of the construction sector, we conducted the analysis with industrial production, but excluding production. The results are very similar and available upon request.

[9] The variables *Output*, *Prices*, *Commodities*, and *Stock prices* are in logarithmic form. All four variables are seasonally adjusted.

uals are large enough to minimize the risk of biased estimates (see (2.11a)). This is probably because we compare both short- and long-term measures. Therefore, $DE2Y$ is a reasonable compromise.[10]

The Akaike information criterion suggests a maximum of $J = 3$ lags, which seems realistic compared to other VAR studies for the euro area (Gambacorta et al., 2014; Boeckx, Dossche, and Peersman, 2017).

Table 1.1: Data Overview

Variable	Proxy	Source	Seasonal adjusted and logarithms
Output	industrial production excluding construction	ECB	yes
Prices	harmonized index of consumer prices	ECB	yes
Commodities	Primary Commodity Price index	IMF	yes
Stock prices	Euro Stoxx 50	ECB	yes
Uncertainty	Composite Indicator of Systemic Stress (CISS)	ECB	no
Bond	2-year German government bonds	Bundesbank and Altavilla et al. (2019) Replication data [Link]	no

The idea is to use different surprises in our model, to compare the impacts on economic variables. Our instruments Z will be the monetary policy surprises from the previous section. Therefore, we will estimate our model with one instrument each, where our instrument is alternately one of the surprises found before.[11]

When it comes to instrument estimations, the challenge is to find a suitable in-

[10] We also checked other possible candidates that could have similar properties, that is, Euribor rates, OIS, other euro area countries' bonds, and different maturities. The $DE2Y$ model performed the best in this respect. The results for the other variables are available from the authors upon request.

[11] For the QE surprises, the series will start in 2014:10 due to design.

strument that meets conditions (2.11b) and (2.11a). Condition (2.11b) should be fulfilled by our choice of instruments, as described above. Condition (2.11a) means that the instrument should be correlated with our monetary policy shock and therefore have explanatory power. To test whether our instruments are suitable, we regress the five-year German government bond ($DE5Y$) residual ($\hat{v}_t$) on our factors separately. Table 1.2 reports the regression results for each shock, as well as the unconventional and total shocks, as described above. It should be noted that F-statistics do not reflect the importance of the factor in the period, but only the strength of the link between the shocks and the residuals of the model.

Table 1.2: F-Statistic of the regression of residuals on Z

	Dependent variable:					
	residual DE5Y					
	(1)	(2)	(3)	(4)	(5)	(6)
Conventional	0.013^{***} (0.004)					
Timing		0.022^{***} (0.007)				
Forward-guidance			0.014^{***} (0.003)			
QE				0.005^{***} (0.002)		
Unconventional					0.016^{***} (0.003)	
Total						0.014^{***} (0.002)
Constant	0.001 (0.009)	0.001 (0.009)	0.0001 (0.009)	0.001 (0.005)	0.001 (0.008)	0.002 (0.009)
R-squared	0.028	0.101	0.096	0.069	0.178	0.191
robust F-statistic	9.668	11.063	21.162	10.992	36.624	50.105
Note:						*p<0.1; **p<0.05; ***p<0.01

The robust F-statistic is above the value of 10 for all factors, except conventional ones.[12] This is a guideline for making a strong instrument (Stock and Yogo, 2001). Therefore, we conclude that our factors are suitable instruments. The combination of individual shocks (unconventional and total) is also highly significant and therefore provides a powerful instrument for the ECB's overall monetary policy strategy. The fit of the data as modeled by the $R-squared$ value is similar to that in papers using US data and the same methodology (Gertler and Karadi, 2015). To avoid the risk of a weak instrument bias, we use the robust confidence intervals developed by Montiel Olea et al. (2020). These intervals are not affected

[12]We tried other variables and other VAR specifications. In the few cases in which the F-statistics increase slightly for conventional factors, they drop sharply for the other factors. To establish comparability, we stick to DE2Y in our analysis. A similar VAR, with DE5Y as the monetary policy variable, yields a sufficiently large F-statistic for the conventional factor in the full sample that the risk of a weak instrument can be ruled out, and it provides very similar impulses responses and confidence intervals. The results are available from the authors upon request.

by instrument strength and convergence toward the standard confidence set when the instrument is vital. The Wald statistic for the covariance between the instrument and the normalized variable is high enough that the robust confidence set will be a bounded interval for every horizon.

1.4 Results

To present the results, we use a general-to-specific approach. Thus, we begin by presenting the influence of the total factor shock before disentangling it into conventional and unconventional shocks in a second step and splitting up the unconventional shock into the three subcategories (timing, forward guidance, and QE).

1.4.1 Total shock

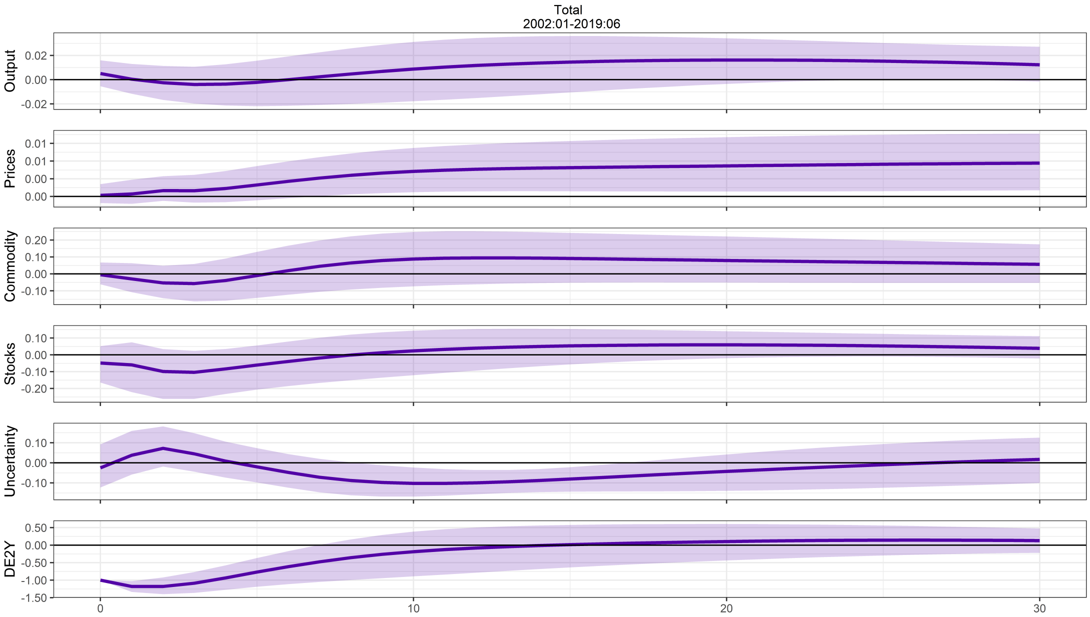
Starting with the total effect of monetary policy shocks in the euro area (Figure 1.2), we find the expected results for the full sample. An expansive monetary policy shock lowers $DE2Y$ on impact. Uncertainty declines in the medium term, and inflation increases with a short time lag and is significant at the 90% confidence level.

1.4.2 Conventional versus unconventional shocks

In the next step, we distinguish monetary policy shocks into conventional and unconventional policies. We also examine the difference between the press release announcement and the subsequent press conference. Therefore, we calculate two different VAR models (Figure 1.3). The point estimates for conventional measures are always higher than those for unconventional measures. Moreover, the aggregated unconventional monetary policy shocks appear to be even nonsignificant throughout the entire period.

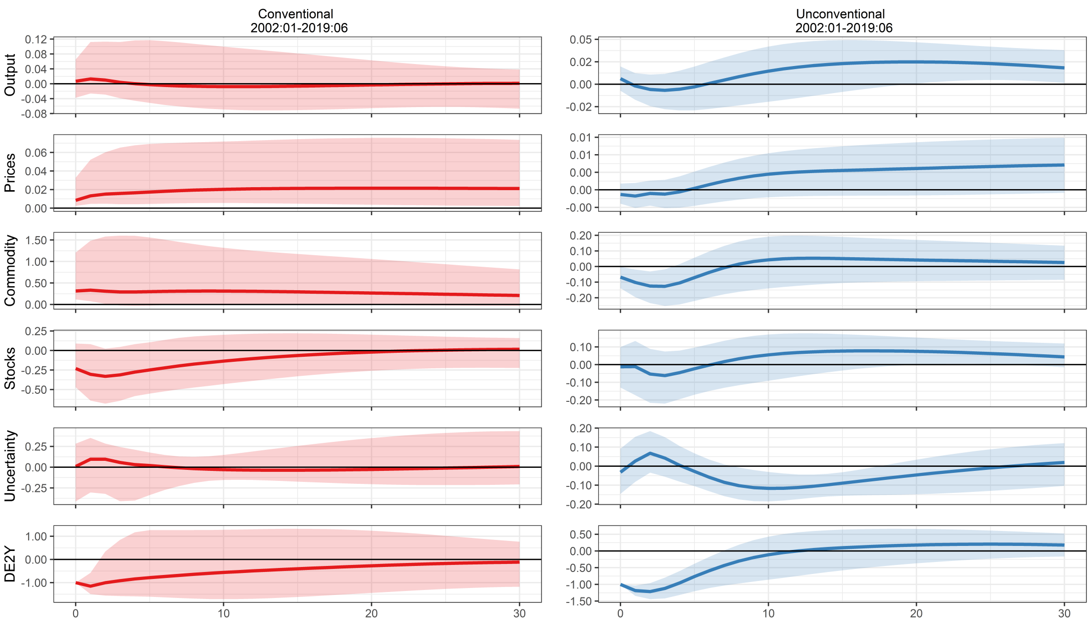
Therefore, it must be concluded at this point that the price increase observed for the overall surprises is exclusively due to the conventional measures and that the accumulated unconventional measures have no joint influence on the price level. We will discuss a possible explanation for this in subsection (1.6).

Figure 1.2: Effect of total expansive monetary-policy shock



Notes: The shaded areas show the upper and lower bands of the 90% of the confidence intervals. The intervals shown are robust for weak instruments (Montiel Olea et al., 2020).

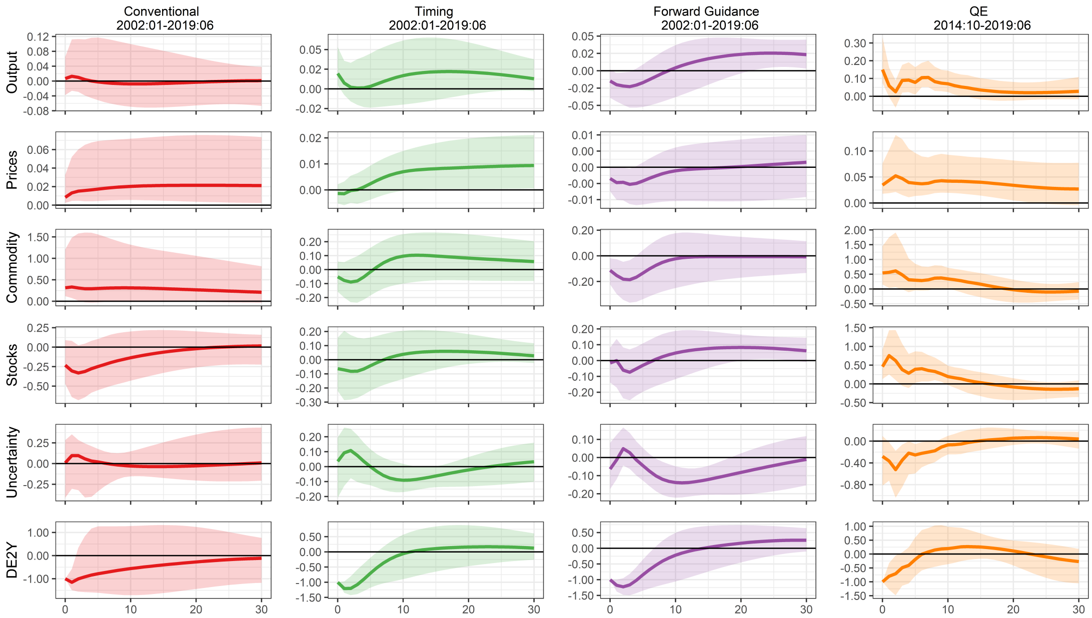
Figure 1.3: Effect of conventional and unconventional monetary-policy shock



Notes: The shaded areas show the upper and lower bands of the 90% of the confidence intervals. The intervals shown are robust for weak instruments (Montiel Olea et al., 2020).

However, the fact that unconventional measures do not affect inflation does not necessarily mean that all sub-measures also do not affect inflation. Figure 1.4 shows the result of splitting the unconventional measures into three individual surprises (timing, forward guidance, and QE).

Figure 1.4: Effect of monetary-policy shocks



Notes: The shaded areas show the upper and lower bands of the 90% of the confidence intervals. The intervals shown are robust for weak instruments (Montiel Olea et al., 2020).

Timing and forward guidance are relatively similar and have no significant impact on prices. The effect of QE is quite different compared to the other measures. A positive QE shock, such as the unexpected introduction of a bond-buying program, lowers uncertainty and almost instantly increases stock prices. The reason could be that QE has already been tested in the United States and the markets considered it a suitable reaction by the central bank. Thus, markets have experience with these kinds of measures. When it comes to the inflation response, the QE reaction differs entirely from those of the other two unconventional measures. While the latter are somewhat similar and found to have no significant impact on inflation, the effectiveness of QE moves at the level of conventional measures and is significantly different from both zero and the other unconventional shocks, at least in the first three months.[13]

[13] Note that a shorter data set had to be used for the QE analysis, so the results are not fully

1.5 The financial crisis

So far, we have considered the entire ECB period from 2002 to 2019 as a whole. However, the financial crisis of 2008-2009 led to significant changes in the economy. The ECB reached the ZLB with a strongly expansionary policy and implemented new measures. All these changes could indicate a potential structural break in our model. To take into account the risk of a structural break around 2008-2009, we split our data set into two samples. We estimate the model with the finest distinction for each shock from 2002:01 to 2009:05 (pre-crisis) and from 2009:06 to 2019:06 (post-crisis), which allows us to compare the impacts of different measures on macroeconomic developments before and after the financial crisis.

Table 1.3: F-Statistic of the regression of residuals on Z (Subsample)

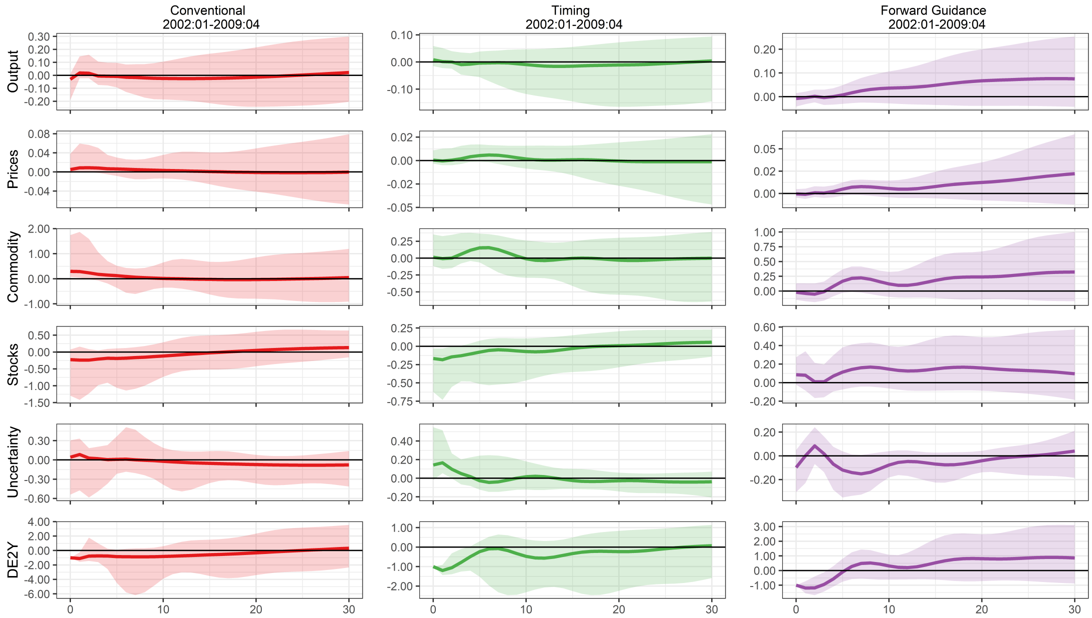
	Full		Pre-crisis		Post-crisis	
shock	F-robust	obs	F-robust	obs	F-robust	obs
Conventional	9.668	207	2.367	85	27.949	119
Timing	11.063	207	2.960	85	5.910	119
Forward-guidance	21.162	207	7.253	85	7.300	119
QE					10.992	54

The F-statistics of the shocks in the subsamples vary. Especially in the period before the crisis, there is less correlation between the conventional, timing, and forward guidance shocks and the DE2Y residuals. Unconventional and total shocks are significantly correlated, at levels above 10. In the subsample, timing and forward guidance shocks are also correlated, but at levels below 10, whereas the correlation with conventional shocks is significantly above 10. The smaller correlation can potentially be explained by the smaller sample size, which is due to its design. We use robust intervals to avoid the risk of a weak instrument bias Montiel Olea et al. (2020).

The results for the period before the financial crisis are not particularly meaningful. The small number of observations leads to large confidence intervals, making reliable statements about the effects difficult. However, the period that is more

comparable. Nevertheless, the data provides very interesting preliminary results.

Figure 1.5: Effect of monetary-policy shocks 2002:01-2009:04



Notes: The shaded areas show the upper and lower bands of the 90% of the confidence intervals. The intervals shown are robust for weak instruments (Montiel Olea et al., 2020).

important for this paper allows for more precise results. In Figure 1.6, it is first noticeable that the impulse responses do not differ significantly from the results in Figure 1.4. Expansive conventional and QE shocks increase prices, while timing and forward guidance shocks make no significant impact.

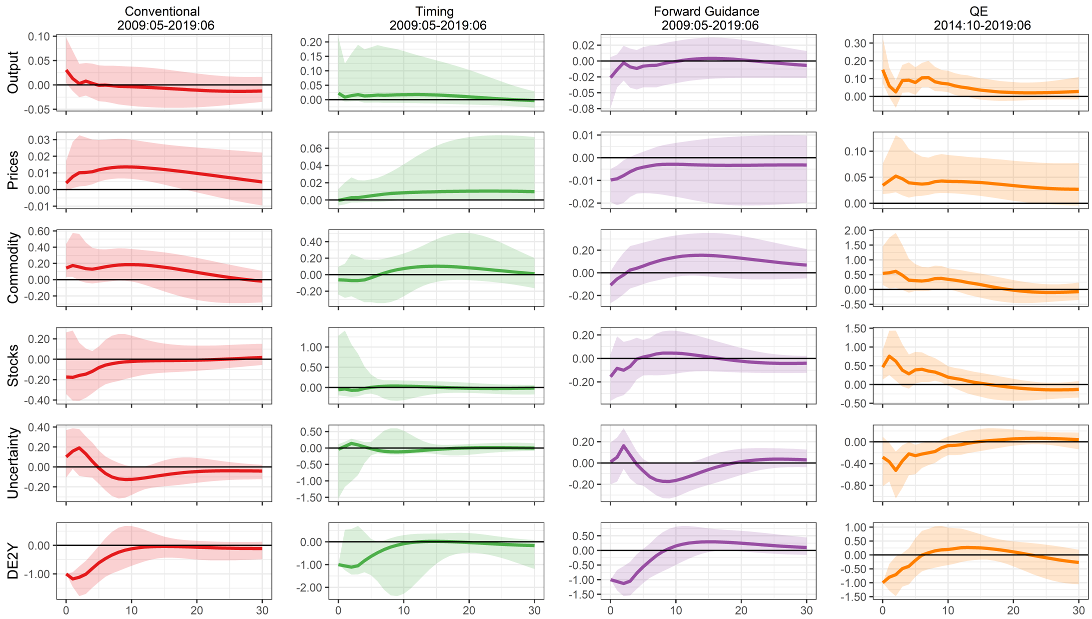
However, there are small interesting differences. On the one hand, the effect of conventional shocks is not as persistent as in the whole sample. On the other hand, for timing surprises, the effects on prices are much higher, but they are still not significantly different from zero. This result is surprising, since, especially after the financial crisis, particular emphasis has been placed on these forward-looking expectation-building measures.

All in all, this robustness test shows that our results from the total sample are not significantly altered by the structural break of the financial crisis. Even after the crisis, the ECB still could influence prices through its policies.

1.6 Delphic and Odyssean shocks

The question arises as to why timing and forward guidance shocks not affect inflation, whereas QE and the conventional shock demonstrate theory-conforming

Figure 1.6: Effect of monetary-policy shocks 2009:05-2019:06



Notes: The shaded areas show the upper and lower bands of the 90% of the confidence intervals. The intervals shown are robust for weak instruments (Montiel Olea et al., 2020).

behavior. One reason could be that the former are not correctly identified. It can be shown in high-frequency data that some central bank decisions cause an unusual reaction, where an expansive central bank shock lowers interest rates, but stock prices fall as well. This contradicts the results of Bernanke and Kuttner (2005), that falling interest rates lead to rising stock prices, and vice versa. A possible explanation for this result is provided by Campbell et al. (2012) and Andrade and Ferroni (2021). These authors develop a theory based on the idea that forward guidance shocks can have different effects, depending on how financial market participants interpret them. The first interpretation is an Odyssean forward guidance shock.

In this case, the central bank is completely credible and clear in its communication. Thus there is no reason for the markets to deviate from the signals coming from the central bank. The name Odyssean goes back to the Greek mythology where Odysseus bound himself to his ship facing the sirens. In an Odyssean forward guidance shock, the markets behave as the central bank expects. If the central bank communicates expansionary forward guidance, such as keeping the interest rate lower for longer, the markets react to it by investing more, for example, in stocks or other assets.

In contrast, Delphic forward guidance shocks work the other way around. If the central bank commits to keeping the interest rates lower for longer, the markets judge this as a signal that the economic situation is even worse than expected and sell assets. The term Delphic refers to the Oracle of Delphi making predictions that need to be interpreted by the individuals, and thus trigger actions.

So, Odyssean shocks could be expected to increase inflation. Simultaneously, the reverse is true for Delphic shocks.[14] However, a new study by Bauer and Swanson (2020) finds results that cast doubt on the theory's basic assumptions. A survey of US forecasters shows that they have never improved their forecast after a restrictive shock.

Although we cannot distinguish Delphic channel's origin, it seems reasonable to distinguish between these two kinds of forward guidance shocks. To do so, we use the "poor man's sign restrictions", which create very similar results compared to more complex procedures (Jarociński and Karadi, 2020). The idea is to determine from the markets' immediate reaction whether they interpret a shock as Delphic or Odyssean, according to the following:

$$(1.12) \quad C_{i,w} = \begin{cases} sgn(OIS2Yd_{i,w}) \neq sgn(STOXX50d_{i,w}) & \rightarrow \text{Odyssean event} \\ sgn(OIS2Yd_{i,w}) = sgn(STOXX50d_{i,w}) & \rightarrow \text{Delphic event} \end{cases}$$

$$w = \{\textit{press release, press conference}\}$$

For each monetary policy decision i, we compare the reaction in DE2Y and Euro Stoxx 50 around the high-frequency window w. If both reactions show the same sign, we label this event as Delphic, and Odyssean otherwise.[15] First, we look at the unconventional measures and therefore use $w =$ *press conference window*. This gives us four new factors: Odyssean timing surprises, Delphic timing sur-

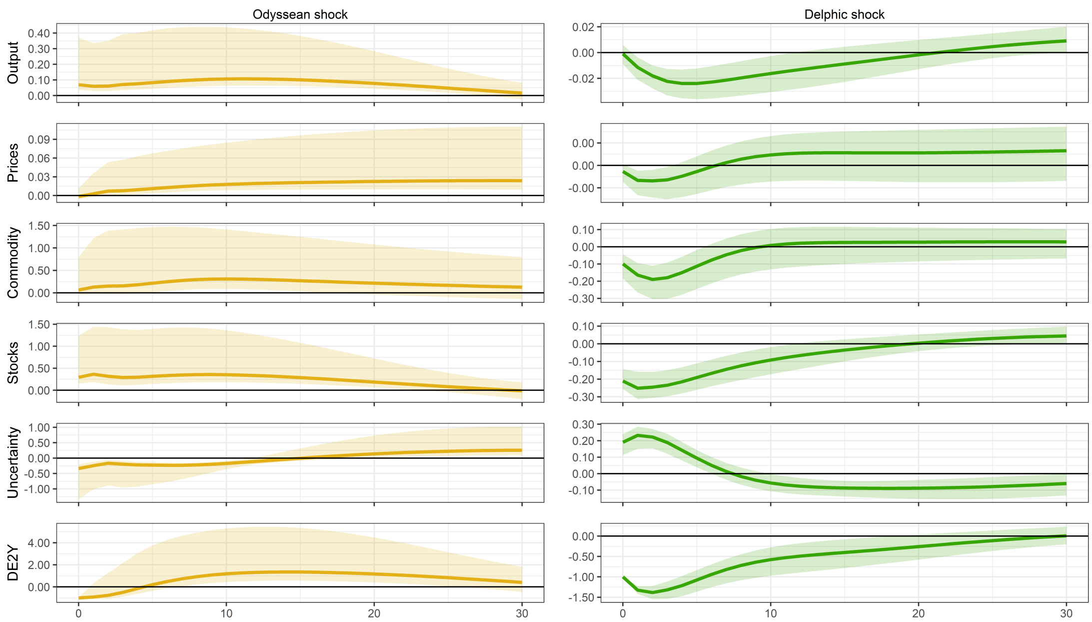
[14] Other terms in the literature for Delphic and Odyssean shocks are information and monetary shocks, respectively.

[15] We stick to this simple identification scheme based on Jarociński and Karadi (2020) and do not include inflation expectations, as Altavilla et al. (2019). This scheme has the advantage that each decision is uniquely assigned to either an Odyssean or Delphic shock. Additionally, we can use the EA-MPD, which excludes other effects due to the narrow time window around the decision.

prises, Odyssean forward guidance surprises, and Delphic forward guidance surprises.

With these four new surprises, we re-estimate our external instruments VAR with the same variables as our baseline model: *Output* (ECB industrial production), *Prices* (ECB Harmonised Index of Consumer Prices), *Commodities* (IMF Primary Commodity Price index), *Stock prices* (Euro Stoxx 50), *Uncertainty* (CISS), and *DE2Y*. The results are shown in Figures 1.7 and 1.8.[16] The timing surprises' impulse responses show a different course, depending on whether the shock is Delphic or Odyssean. Both Delphic and Odyssean forward guidance shocks lower bond yields. However, if the announcement is Delphic, this influences the markets negatively in various ways: uncertainty rises, stock prices collapse, and commodity prices decrease, possibly because of demand-side effects. This lowers output and has even a significant negative impact on inflation. An Odyssean timing shock shows exactly the opposite behavior. A price increase results, with a short time lag but roughly at the level of a conventional or QE shock.

Figure 1.7: Effect of Odyssean and Delphic timing shock

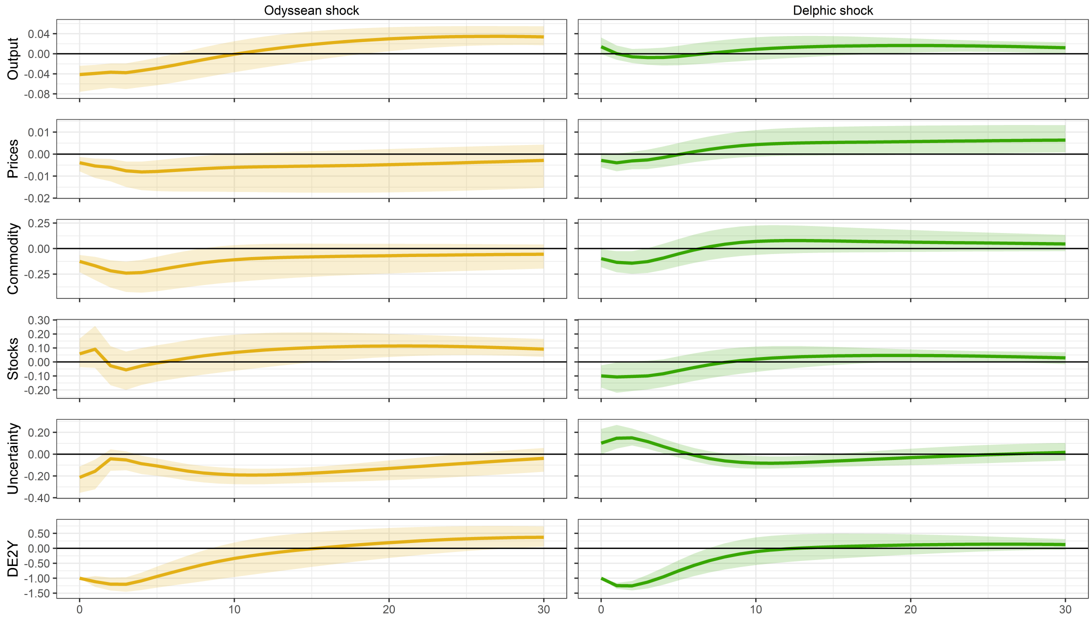


Notes: The shaded areas show the upper and lower bands of the 68% of the confidence intervals. The intervals shown are robust for weak instruments (Montiel Olea et al., 2020).

[16] For statistical reasons, we do not create subsamples here. The results from the previous chapter also show that the financial crisis does not significantly influence the results.

The reactions differ from the preceding ones in terms of the forward guidance shocks (Figure 1.8). Again, we can observe the different behaviors of Odyssean and Delphic shocks in uncertainty and stock prices. However, in contrast to Odyssean timing shocks, Odyssean forward guidance shocks do not lead to an increase in commodity prices. The output does not increase on impact, but only after some time. There is now a negative effect on prices. Longer-term expectation management by the central bank does not appear to have the desired effect on inflation.[17]

Figure 1.8: Effect of Odyssean and Delphic forward-guidance shock



Notes: The shaded areas show the upper and lower bands of the 68% of the confidence intervals. The intervals shown are robust for weak instruments (Montiel Olea et al., 2020).

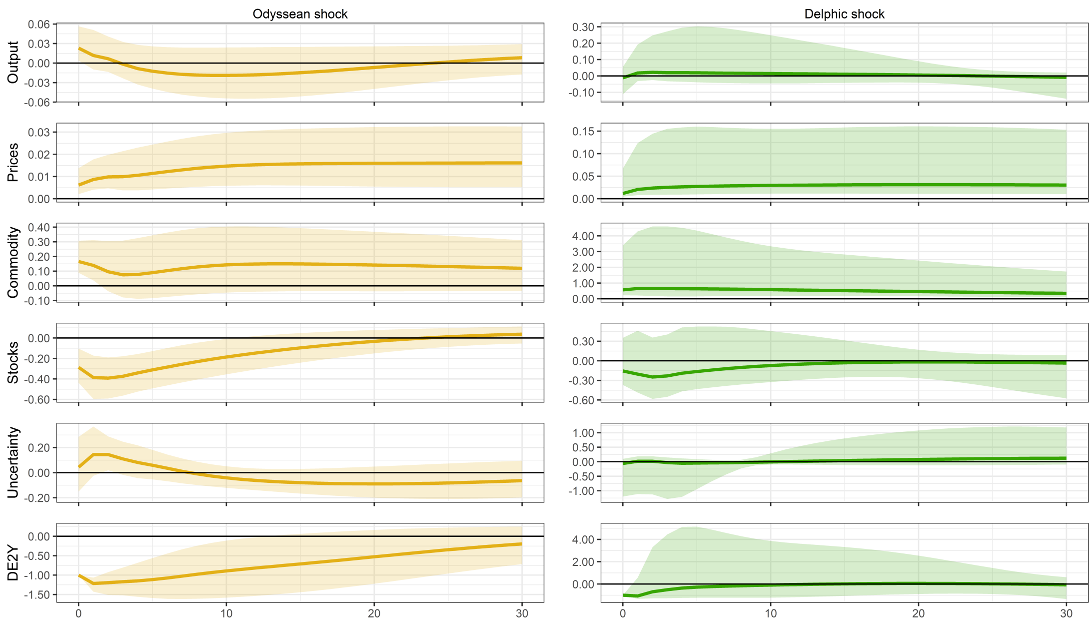
We conclude from this result and following the literature, that a more precise distinction between timing and forward guidance shocks is reasonable. It seems that the ECB can influence its primary target inflation more successfully if it influences short-term expectations. A prerequisite for this is, however, that the central bank can consciously send an Odyssean shock. Whether a central bank can influence its shocks as being viewed as Odyssean or Delphic has not yet been investigated, to the best of our knowledge. That topic would be a promising starting point for further research.

[17] This finding is in line with the study by McKay et al. (2016).

Our data set allows us to investigate another interesting point. So far, it is not clear in the literature where exactly a distinction between Odyssean and Delphic shocks can be useful. While Campbell et al. (2012) and Andrade and Ferroni (2021), by assumption, only refer to forward guidance shocks in their analysis of Delphic shocks, Jarociński and Karadi (2020) examine a monetary policy aggregate effect. Therefore, the question arises as to whether the central bank, by setting interest rates, also discloses information on its assessment of the economic situation. If this were the case, the conventional policy would also have a Delphic component.[18]

We therefore slightly adjust the above-mentioned poor man's sign restriction. To distinguish the conventional surprises, we now use changes in the high-frequency variables around the publication of the press release $w =$ *press release window*. We again apply the resulting new surprise series in our VAR framework. Figure 1.9 shows the resulting impulse response functions.

Figure 1.9: Effect of Odyssean and Delphic conventional shock



Notes: The shaded areas show the upper and lower bands of the 68% of the confidence intervals. The intervals shown are robust for weak instruments (Montiel Olea et al., 2020).

The effects on inflation and output differ slightly, but not in their sign. A Delphic

[18]The same would be conceivable for the QE components, but the subdivision of the data set makes a reliable estimate impossible for such a short time. Therefore, we postpone this analysis for future research.

shock has a slightly smaller effect on inflation. The classification scheme is not appropriate for conventional policy, which we see as an indication that Delphic shocks are indeed mainly reflected in forward guidance. In this respect, we can empirically support the assumption of Campbell et al. (2012) and Andrade and Ferroni (2021). A difference in timing shocks also exists in the euro area.

1.7 Conclusion

This paper distinguished the responses of conventional and unconventional monetary policy measures on macroeconomic variables using a high-frequency data set that measures the impact of the ECB's monetary policy decisions. Our framework allows us to estimate the various macroeconomic effects of central bank policies using a single methodology, facilitating policy comparisons. We show that unconventional and conventional monetary policy measures are somewhat similar in terms of their influence on uncertainty and output, but differ considerably concerning commodity prices and the ECB's primary target, the inflation rate. While conventional measures show the expected response of an increase in inflation following an expansionary monetary policy shock, unconventional measures appear to have no significant influence.

In detail, this result holds for timing and forward guidance shocks, but not for QE, which is found to have an influence on inflation equivalent to that of conventional interest rate changes. To explain this finding, timing shocks and forward guidance are divided into two parts. We show that there is indeed a difference for the short-term timing shock, depending on how the markets interpret the signal given by the ECB. Whereas Odyssean shocks exhibit the expected behavior in this case—that is, an expansionary shock tends to increase inflation—Delphic shocks show no effect or even a negative effect on inflation. Even worse, concerning medium-term forward guidance shocks, both Odyssean and Delphic shocks tend to decrease inflation if the ECB wants to send an expansionary signal. Furthermore, we can show that this classification by high-frequency variables for conventional shocks does not allow for a clear distinction. We conclude that the assumption that Delphic shocks are a forward guidance–specific phenomenon is justified and empirically verifiable.

What do these results mean for monetary policy? We would call for central banks, such the ECB, to conduct a conventional monetary policy for as long as possible, which the ECB did in large parts of the crisis period. The preferred measure among these is QE, because the "gentler" communication measures always carry the risk of a Delphic shock. It is unclear whether the central bank can precisely control the effect of its announcement and thus intentionally trigger an Odyssean shock. Only in this case would an expansionary shock indeed raise inflation. This result suggests that the use of communication measures as a whole cannot steer the markets in the way the ECB expects. Therefore, a safe option would be to focus on quantitative measures such as conventional policies and QE, since the risk of Delphic shocks is much lower in these cases.

2 Information shocks and Uncertainty.*

Martin Baumgärtner

Abstract

This paper studies the effect of uncertainty on conventional and unconventional monetary policy shocks in the euro area. The analysis shows that various forms of forward guidance under high uncertainty produce results similar to those attributed to information shocks in the literature. This effect can be seen in high-frequency variables and VAR models with external instruments. The results suggest that uncertainty is related to information shocks and can potentially explain the timing of these shocks.

Keywords: Uncertainty, high-frequency identification, structural VAR, ECB

JEL classification: E44, E52, E58, G14

*This essay was presented by me at the MAGKS Doctoral Colloquium and the 21st IWH-CIREQ-GW Macroeconometric Workshop: Forecasting and Uncertainty (Poster). The Paper is currently under review at the Journal of International Money and Finance.

2.1 Introduction

In reaction to the eurozone crisis, the ECB has adjusted its communication in the aftermath of the financial crisis. The main reason for this increased communication is that forward guidance is seen as an important policy tool to shape the expectations of market participants. However, this does not always seem to produce the desired effects. In high-frequency data around central bank announcements, we find a pattern that cannot be explained with basic economic theory. Contrary to theory, stock prices do not rise in the case of an expansionary announcement but fall. This phenomenon is known as *Delphic* or *information shock*. The literature agrees that this pattern exists and has important economic consequences (e.g. Jarociński and Karadi, 2020). Nevertheless, there is an ongoing debate about the reasons for this market reaction. A theory proposes that a central bank discloses private information together with its monetary policy decision (Campbell et al., 2012). Therefore, an expansive monetary surprise would suggest that the economic situation is worse than expected. Financial market participants revise their forecasts downwards because of this new information, and stock prices decrease. At the same time, Bauer and Swanson (2020) find that this forecast adjustment does not happen in reality. Leading US forecasters have never raised their forecasts in response to a restrictive central bank announcement.

This paper shows that the level of uncertainty surrounding a central bank decision significantly influences the effectiveness of monetary measures. I combine the findings that information shocks occur more frequently in crises[19] with studies showing that the effectiveness of monetary policy varies with the level of uncertainty. To account for this, I extend the analysis of Bauer and Swanson (2020) and Jarociński and Karadi (2020) with the level of uncertainty in which a monetary decision takes place. The resulting patterns are consistent with those in the information shock literature, namely that an expansionary shock can lower prices under certain circumstances. The findings suggest that uncertainty may play a role in explaining information shocks.

[19] See Altavilla et al. (2019).

I use the EA-MPD by Altavilla et al. (2019) to construct the ECB's monetary policy shocks and to measure the reaction of the financial market participants to the ECB decision. Euro-area data has the advantage that the ECB conference structure can be used to split expectation-forming policies. I consider short-term expectation-forming surprises, called timing, and longer-term forward guidance surprises.

If uncertainty is added in high-frequency estimations, timing and forward guidance shocks both show a positive correlation with stock prices, suggesting that restrictive timing and forward guidance shock can lead to rising stock prices if there is a high degree of uncertainty.

To confirm that this finding is more than a high-frequency phenomenon, I estimate a complementary VAR model with external instruments. I use a modified poor-man sign restriction approach, which identifies shocks depending on the uncertainty present before the announcement. Restrictive timing and forward guidance shocks under high uncertainty increase stock prices and increase output. Thus, despite a different identification strategy, the impulse responses contain both the identification pattern of Jarociński and Karadi (2020) and results comparable to the information shock literature.

The remainder of the paper proceeds as follows: Section (2.2) gives a literature review that includes an overview of the interaction between uncertainty and monetary policy. Section (2.3) describes the methodology, the construction of the monetary policy surprises, the regression model and the variables included. Section (2.4) presents the estimation results, showing that uncertainty is an important variable in the high-frequency context. In Section (2.5), I demonstrate that the observed effects are important in VAR models and influence monetary policy's economic effectiveness. To ensure that the observed patterns are due to uncertainty, I conducted a robustness check in Section (2.6), which shows that the results are robust to a decomposition of implied volatility. Finally, Section (2.7) concludes the paper.

2.2 Literature

The paper builds on the large strand of literature that deals with the issue how monetary policy shocks can be calculated from the reaction of financial markets directly before and after a central bank decision. For example, Gürkaynak et al. (2005), Brand et al. (2010), Swanson (2021), Andrade and Ferroni (2021), Cieslak and Schrimpf (2019), and Altavilla et al. (2019) use shocks identified by the change in high-frequency variables around central bank events.

The paper complements research that aims to explain the unusual reaction of the financial markets in the context of monetary policy shocks. Bernanke and Kuttner (2005) find that a restrictive monetary policy shock leads to a lower present value of stocks. However, the reverse is true for some events. A possible explanation for this is given by Campbell et al. (2012) and Nakamura and Steinsson (2018). The underlying idea is that the central bank provides market participants with information on future economic developments at the same time as a monetary policy shock. A restrictive monetary policy could thus convey information about a robust economic situation. This additional "information shock" leads market participants to adjust their forecasts upwards (Campbell et al., 2012). These improved market prospects may lead to a reaction of the high-frequency variables that do not correspond to theory at first sight.

The basis for this is a belief that central banks are in a superior position to predict economic developments than other market participants. Romer and Romer (2000) conclude that this superiority is not due to private information but because central banks invest considerably more in forecasting efforts. Rossi and Sekhposyan (2016) find similar results for the inflation forecast, but at the same time discover that the central bank and other market participants systematically over- or underestimate the forecasts in some periods. However, Faust et al. (2004), D'Agostino and Whelan (2008), and Hoesch et al. (2020) find evidence that the superiority of the central bank forecasts does not exist today, or at least does not occur in all periods.

The knowledge about information shocks has frequently been used in the literature to refine the identification of monetary policy shocks in VAR models.

Miranda-Agrippino and Ricco (2021) use an instrument that is robust to information shocks, while Jarociński and Karadi (2020), Andrade and Ferroni (2021), Kerssenfischer (2019), and Baumgärtner and Klose (2021) separate the shocks by sign restrictions on high frequency variables. All studies find large differences between the shocks affecting output and inflation. Although information shocks have substantially different effects on the economy, theoretical models suggest they are unsuitable for concrete policy instruments. Fujiwara and Waki (2019) show in a Dynamic Stochastic General Equilibrium (DSGE) model that these deliberately induced shocks would be associated with a high degree of uncertainty, which would contradict the central bank's intentions.

However, a study by Bauer and Swanson (2020) casts doubts on the theory of information shocks. The authors demonstrate that restrictive monetary shocks reduce stock prices around FED decisions. This pattern also applies to periods that provide the most substantial evidence of information shocks. The authors conclude that information effects are weak if they exist at all. Besides, the authors have surveyed major United States blue-chip forecasters, with the result that they have never raised their forecasts in the past due to a restrictive shock. They are thus fundamentally contradicting the theory of information shocks.

The second strand of the literature, which is relevant for this paper, focuses on how uncertainty can be measured and the interaction between uncertainty and monetary policy. Bloom (2009) shows that higher implied stock market volatility lowers output and employment. This measure is adopted by Bekaert, Hoerova, and Lo Duca (2013). They argue that the volatility indices contain a risk component in addition to the uncertainty component. By decomposing the index, they can separate the two components and express uncertainty in the sense of Knight (1921). Using the decomposition, they show that expansionary monetary policy can lower uncertainty.

Complementary to financial market indices, there are approaches to approximate uncertainty through text analysis. Baker et al. (2016) use various keywords in newspaper articles to construct uncertainty indices. Azqueta-Gavaldon et al. (2020) extend this approach with unsupervised machine learning procedures.

In addition to the literature on the effects of an uncertainty shock and the effects of monetary policy on uncertainty, studies examine the effectiveness of monetary policy changes with varying uncertainty. Aastveit et al. (2013) establish a theoretical model which, based on Dixit and Pindyck (1994), explains the role of uncertainty in the effectiveness of monetary policy and shows that monetary policy might be less effective under high uncertainty. A DSGE model by Castelnuovo and Pellegrino (2018) comes to similar results. Eickmeier et al. (2016), Pellegrino (2018), and Pellegrino (2020) can provide empirical evidence for the theoretical considerations through VAR models. In an uncertain market situation, monetary policy becomes less effective.

2.3 Methodology

To analyze the effect of monetary policy on interest rates and stock prices, I run a regression with high-frequency data based on the approach of Bernanke and Kuttner (2005).

$$\Delta x_t = \alpha + \beta_1 mp_t + \epsilon_t \tag{2.1}$$

where x is any high-frequency variable, like interest rates or stock prices, and mp_t are monetary policy surprises. I will show in Section (2.3.1) how the monetary policy surprises mp_t for Equation (2.1) can be calculated. Subsequently, Section (2.3.2) describes how this equation is extended with uncertainty.

2.3.1 Monetary policy surprises

Following the literature, I use the EA-MPD and framework by Altavilla et al. (2019) to calculate monetary policy surprises.[20] The database contains high-frequency deviations of financial variables around ECB press releases and press conferences. These two time windows are a unique feature of the euro area. Unlike the Fed, a central bank decision is communicated in two parts. First, the ECB publishes a brief statement at 1:00 p.m. outlining the relevant interest rate decision. This decision is explained in more detail in the second part, the

[20] The database currently covers the observation period from 2002:01 to 2019:12, which is, therefore, my observation period.

press conference at 2:00 p.m.. In addition, the president announces the relevant unconventional measures. The different windows allow dividing the central bank surprises more precisely.
Based on Gürkaynak et al. (2005), Brand et al. (2010), and Swanson (2021), I use factor analysis with imposed restrictions:

$$X^w = F^w \Lambda^w + \epsilon^w$$

$$\text{with w in } \{\textit{press release, press conference}\}$$

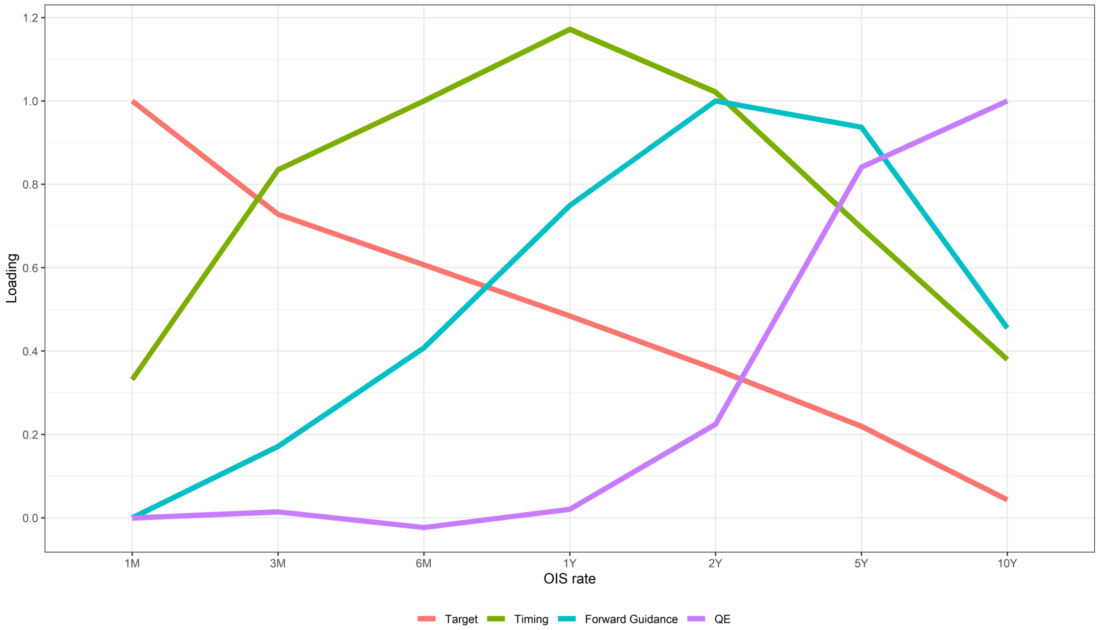
where X^w is the change of seven OIS with maturities from 1 month to 10 years, F^w is a $(N \times T)$ matrix of latent factors, Λ are the factor loadings, and ϵ^w is the idiosyncratic variation. I can estimate the latent factors F^w by using principal components on X^w. The matrix rank test of Cragg and Donald (1997) finds one statistically significant factor for the press release window. There are two significant factors in the conference window before the financial crisis and three significant factors for the entire sample. Therefore, for *w=press release*, I use one principal component, and for *w=press conference* I use three principal components.
The factors F^w cannot be interpreted directly as each factor is usually correlated with all OIS futures:

$$X^w = \tilde{F}^w \tilde{\Lambda} + \epsilon^w$$

with $\tilde{F}^w = FU$, $\tilde{\Lambda} = U'\Lambda$ and where U is a 3×3 orthogonal matrix. Any combination of U that satisfies $UU' = I$ solves the equation.

By rotating F^w, I get interpretable factors. The restrictions for the rotation are based on Gürkaynak et al. (2005), Brand et al. (2010), and Swanson (2021): (1) The second and third factor in each window should not be correlated with the one month OIS rate and (2) the third factor should have the smallest variance in the pre-crisis period. The factor names are determined by the influence of the individual factors on corresponding interest rates and following Altavilla et al. (2019): Target, timing, forward guidance and QE. Figure 2.1 presents the resulting factor loadings of the rotated factors.

Figure 2.1: Factor loadings



Notes: The figure shows the resulting factor loadings in basis points after rotation. All factors are scaled so that each has a unit effect on the corresponding OIS rate. Based on Altavilla et al. (2019).

The press release factor loads heavily on the very short-term OIS rates. Since the central bank's conventional interest rate policy affects this end of the scale, the factor is named *target factor*.

The loadings of the first factor in the press conference window show that it mainly influences the OIS rates under one year. At the same time, the effect on the one-month OIS rates is negligible. Accordingly, this factor does not coincide with the target factor from the press release window but is forward guidance with a short horizon. It is aimed at the following ECB decisions and is therefore called the *timing factor*.

The second resulting factor loadings show that this factor mainly affects the medium term, i.e. two to five-year interest rates. This factor is referred to as *forward guidance (FG) factor*.

The third factor loads on the long end of the yield curve, consistent with QE's expected effects. Therefore, it is called *QE factor*.

As the press conference window factors are orthogonal to each other by construction, the combination of the three press conference factors can be summed up.

The resulting factor is used to evaluate all measures that take effect during the press conference and do not affect short-term OIS rates. It is comparable to the *path factor* used in several US studies and is therefore called accordingly (e.g. Gürkaynak et al., 2005; Campbell et al., 2012; Bauer and Swanson, 2020).
All surprises are scaled that an increase of the respective value corresponds to an OIS rate increase. Therefore, positive surprises correspond to monetary tightening.

2.3.2 High-frequency estimation

To study the effect of monetary policy shocks on high-frequency variables, an essential point in estimating equation (2.1) is that the surprises affect different time windows. The target factor describes the reaction of the OIS rates around the release of the ECB decision. Timing, forward guidance and QE surprises result from OIS rates' reactions around the press conference. To distinguish the effects, I consider both time windows separately. This leads to the following periods: $\Delta x_{release}$ for the difference in the variable x before and after the ECB decision release, $\Delta x_{conference}$ for the difference in x before and after the press conference.

This results in the following equations:

$$\Delta x_{release,t} = \alpha + \beta_2 target_t + \epsilon_t \tag{2.2}$$

$$\Delta x_{conference,t} = \alpha + \beta_2 path_t + \epsilon_t \tag{2.3}$$

$$\Delta x_{conference,t} = \alpha + \beta_3 timing_t + \beta_4 FG_t + \beta_5 QE_t + \epsilon_t \tag{2.4}$$

where t indexes ECB announcements, $target_t$, $path_t$, $timing_t$, FG_t and QE_t describe the monetary policy surprises for event t and $x_{w,t}$ describes the change of the high frequency variables in the corresponding time window w at event t. Here, equations (2.2) and (2.3) are similar to those from Bauer and Swanson (2020). However, equation (2.4) allows a finer distinction between the various unconventional monetary measures.
In the estimation, I will focus on overnight interest rate swaps and stock prices,

as these are used in the literature to identify information shocks in VAR models (e.g. Altavilla et al., 2019; Jarociński and Karadi, 2020). The OIS rates are directly linked to central bank policy through overnight interest rates. Therefore, if the central bank has a credible policy, then the change of OIS rates should reflect only the monetary shock, regardless of the economic situation, since the central bank has committed itself to maintain this course. A restrictive central bank policy is expected to increase OIS rates, raising expectations for main refinancing rates in the future. In contrast, stock prices are not exclusively linked to monetary policy but serve as a benchmark for financial market participants' expectations. Monetary theory suggests that a restrictive monetary policy leads to falling stock prices. As a proxy for stock prices, I use the STOXX50 index ($STOXX$).

The next step is to extend equations (2.2)-(2.4) with uncertainty and euro specific control variables. Overall, four variables are added: uncertainty at the time of the announcement, a dummy whether the central bank publishes a forecast or not, the publication of the US Initial Jobless Claims, and a proxy for the current state of the economy.

Uncertainty is an essential factor for monetary policy because it affects investments. In phases of high uncertainty, investment stimulating measures, such as monetary policy, have a smaller effect on investor decisions as there are more risks for investors (Aastveit et al., 2013). At the same time, there is evidence for a link to information shocks. Altavilla et al. (2019) find that information shocks occurred particularly often in times of crisis, which are generally associated with high uncertainty.

I use implied volatility as a measure of uncertainty at the time of the decision following Bloom (2009). The VSTOXX index measures the implied volatility of option prices starting from one month up to two years. Its daily availability allows to measure uncertainty immediately before the announcement, rule out endogeneity problems, and minimize the risk of other events distorting the result. I achieve this by using the VSTOXX index closing price on the previous day of the decision, labelled as $VSTOXX_{pd,t}$. To account for the influence of uncertainty on the effectiveness of monetary policy decisions, I include the variable into the equations (2.2)-(2.4). By adding the respective interaction terms, I control for the varying effectiveness of individual monetary policy measures with varying degrees of uncertainty.

As the first control variable, I measure the impact of the ECB's publication of forecasts. These are published by the ECB every quarter and appear simultaneously with the press release. At these times, it should be easier for financial participants to get an intuition about the central bank's private forecasts. Thus, if there are information effects, the central bank decisions that coincide with the publication of projections are of particular interest. Therefore, I use a dummy to measure possible differences between the decisions. Like uncertainty, I include interaction terms for the individual policy measures to allow for separately possible effects on each variable. The projections are published after the press conference. However, the important results will already be announced at the press conference. Therefore, the dummy is only used in the regressions (2.3)-(2.4) and not in (2.2).

Since US Initial Jobless Claims are released every week at 14:30 CEST, they are a variable that could affect eurozone stock prices.[21] An increase in US unemployment could be a signal that the economic situation is deteriorating. Therefore, this would be an explanation for a possible negative reaction to stock prices. US jobless claims are seasonally adjusted and used in logarithm. Likewise, the jobless claims are only used in the equations (2.3)-(2.4) and not in (2.2), because the US jobless claims are published after the ECB's press releases.

A last potentially omitted variable is the current economic situation. It is conceivable that financial market participants react differently to central bank decisions, depending on the business cycle. Like uncertainty, I use the STOXX50 closing price on the previous evening of the decision as an additional control variable to monitor the real economic situation. The variable is abbreviated with $STOXX_{pd,t}$.

[21] Brand et al. (2010) and Altavilla et al. (2019) review the effect of US Initial Jobless Claims on OIS rates and find no significant impact.

2.4 Results

The results of the regressions are shown below in Table 2.1. I start with the simple estimates of equation 2.2-2.4 before turning to the results with control variables.

Table 2.1: Regression of $\Delta OIS/\Delta STOXX$ on monetary policy surprises

	release		conference			
	ΔOIS_{2year}	ΔSTOXX	ΔOIS_{2year}	ΔSTOXX	ΔOIS_{2year}	ΔSTOXX
Target	37.21^{***}	-3.72^{**}				
	(9.36)	(1.65)				
Path			92.72^{***}	-2.22^{*}		
			(3.57)	(1.25)		
Timing					100.57^{***}	0.45
					(2.22)	(1.83)
FG					99.90^{***}	-2.01
					(1.54)	(1.27)
QE					21.79^{***}	-11.35^{**}
					(1.58)	(5.48)
Adj. R^2	0.31	0.06	0.94	0.02	0.99	0.05
Num. obs.	195	195	190	190	190	190
F statistic	89.72	12.30	3177.41	5.29	5085.21	4.52

Note:$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$. The table shows the relationship between high-frequency variables and monetary surprises over a period from 2002-01-01 to 2019-12-31. The dependent variable is the change in either two-year OIS rates or STOXX50 stock prices within the two ECB announcement windows. Target, timing, forward guidance, and QE describe the level of the respective monetary surprise. For better readability, the LHS variable is multiplied by 100. The test statistics are calculated with heteroscedasticity and autocorrelation robust (HAC) standard errors.

The results for the baseline estimation confirm the theoretical considerations. A restrictive monetary policy raises OIS rates and vice versa. Since target surprises influence the short end of the yield curve, the effect on the two-year OIS rate is somewhat weaker than for the unconventional surprises.[22] At the same time, restrictive monetary policy lowers stock prices in the release window.

A similar pattern is present in the conference window: Interest rates rise, and stock prices fall due to a restrictive shock. This is similar to the situation that Bauer and Swanson (2020) report for the US. However, if we look at the individual surprises, there are differences: Although the effects of all variables on

[22]For reasons of comparability between the measures, only the effect on two-year OIS rate is given here. The results for shorter maturities than two years are similar and available upon request.

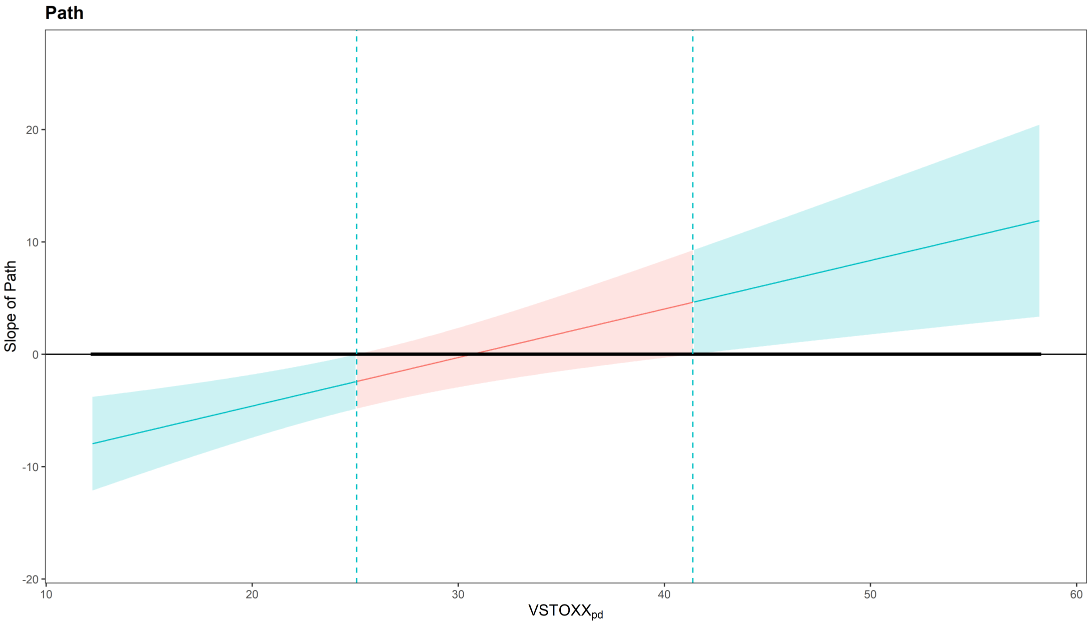
OIS rates are positive, the response of stock prices differs. The QE surprises show a negative correlation, but timing and forward guidance surprises are not significantly different from zero.

Therefore, the results of Bauer and Swanson (2020) need to be considered more differentiated, at least for the euro area: For communicative measures, in contrast to target and QE measures, evidence of behaviour can be found, which cannot be explained by pure monetary shocks. In the following, I will show that uncertainty can explain this pattern. The extended model results are illustrated in Table 2.2. Again, we expect effective restrictive monetary policy to cause higher OIS rates and lower stock prices. The reaction of the OIS rates in the release window is similar to the previous results. However, due to interaction terms, the target shock does not reflect the full monetary shock effect. A target shock's total effect is composed of two coefficients, one of which is dependent on another variable, such as $VSTOXX_{pd,t}$. Overall, a restrictive target shock increases OIS rates, similar to Table 2.1.

In the release window, the level of uncertainty determines the effect of stock prices. The higher the level of uncertainty at the time of publication, the more negative the stock prices' reaction during the release window. In this context, the effect of a target shock is not clear-cut. Neither the base effect nor the interaction term is significantly different from zero.

If we look at the conference window, we see that unconventional monetary policy's base effect is still in line with expected behaviour. Without uncertainty, unconventional monetary policy increases overnight interest rates and lowers stock prices. However, the interaction terms with uncertainty show that this is not fully persistent. The interaction effect between the path factor and uncertainty is significantly positive. If uncertainty increases, the total coefficient becomes less negative. Figure 2.2 illustrates the parameter's change with increasing uncertainty. The confidence intervals presented describe the 95% Johnson-Neyman intervals according to Johnson and Fay (1950) and Bauer and Curran (2005). They indicate at which uncertainty values the parameters deviate significantly from zero. Not only does monetary policy become less effective, but it also shows a significant positive reaction of stock prices with expansionary measures in periods of high uncertainty. The critical value for a reversal of the coefficient is around 30.

Figure 2.2: Change in the effect of unconventional monetary policy with varying uncertainty



Notes: The plotted line is calculated using the estimated coefficients of Equation (2.3) and control variables: $\Delta STOXX_{conference,t} = mp_{surprise} + mp_{surprise} * VSTOXX_{pd,t}$. The confidence interval describes the "Johnson-Neyman" interval at the 95% significance level. It indicates at which uncertainty values the parameters deviate significantly from zero. The distribution of uncertainty is marked by the thick black line.

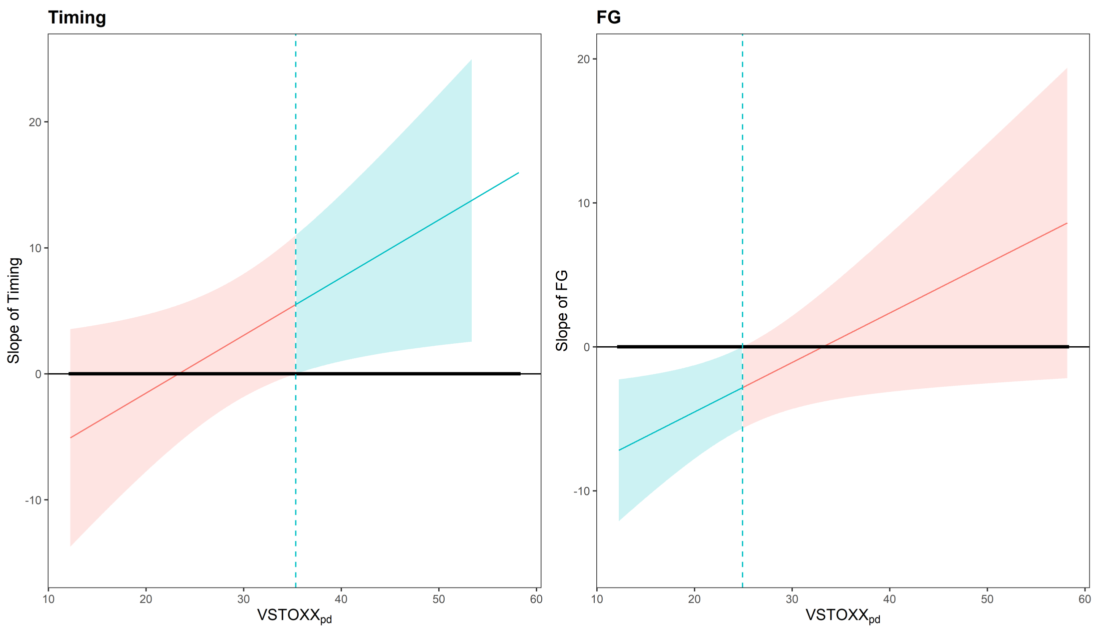
Also, there is a significantly positive effect in connection with the previous day's stock prices. Stock prices tend to rise during the ECB conference when the economic outlook is positive.

The picture is similar when looking at the individual measures in the conference window. Without the influence of uncertainty, all restrictive measures increase OIS rates. Interestingly, the effect of a timing shock on OIS rates is dependent on whether the central bank discloses its forecasts. However, the effects are not particularly large compared to the baseline effects of timing. The overall effect of uncertainty in the previously observed data would never be high enough that the sign for OIS rates would turn positive. The OIS results are an indication that financial markets consider the ECB credible to implement the policies it announces consistently.

In the case of stock prices, it is apparent that uncertainty plays a significant role. The coefficients of timing and forward guidance are, without uncertainty, significantly negative. With increasing uncertainty, the coefficients move into the positive range. Besides, the stock prices level before the ECB announcement, i.e. the state of the economy, positively influences the development of stock prices in the considered press conference window.

The publication of the ECB projections does not appear to have a systematic impact on stock prices. The projections are the most obvious point where the central bank discloses its information and forecasts. On these dates, the central bank should be the most reliable to read. Nevertheless, on these dates, the regression does not show any effect. Neither the individual coefficient nor the interaction terms with monetary policy surprises are significantly different from zero. Thus, there is no indication that forecast releases provide essential information to the stock markets. This result shows that it is not necessarily information that drives the pattern of so-called information effects.

Figure 2.3: Change in the effect of timing and forward guidance with varying uncertainty



Notes: The plotted line is calculated using the estimated coefficients of Equation (2.4) and control variables: $\Delta STOXX_{conference,t} = mp_{surprise} + mp_{surprise} * VSTOXX_{pd,t}$. The confidence interval describes the "Johnson-Neyman" interval at the 95% significance level. It indicates at which uncertainty values the parameters deviate significantly from zero. The distribution of uncertainty is shown by the thin black line.

Figure 2.3 shows the heterogeneity between timing and forward guidance surprises. While the timing coefficient is significantly negative with very low uncertainty, the values turn positive for high uncertainty. It can be seen that uncertainty makes timing measures less effective, but the coefficient becomes significantly positive when uncertainty is high.

A similar pattern is apparent for forward guidance. There is a significant negative effect at low uncertainty, which gets insignificant with increasing uncertainty. In principle, a restrictive monetary policy has a dampening effect on stock prices. However, as soon as there is a high uncertainty level, this is no longer the case. The total effects become positive above a certain level of uncertainty in an economy, with a volatility index value of 26.2 for timing and 31.2 for forward guidance. In such cases, restrictive monetary policy can increase stock prices, consistent with the information shock literature's observations.

Therefore, uncertainty could be a possible explanation for the pattern observed by Bauer and Swanson (2020), who claim that financial market participants price in past market events with a delay at the time of the central bank decision. Similar to the incentive for firms to postpone investment decisions for some time when uncertainty is high, it may be reasonable for financial market participants without useful benchmarks to wait for the central bank's reaction and postpone the pricing of bad news until the central banks' decision. If the central bank responds to a crisis with conventional measures such as interest rate policy, this calms the markets. The negative effect of economic news is (over)compensated by the positive monetary policy shock.[23] This results in a textbook reaction as long as the monetary policy shock is big enough. However, suppose the central bank opts for softer measures. In that case, the central bank will not dampen the market's uncertainty because timing and forward guidance become less effective while uncertainty increases. The markets also evaluate the previous economic news and monetary policy together, but the monetary policy does not balance the economic news. Accordingly, the stock price reacts unusually.

[23] In order to consider the effect of the effective lower bond in the analysis, I have estimated this analysis with data up to 2011, when the the effective lower bond did not yet apply in the euro area. The effects are similar and available on request.

Table 2.2: Regression of $\Delta OIS/\Delta STOXX$ on monetary policy surprises including controls

	release		conference			
	ΔOIS_{2year}	$\Delta STOXX$	ΔOIS_{2year}	$\Delta STOXX$	ΔOIS_{2year}	$\Delta STOXX$
Target	65.84^{***}	5.24				
	(21.09)	(7.22)				
Path			80.69^{***}	-13.25^{***}		
			(9.13)	(3.51)		
Timing					106.11^{***}	-10.70^{**}
					(6.17)	(4.76)
FG					102.71^{***}	-11.40^{**}
					(7.45)	(4.40)
QE					17.88^{***}	-4.23
					(6.75)	(14.96)
$\text{VSTOXX}_{\text{pd}}$	-0.04	-1.05^{***}	1.04	0.11	0.56	-0.11
	(1.27)	(0.30)	(1.03)	(0.53)	(0.46)	(0.56)
$\text{VSTOXX}_{\text{pd}} * \text{Target}$	-0.75	-0.22				
	(0.60)	(0.17)				
$\text{VSTOXX}_{\text{pd}} * \text{Path}$			0.55^{*}	0.43^{***}		
			(0.30)	(0.13)		
$\text{VSTOXX}_{\text{pd}} * \text{Timing}$					-0.10	0.46^{***}
					(0.18)	(0.16)
$\text{VSTOXX}_{\text{pd}} * \text{FG}$					-0.14	0.34^{*}
					(0.24)	(0.18)
$\text{VSTOXX}_{\text{pd}} * \text{QE}$					0.21	-0.40
					(0.28)	(0.78)
STOXX_{pd}	0.01	-0.00	-0.02^{*}	0.02^{*}	-0.02^{**}	0.02^{*}
	(0.01)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)
US jobless claims			-53.27	-4.51	-38.29^{**}	4.11
			(47.54)	(22.39)	(15.68)	(23.26)
Projection Dummys	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R^2	0.32	0.14	0.95	0.07	0.99	0.08
Num. obs.	195	195	190	190	190	190
F statistic	23.68	8.97	482.10	2.92	1262.90	2.18

Note: $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$. The table shows the relationship between high-frequency variables and monetary surprises over a period from 2002-01-01 to 2019-12-31. The dependent variable is the change in either two-year OIS rates or STOXX50 stock prices within the two ECB announcement windows. Target, Timing, Forward Guidance, and QE describe the level of the respective monetary surprise. $VSTOXX_{pd}$ describes the closing price of the index of implied volatility on the day before the decision. All estimations control the stock price at the time of the decision, the release of the U.S. unemployment data, and the ECB projections' release. For better readability, the LHS variable is multiplied by 100. The test statistics are calculated with heteroscedasticity and autocorrelation robust (HAC) standard errors.

2.5 VAR approach

Uncertainty at the ECB decision may explain why financial market participants react atypically and why a restrictive monetary policy leads to rising stock prices. Therefore, the question arises whether this finding in the high-frequency data is a short-term anomaly or whether this behaviour influences longer-term macroeconomic variables. To investigate this, I estimate a complementary VAR model to show the influence of uncertainty on monetary measures. The instruments will be the policy surprises identified above.

I modify the identification by poor-man sign restriction from Jarociński and Karadi (2020). The authors use a *VAR model with external instruments and poor-man sign restriction* to identify monetary policy and divide a shock series into either monetary shocks or information shocks based on high-frequency variations.[24]

I adapt this procedure and use the constructed surprises above: For timing and forward guidance surprises, I divide the time series into two sub-series: one with surprises at low uncertainty and one at high uncertainty. Therefore, it is possible to compare the effectiveness of timing and forward guidance surprises at different levels of uncertainty. Again, the proxy for uncertainty is the VSTOXX index closing price on the day before the ECB decision.

2.5.1 Econometric Model

The econometric model used here is based on the work of Stock and Watson (2012), Mertens and Ravn (2013), and Gertler and Karadi (2015). It is identical with the model in Baumgärtner and Klose (2021). Let Y_t be a $(N \times 1)$ vector of N economic variables. Consider a VAR in general structural form:

$$AY_t = C + \sum_{j=1}^{J} C_j Y_{t-j} + \epsilon_t \tag{2.5}$$

where C represents a vector of constants, while A and C_j form the coefficient matrices, including J lags. Premultiplying both sides with the inverse of A leads

[24] In addition to this, they identify the shocks by a high-frequency approach, that leads to the same results.

to

$$Y_t = SC + \sum_{j=1}^{J} SC_j Y_{t-j} + v_t \tag{2.6}$$

with v_t denoting the reduced-form residuals and $S = A^{-1}$. They are connected to the structural shocks ϵ_t by

$$v_t = S\epsilon_t \tag{2.7}$$

because they are a linear combination of structural shocks. Inserting (2.7) in (2.6) gives:

$$Y_t = SC + \sum_{j=1}^{J} SC_j Y_{t-j} + S\epsilon_t \tag{2.8}$$

I am especially interested in estimating one column of S. The column s^p indicates how the reduced form residuals v_t changes in response to a unit increase in the structural shock ϵ_t^p. I follow Gertler and Karadi (2015) and focus the analysis on column $s^{mp} = S_{.,mp}$, which reflects the reaction of our variables to a monetary policy shock. All other columns are represented by $s^q = S_{.,q}$. Together with (2.7) it follows:

$$v_t^{mp} = s^{mp}\epsilon_t^{mp} \tag{2.9a}$$

$$v_t^{q} = s^{q}\epsilon_t^{mp} \tag{2.9b}$$

These can be solved for v_t^q by

$$v_t^q = \frac{s^q}{s^{mp}} * v_t^{mp}. \tag{2.10}$$

The fraction represents a unit effect normalization. A unit shock in ϵ_t^{mp} increases v_t^{mp} by the very same amount. All other effects on variables are expressed in proportion. If we seek to solve this equation, we face an endogeneity problem. To resolve this, I use a two-step approach with an instrument Z. According to Stock and Watson (2018), a good instrument requires the following characteristics to obtain consistent estimates:

$$E[\epsilon_t^{mp} Z'] = \alpha \neq 0 \qquad \text{(relevance)} \tag{2.11a}$$

$$E[\epsilon_t^{q} Z'] = 0 \qquad \text{(exogeneity w.r.t. other current shocks)} \tag{2.11b}$$

Therefore, an instrument is needed which is highly correlated with the monetary policy shock ϵ_t^{mp} but is uncorrelated with any other shock ϵ_t^q. With a feasible instrument and the reduced form variance-covariance matrix Σ, I get a consistent estimate of s by using a two-stage approach. In the first stage I regress v_t^{mp} on Z to estimate the fitted value $\hat{v}_t^{mp}$. The result is the part of the variation in v_t^{mp} which relies on the structural shock ϵ_t^{mp}. Inserting this in (2.10) gives

$$v_t^q = \frac{s^q}{s^{mp}} * \hat{v}_t^{mp} + \xi_t. \tag{2.12}$$

The second stage regression (2.12) leads to a consistent estimation of $\frac{s^q}{s^{mp}}$. With Σ I can determine all components of s^{mp}, which in turn allows me to estimate impulse responses from our partially identified structural VAR model:[25]

$$Y_t = SC + \sum_{j=1}^{J} SC_j Y_{t-j} + s\epsilon_t^{mp} \tag{2.13}$$

2.5.2 Data

The endogenous variables Y_t in our model consist of a proxy for *Output* (ECB industrial production excluding construction), *Prices* (ECB harmonised index of consumer prices), *Commodities* (IMF Primary Commodity Price index), *Stock prices* (Euro Stoxx 50), *Financial stress* (ECB Composite Indicator of Systemic Stress, CISS) and 2-year German government bonds (*DE2Y*).[26] The monetary policy surprises shown above must be transformed into a monthly frequency. Following Gertler and Karadi (2015), I use monthly average surprises: The shock values of the elapsed 31 days are added up, and then the arithmetic mean of all

[25] See Gertler and Karadi (2015) for a detailed derivation.

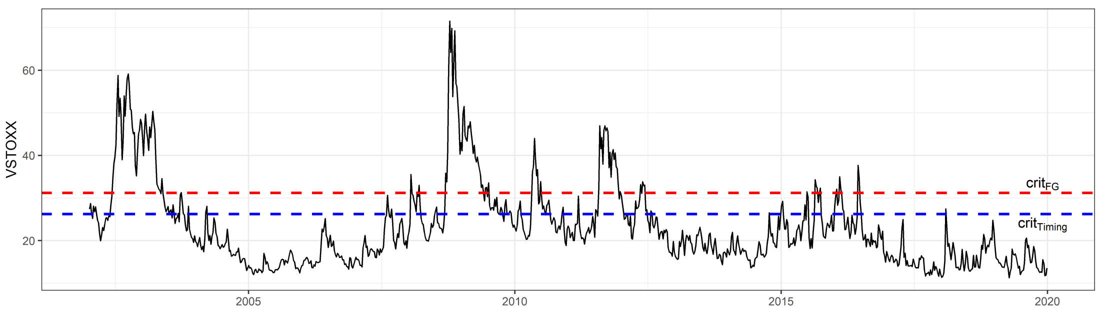
[26] The variables *Output*, *Prices*, *Commodities*, and *Stock prices* are used in logarithms. All four variables are seasonally adjusted.

accumulated values in each month is formed. This gives surprises at the beginning of the month a higher weight within that month than surprises at the end of the month, thus balancing variable meeting dates. The data covers the period from 2002:01 to 2019:12, which is due to EA-MPD availability. On the one hand, this broad sample size ensures efficient estimation using external instruments. On the other hand, the sample includes periods of high uncertainty, such as the Iraq war, the financial crisis, and the euro crisis, making it possible to distinguish effectively between periods of high and low uncertainty.

2.5.3 Poor-Man Sign Restriction

I aim to find out how expectation-forming monetary policy affects the economy in different states of uncertainty. The idea to identify the different shocks is based on the methodology of Jarociński and Karadi (2020).[27] For both surprise series, timing and forward guidance, I consider the index of the euro area's implied volatility and divide the series into two regimes: one with a high degree of uncertainty and one with low uncertainty. The results of the high-frequency estimation serve as reference values. With an index value of $crit_{Timing} = 26.2$ for timing and a value of $crit_{FG} = 31.2$ for forward guidance, there is an overall positive effect of the corresponding surprise on stock prices in the high-frequency data in Section (2.4).

Figure 2.4: Development of the VSTOXX index and the critical values for timing and forward guidance



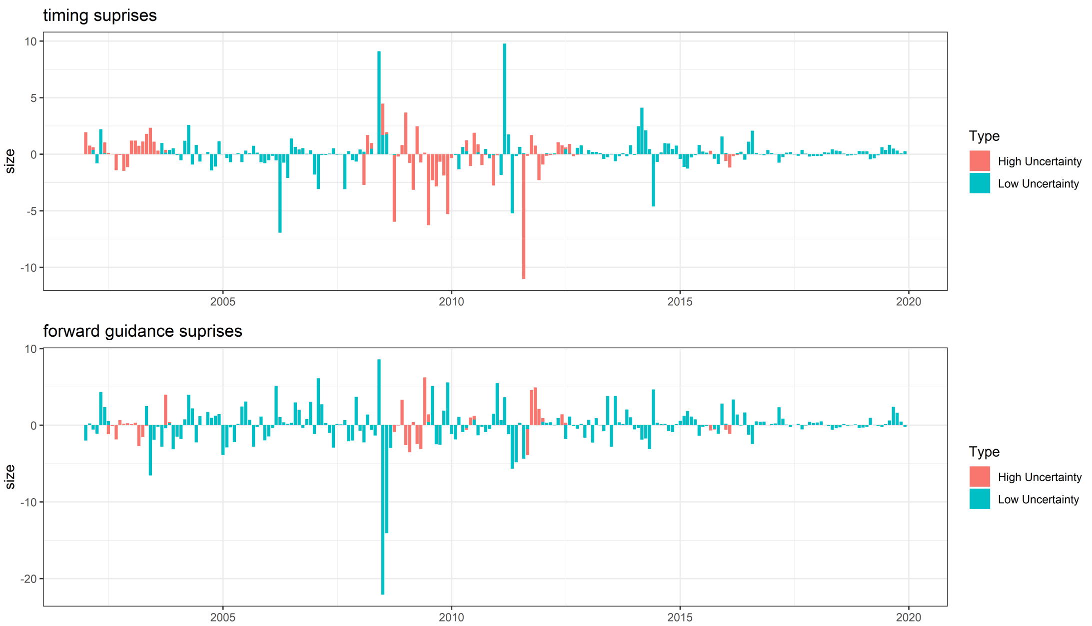
Therefore, all monetary policy surprises announced in an environment above these values are classified as high uncertainty states, all below as low uncertainty states.

[27] See Baumgärtner and Klose (2021) for an application of the original method with Euro area data.

$$
c_{t,j} = \begin{cases} crit_j > VSTOXX_{pd,t} & \rightarrow \text{low uncertainty} \\ crit_j < VSTOXX_{pd,t} & \rightarrow \text{high uncertainty} \end{cases} \tag{2.14}
$$

Here $c_{t,j}$ describes the event classification for monetary policy event t and measure j, $VSTOXX_{pd,t}$ the implied volatility from the day before the ECB decision and $crit_j$ as critical cut-off value for the surprise series j = timing, forward guidance. Figure 2.5 displays the resulting shock series.[28]

Figure 2.5: Classification of surprises according to the Poor-man restriction



Uncertainty is exceptionally high during periods of crisis, such as the financial crisis (2007-08) and the euro crisis (2011). Accordingly, surprises concentrate in these periods, and monetary shocks with low uncertainty are the standard case, and high uncertainty is the exception. The different values for $crit_j$ have the consequence that the number of surprises under high uncertainty is lower for forward guidance than timing surprises.

[28] Since I first classify the events and then convert them into monthly data using the average monthly surprises described above, it follows that both shocks can occur in the same month.

2.5.4 Instrument Validity

In the next step, the surprises are used as instruments to identify the VAR model. For a reliable estimation, two conditions must be met: According to (2.11b), the instrument must not be correlated with other shocks. This assumption seems reasonable, as the surprises observed are from a tight time window around the ECB's announcement (Kuttner, 2001). There are no indications that other events had a notable influence during this period (Brand et al., 2010; Altavilla et al., 2019).

Table 2.3: Regression of Residuals on Z

	Dependent variable:			
	residual DE2Y			
	High Uncertainty	Low Uncertainty	High Uncertainty	Low Uncertainty
$Timing_{high}$	0.029*** (0.007)			
$Timing_{low}$		0.015** (0.007)		
FG_{high}			0.014* (0.008)	
FG_{low}				0.013*** (0.003)
Constant	0.004 (0.008)	−0.001 (0.008)	−0.0004 (0.008)	0.0004 (0.008)
R-squared	0.085	0.026	0.012	0.071
robust F-statistic	15.734	4.447	2.992	17.832

Note: ***$p < 0.01$; **$p < 0.05$; *$p < 0.1$. The F statistics are calculated with heteroscedasticity and autocorrelation robust (HAC) standard errors.

Furthermore, the instruments must be highly correlated with monetary policy shock (2.11a) and therefore have explanatory power. To test whether the instruments are suitable, I regress the DE2Y residual ($\hat{v}_t^{mp}$) on my factors separately. Table 2.3 reports the regression results for each shock. I use the heteroscedasticity and autocorrelation consistent (HAC) covariance matrix for the F-statistic.

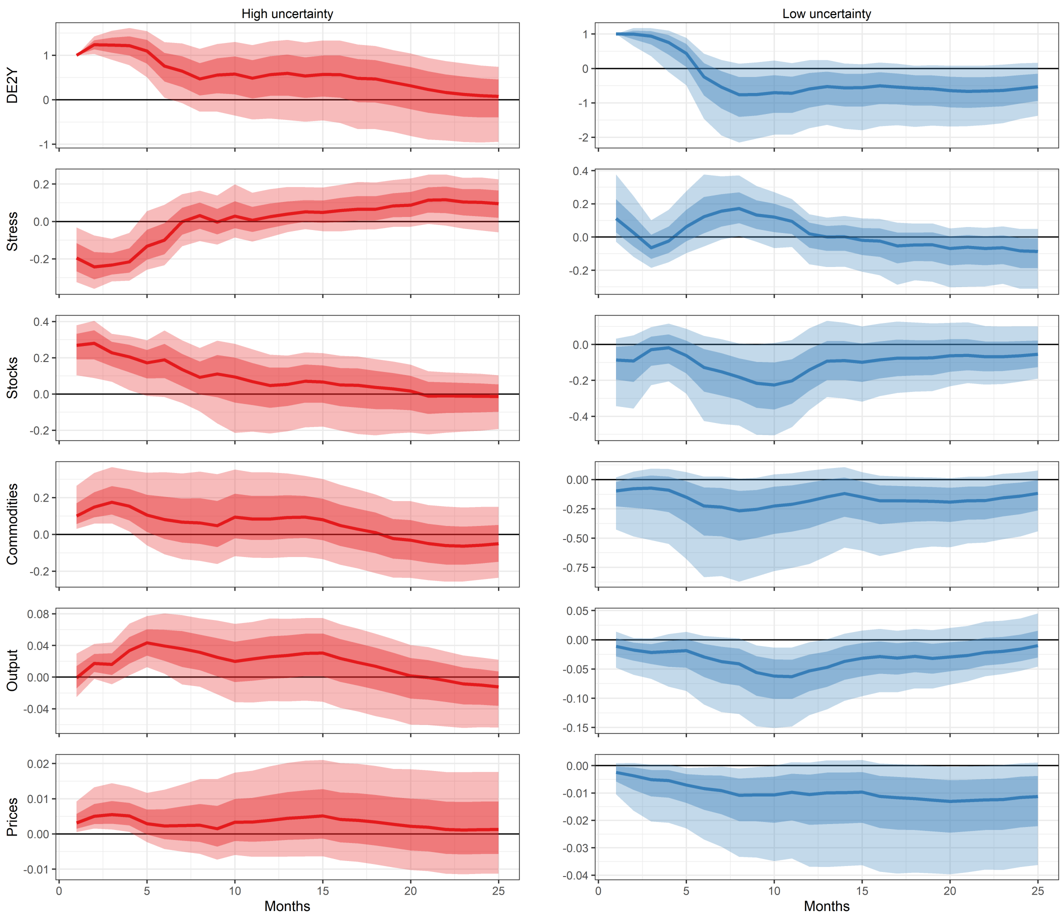
In the literature, an F-statistic of 10 is commonly considered the critical limit for the admissibility of the instrument (Stock and Watson, 2018). For all values above, the confidence intervals have the correct size. All values below this limit are at risk of the confidence intervals being too small. The results in Table 2.3 show parallels to Figure 2.3. Based on the F-statistic, there are no objections to an instrument estimate for timing shocks with high uncertainty and forward guidance shocks with low uncertainty. Figure 2.3 shows the areas where the parameter differs significantly from zero. There is a significant correlation between the timing shocks under low uncertainty and the forward guidance shocks un-

der high uncertainty, but it is not above an F-statistic of 10. A possible reason could be that, due to the different values for $crit_{Timing}$ and $crit_{FG}$, there are not enough observations in one of the two classes and that this leads to an inaccurate estimation.
Therefore, the entire estimation uses impulse response functions with robust confidence bands to consider the uncertainty caused by possible weak instruments (Montiel Olea et al., 2020). These bands remain valid under weak instruments concerns but have the correct size and coverage probability when the shock is highly correlated.

2.5.5 VAR Results

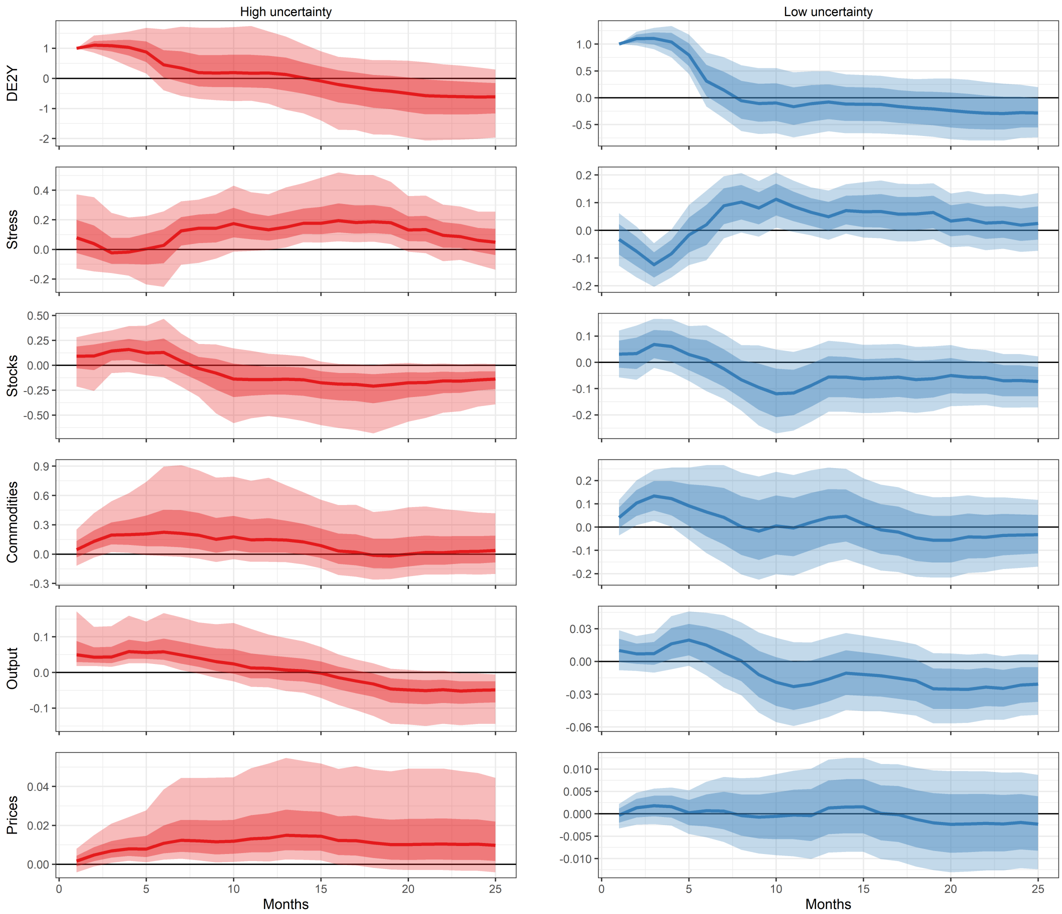
Comparing the impulse responses after a restrictive monetary policy shock, we find that the results differ considerably. Figure 2.6 shows the impulse responses of a restrictive timing shock in the estimated VAR models.
It is evident that the right column has the expected textbook behaviour with low uncertainty: Tight monetary policy lowers stock prices, in line with results from high-frequency data. Market participants expect the policy intervention to weaken future earnings, which reduces production and leads to a decline in prices. In contrast, a timing shock at high uncertainty produces different reactions. Interest rates rise, but the CISS decreases and stock prices rise, as do commodity prices. The impact on the output is positive. Contrary to the theory, prices do not fall but rather rise.
With forward guidance (Figure 2.7), the pattern of timing shocks does not repeat: Although a restrictive shock at high uncertainty leads to a response that is inconsistent with the theory, output and prices rise, no response is found in either variable at low uncertainty. Also, the pattern of rising stock prices with rising interest rates familiar from (Jarociński and Karadi, 2020) does not show up here. There are several possible explanations for this: First, long-term announcements might be generally ineffective (McKay et al., 2016; Baumgärtner and Klose, 2021). Second, it could be that the underlying cause of the different responses is different. The subsequent Chapter 2.6 elaborates on this idea by decomposing the VSTOXX.

Figure 2.6: Impulse responses of timing shocks



Notes: The shaded areas show the upper and lower bands of the 68% and 90% of the confidence intervals. The intervals shown are robust for weak instruments (Montiel Olea et al., 2020).

Figure 2.7: Impulse responses of forward guidance shocks



Notes: The shaded areas show the upper and lower bands of the 68% and 90% of the confidence intervals. The intervals shown are robust for weak instruments (Montiel Olea et al., 2020).

Overall, the results of the VAR model fit well with the observations in the high-frequency data. The pattern of timing impulse response functions are similar to that of information shocks (Jarociński and Karadi, 2020). However, the shocks are identified differently, using information already known before the ECB's announcement. In the instantaneous response of the impulse responses, I find similar behavior to the high-frequency variables, reinforcing the conclusion that uncertainty at the time of decision is the real driver of the information shock pattern.

2.6 Robustness

For now, the VSTOXX index has been the central proxy to measure uncertainty based on Bloom (2009). Simultaneously, there are doubts in the literature that this measure, based on financial markets, actually measures uncertainty or merely represents the effect of a fluctuating business cycle (Carriero et al., 2018; Ludvigson et al., 2021). To account for this and to check whether the observed results are due to uncertainty, I use a decomposition of the VSTOXX by Bekaert, Hoerova, and Lo Duca (2013) and Bekaert, Hoerova, and Xu (2021).[29] The authors show that the index contains an uncertainty component (UC) as well as a risk aversion component (RA).

In order to verify that the observed results are indeed due to uncertainty I integrate both components in Equation (2.4). In model (A) $VSTOXX_{pd}$ is replaced by UC_{pd}. Model (B) additionally integrates the risk component RA_{pd} to separate uncertainty and risk effects. Model (C) excludes all non-significant controls for a more efficient estimation. Table 2.4 shows the results.

First, we notice that the interaction term's coefficient between *timing* and UC_{pd} is positive across all specifications. This confirms the previous results. In periods of high uncertainty, tightening in the timing surprise is associated with rising stock prices. Second, we see that the positive coefficient of FG and UC_{pd} is present in model (A). However, models (B) and (C) show that this effect might be due to the omission of risk. The significant effect disappears by including risk; instead, the interaction term between FG and RA_{pd} is relevant. The dampening effect of forward guidance, therefore, seems to be related to risk rather than uncertainty.

[29] I thank Marie Hoerova for providing the data series.

Table 2.4: Regression of $\Delta STOXX$ on monetary policy surprises including Bekaert, Hoerova, and Xu (2021) uncertainty

	$\Delta STOXX$		
	(A)	(B)	(C)
Timing	-5.45	-4.38	-6.46^{***}
	(3.54)	(3.58)	(2.20)
FG	-7.18^{***}	-4.72^{*}	-4.07^{*}
	(2.46)	(2.64)	(2.25)
QE	-13.03^{*}	0.33	2.58
	(7.33)	(9.17)	(11.81)
$\mathrm{UC_{pd}}$	0.03	-0.61	-0.46
	(0.16)	(0.42)	(0.39)
$\mathrm{RA_{pd}}$		0.80	0.66
		(0.55)	(0.51)
$\mathrm{UC_{pd}} *$ Timing	0.15^{***}	0.19^{**}	0.23^{***}
	(0.05)	(0.08)	(0.07)
$\mathrm{UC_{pd}} *$ FG	0.11^{*}	-0.10	-0.12
	(0.06)	(0.11)	(0.10)
$\mathrm{UC_{pd}} *$ QE	0.04	0.34	0.18
	(0.20)	(0.25)	(0.23)
$\mathrm{RA_{pd}} *$ Timing		-0.08	-0.11
		(0.09)	(0.09)
$\mathrm{RA_{pd}} *$ FG		0.28^{**}	0.30^{**}
		(0.14)	(0.13)
$\mathrm{RA_{pd}} *$ QE		-2.54	-1.68
		(1.56)	(1.22)
$\mathrm{STOXX_{pd}}$	0.02^{*}	0.01	0.01^{*}
	(0.01)	(0.01)	(0.01)
US jobless claims	-2.83	5.38	
	(22.18)	(22.02)	
Projection Dummys	Yes	Yes	No
Adj. R^2	0.08	0.12	0.13
Num. obs.	189	189	189
F statistic	2.19	2.55	3.31

Note: $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$. The table shows the relationship between high-frequency variables and monetary surprises over a period from 2002-01-01 to 2019-12-31. The dependent variable is the change in either two-year OIS rates or STOXX50 stock prices within the two ECB announcement windows. Target, Timing, Forward Guidance, and QE describe the level of the respective monetary surprise. $VSTOXX_{pd}$ describes the closing price of the index of implied volatility on the day before the decision. All estimations control the stock price at the time of the decision, the release of the U.S. unemployment data, and the ECB projections' release. For better readability, the LHS variable is multiplied by 100. The test statistics are calculated with heteroscedasticity and autocorrelation robust (HAC) standard errors.

This could be a possible approach to explain why VAR models' results differ between timing and forward guidance.[30] Here, there seems to be a difference between monetary policy in phases of high uncertainty and phases of high risk.

2.7 Conclusion

In this paper, I show that the level of uncertainty at central bank events is of considerable importance for the impact of expectation-building measures. The high-frequency results suggest that uncertainty impacts stock prices' response after a timing or forward guidance shock. It turns out that uncertainty could explain the pattern used to identify information shocks: When interest rates increase and uncertainty is high, stock prices increase.

Similar results can be seen in VAR models. When uncertainty is integrated into the analysis, the impulse responses of timing shocks are similar to those in the information shock literature. When uncertainty is high, an expansionary (restrictive) timing shock lowers (raises) the price level. Additionally, under high uncertainty, we find the well-known pattern in the immediate responses of interest rates and stock prices: When interest rates rise, stock prices rise.

Furthermore, the analysis with euro area data shows that it may be useful to separate the channel known in the literature as path factor more precisely. There are apparent differences between individual measures. QE is not affected by the level of uncertainty, but the expectation-forming measures, timing, and forward guidance show significantly different effects in their impact on inflation. These two expectation-forming instruments also appear to have different effects. While timing shocks are influenced by uncertainty, forward guidance shocks react to prevailing risk. In general, the latter seem to have little effectiveness on the price level.

All in all, the results can potentially explain when information shocks occur. The question is what the channel of action is and why these market reactions occur. One possible hypothesis, following Bauer and Swanson (2020), would be that in periods of high uncertainty, it is worthwhile for both financial investors and

[30] To keep the paper straightforward, the VAR model with UC is shown in the Appendix A. The critical limits and, thus, the results do not change considerably.

companies to wait and postpone decisions until the central bank decision. The markets have an incentive to wait because they know that the central bank will act, but not how and to what extent. The markets wait for the evaluation of the crisis by the central bank and then include both shocks, the economic shock and the central bank reaction in their forecasts. In this context, expectation shaping measures are less effective. If the central bank responds to a crisis with timing or forward guidance surprises, this does not calm the markets. The negative effect of economic news is higher than the expansive monetary policy surprises. As a result, stock prices fall simultaneously as a monetary shock, which corresponds to the pattern known in the literature. Thus, if the central bank takes the prevailing level of uncertainty into account when choosing its policy measures, it is possible to assess the effects of the measures in advance and not only afterwards.

Therefore, future research should carefully consider uncertainty's potential effects, such as evaluating central bank measures' effectiveness. It is essential to understand in detail why uncertainty influences monetary policy. Further work in this area could take a more detailed look at uncertainty and further narrow down the source of uncertainty. This framework cannot clarify whether the effect found is limited to financial uncertainty or whether macro uncertainty also plays a role.

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

Martin Baumgärtner 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 banks. Finally, we demonstrate the applicability of embeddings in this paper by several examples in the fields of central bank objectives, financial uncertainty, and gender bias.

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

JEL classification: C45, C53, E52, Z13

*This essay was presented by me at Conference "Advanced analytics: new methods and applications for macroeconomic policy" (BoE, ECB and DAFM) and "Non-traditional Data, Machine Learning and Natural Language Processing in Macroeconomics" (BoC, FED, BoI). The Paper is currently under review at the Journal of Monetary Economics.

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

3.1 Introduction

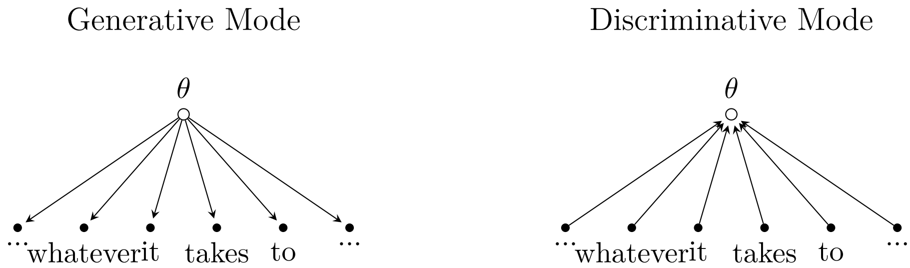
What did ECB president Mario Draghi mean on July 26, 2012, when he stated that *"within [its] mandate, the ECB is ready to do whatever it takes to preserve the euro."*? According to the current literature on central bank communication quantification, this is a neutral sentence. However, the message contained in the statement was nothing short of extraordinary for financial market participants and monetary policy experts; in fact, it marked a turning point in the ongoing euro crisis. We propose a novel language model in this paper that is able to capture such subtleties.

Over the last few decades, there has been an increase in the use of unstructured big data in monetary policy, in particular in the analysis and interpretation of central bank communication (Blinder et al., 2008). This development was certainly accelerated by 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 the desired equilibrium path. As a result, central banks increased their communication substantially. The FOMC, for example, started publishing press conferences in 2011, and the ECB began disclosing monetary policy meeting minutes in 2015.

The analysis of central bank communication is based on the presumption that it contains latent messages (θ) by the monetary policymakers, worth extracting. These messages can be discrete, such as a Bank's stance in a policy debate, or continuous, such as signaling policy direction or communicating the Bank's preferences. While not observable directly, the θ's generate variations in the communication, and hence the words used (W), a process as depicted in Figure 3.1 on the left side. Since only the outcome of this sampling process can be directly observed, it is the receivers' job to infer the underlying message from the variation in W, as illustrated by the right-hand side. This paper aims to provide a representation for words that allows simple models to retrieve the underlying messages from the observed variation in central bank communication ($W \rightarrow \theta$).

Decoding of messages is most effective when the language used is stable, homogeneous, and represented in its richness. The current string in the central bank

Figure 3.1: Communication Model



Notes: The illustration is adapted from Lowe (2021, p. 10).

communication literature uses pre-defined dictionaries, such as Loughran and McDonald (2011), Apel and Grimaldi (2014), and Picault and Renault (2017) to count terms (for example, positive and negative) to extract a single dimension (for example, sentiment) from document. Such a practice equates to an extreme prior of the informativeness of the vast majority of communicated terms, which may only suffice for simple messages, thereby falling short of capturing the domain-specific richness of the representation.

To address these shortcomings, modern linguistics and computer science has turned to machine learning to develop novel *language models*. Such models are estimated from a set of text – the *corpus* –, and an *algorithm* that locates words in a multidimensional vector space. Conceptually similar terms are mapped in close proximity, Meanwhile models such as Mikolov, Yih, et al. (2013) and Pennington et al. (2014), leverage large corpora from a variety of sources, such as Twitter or Google searches, and thus allow for little inference about the technical language used by monetary policymakers, violating the condition of stable and homogeneous 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 one previously used in the central bank communication literature. Second, this paper introduces novel machine learning algorithms for text quantifying. We compare a multitude of different algorithms according to objective criteria. Doc2Vec, an algorithm that leverages the word and document space, outperforms the others in our evaluation. Third, by training the novel algorithm on the novel text corpus, we introduce a language model previously unseen in monetary policy (and likely economics at large). We demonstrate how this language model can be used in various applications throughout this paper, such as comparing central bank objectives, measuring the effect of central bank communication in times of heightened uncertainty, and evaluating gender bias. Finally, by making the language model publicly available,[31] this paper's most important contribution is to provide this new string of research to other researchers, allowing them to incorporate embeddings into their research.

The remainder of this paper is structured as follows. Section 3.2 provides a literature overview of the current state of natural language processing (NLP) in monetary economics. In Section 3.3 we introduce both the text corpus and the algorithms, combining both elements into language models used to represent W. We then evaluate the quality of the resulting embeddings in the central bank context in Section 3.4 before applying the best-performing language model in Section 3.5, essentially providing possibilities of inference ($W \rightarrow \theta$). The final section concludes this paper.

3.2 Related literature

Natural language processing (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,

[31] `sites.google.com/view/whatever-it-takes-bz2021`

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 as captured by (the aggregate consequences of) their behavioral responses to it. Among successful implementations are Gürkaynak et al. (2005), Brand 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, and Tillmann (2021) classifies answers during the ECB press conference's Q&A to estimate a disagreement index.[32]

However, most applications today concentrate on rule-based classification utilizing computers. Precisely, 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 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

[32] One notable shortcoming the quantification literature (and this paper), is the focus on the supply of provided information, omitting potential demand effects. However, Tillmann (forthcoming) shows that market participants react to surprises in the expected manner.

and McDonald, 2011; Apel and Grimaldi, 2014; Picault and Renault, 2017; Correa 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), home biases of central bankers (Hayo and Neuenkirch, 2013) or financial stability objectives (Peek et al., 2016; Wischnewsky 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 omit relevant information. In terms of Figure 3.1, the $\theta \rightarrow W$ relationship is characterized by a prior of zero for the majority of the modeled words. The issue of excluding a substantial portion of text has been articulated before by Harris (1954, p. 156), stating 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"*

In addition, dictionaries are inherently 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 the low dimensionality and the coarseness of the interpretation of the message that comes along with it, 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.

Recent research recognizes and highlights the dictionary approach's limitations to capture the messages' meanings, suggesting either augmenting such an index or combining different dictionaries to improve predictive power. Tadle (2022), 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 contains 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 et al. (2020). A different approach is applied by Azqueta-Gavaldon et al. (2020), Kalamara et al. (2020), Shapiro, Sudhof, et al. (2020), and Gorodnichenko 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 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 as captured by its measure 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, p. 533) quote. Word embeddings are multidimensional word representations that are used to measure similarity in Twitter tweets (Masciandaro et al., 2020), for the improvement of the Euro Area uncertainty index (Azqueta-Gavaldon et al., 2020), for the decomposition of central bank vague talk (Hu and Sun, 2021), for the analysis of FOMC introductory statements (Handlan, 2020), and for measuring central banker disagreement (Apel, Blix Grimaldi, et al., 2019). Economic research in this field relies on general language models trained on a general text corpus such as Wikipedia. Shapiro, Sudhof, et al. (2020), for example, use Pennington et al.'s (2014) 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 in Shapiro, Sudhof, et al. (2020) and may 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. Their embeddings, however, are not a focal part of the paper and are thus not suitable for general-purpose quantifying central bank communication.

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 contributes to two current desiderata in this literature. On the one hand, the development of novel text-representation (Apel, Blix Grimaldi, et al., 2019), and on the other hand, the need to fine-tune these representations for their respective use (Loughran and McDonald, 2011).

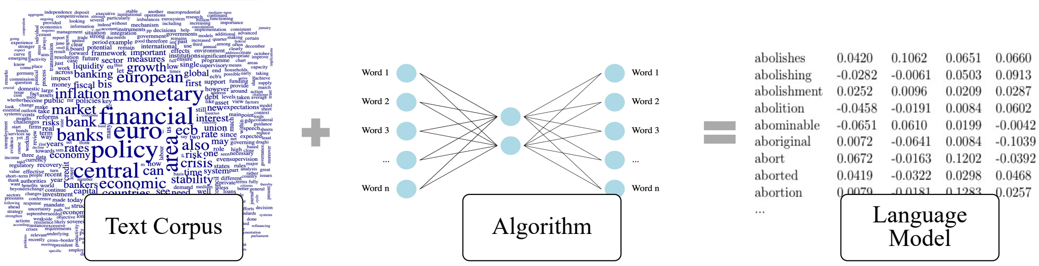
3.3 Methodology

> *"The meaning of words lies in their use. [...] One cannot guess how a word functions. One has to look at its use, and learn from that."*
>
> — 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, derived from training an algorithm on a text corpus. Embeddings, thereby, provide a nuanced representation of the words (W). Our paper proposes a method for text classification that is detached from causal inference, called *transfer learning*. Transfer learning describes process in which specialized knowledge is gained by working on one task and is subsequently applied to a different, but related, task. As a result, we avoid potential conflicts that arise when dimension reduction and the application of dimension-reduced variables are performed simultaneously (e.g. Egami et al., 2018).

Figure 3.2 provides a stylized overview of the procedure how to retrieve a language model. The figure also reflects the structure of the remainder of this section.

Figure 3.2: How to retrieve a language model



3.3.1 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). [33] We complement the 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 literature in their jargon and are thus not official central bank communication. The use of information on the respective institutions 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.

In contrast to the previous NLP applications 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

[33] We determine the language of the individual documents using Google's Compact Language Detector 3 and clean the corpus accordingly.

Table 3.1: 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

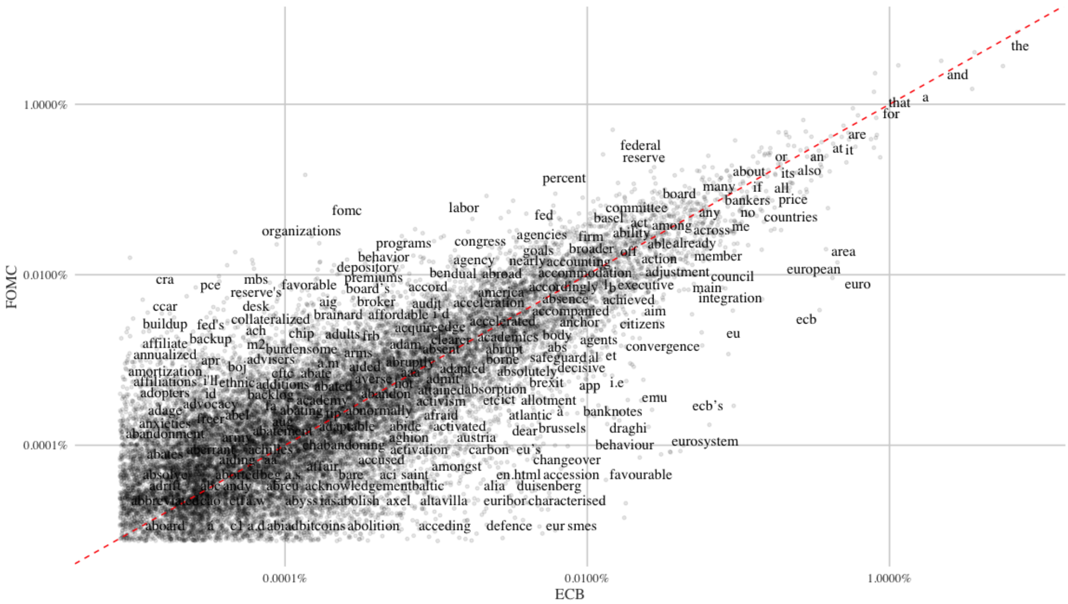
Note: The table summarizes the number of documents (n) by sources in the our text corpus.

done in the embeddings literature (e.g. Mikolov, Yih, et al., 2013) since similar words should be in near proximity 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. 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 words *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 instrument 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

Figure 3.3: 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.

[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 over a period from the 1980s to early 2021.[34]

The result is a corpus of text that, on the one hand, is unprecedented in quantity and diversity in the monetary communication literature, and, on the other hand, contains exclusively highly specific central bank vocabulary.

The corpus' stability with respect to word use is what we address next. We test whether the usage of language can be assumed rhetorically stable across institutions and time, a necessary condition to allow for inference.[35] We use the

[34]The distribution of documents is heavily skewed to the left, as central banks have only increasingly published documents over the Internet since the 2000s. An overview is provided in Figure B.1 in the Appendix.

[35]An example of rhetorical instability is the Google Flu Trends Project (Lazer et al., 2014), which used flu-related Google searches to predict medical appointments. The project was discontinued in 2015 due to severe misjudgment by the algorithm caused by changes in search behavior.

central bank's jargon, the relative word frequency, for the seven most frequent central banks in our sample as an approximation. An illustration of the relative word frequencies for the ECB and the FOMC is provided in Figure 3.3. 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.[36] We conclude from these observations that the institutions do not differ in any relevant way concerning their jargon. In order to test for temporal stability, we compare the jargon of the central bankers across time. The results remain quantitatively the same, illustrating that the usage of language did not change markedly.

3.3.2 Algorithm

Modern language models follow the proposition of linguistic Zellig S. Harris in their pursuit of superior text representation. According to Harris (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 hypothesis 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 – how often words appear together. The following section introduces both methods and their most prevalent techniques.

[36] The precise values are: Federal Reserve (FED): 98% (t = 884.67), Riksbank: 98% (t = 585), Bank of England (BoE): 98% (t = 966), Bank of Japan (BoJ): 98% (t = 668), Bundesbank: 99% (t = 1257), and Central Bank of India: 98% (t = 783). The results are also illustrated in the Appendix in Figure B.2.

Prior to formally introducing the algorithms, we provide a simplistic example to facilitate comprehension between the concepts of target words and context. Following Harris (1954)'s distributional hypothesis, a word's meaning is based on the environment in which it appears. The context of a word, the set of its adjacent 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

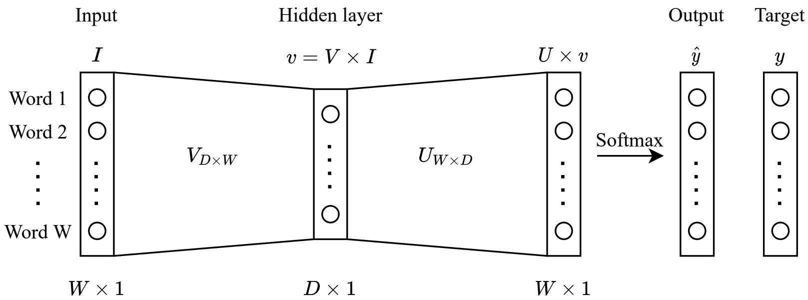
The prediction-based algorithms are generally tasked to predict the target word given the context words, i.e. *P(brighter | this, outlook)*. They then proceed with the next target word, i.e. to predict *P(outlook | brighter, remains)*, then *P(remains | outlook, subject)* and so on.[37] By optimizing 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 distributional semantics 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 ______"* 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 et al. (2003), Collobert and Weston (2008), and

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

Turian 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 algorithms, 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 3.4.

Figure 3.4: Graphical illustration of Mikolov, Yih, et al. (2013)'s 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.

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:

$$L = \frac{1}{T}\sum_{t=1}^{T} \log P(w_t|w_c). \tag{3.1}$$

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

$$P(w_t|w_c) = \frac{exp(u_{w_t}^T v_{w_c})}{\sum_{w=1}^{W} exp(u_w^T v_{w_c})} \tag{3.2}$$

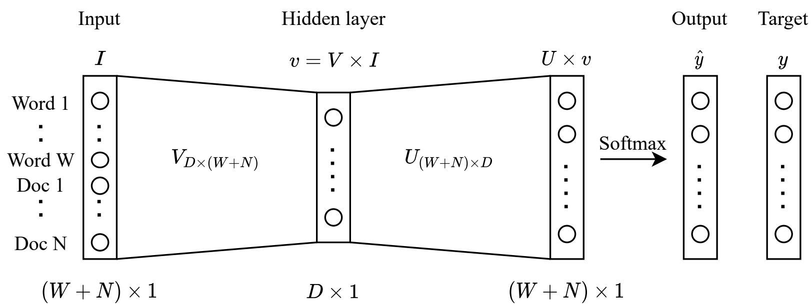
where v_{w_c} is the embedding vector. In other words, the models' functional structure represents a single linear hidden layer linked to a softmax output layer, where the exponential function prevents negative numbers and could be omitted without loss of generality. The objective is maximized using 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, 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 (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 following algorithm) 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.

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 3.5.

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

Figure 3.5: 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 3.4 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.

can be used as a summary of the document in subsequent regressions. For example, in Section 3.4 and Section 3.5, we demonstrate how similarity in document embeddings may approximate in institutional differences by central banks. However, it should be noted that, unlike word embeddings, document embeddings cannot be easily transferred to new corpora.

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. Hansen and McMahon, 2016; Tobback et al., 2017; Hansen, McMahon, and Tong, 2019; Wischnewsky 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 et al. (2015) for an introduction to LDA in monetary policy NLP 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 application 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), although in general, the number of topics does not exceed 50 in the economic literature. 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 GloVe, a global factorization method. Following the success of Word2Vec, Pennington et al. (2014) propose GloVe, which trains a language model on 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:

$$L = \sum_{t,c=1}^{W} f(X_{t,c})(w_t^T w_c + b_c + b_t - \log X_{t,c})^2. \tag{3.3}$$

In Equation (3.3) 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 (3.3) 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.

Table 3.2: 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 et al. (2014).

3.3.3 General corpus models

As mentioned in the introduction, to the best of our knowledge, so far no attempts have been made to train embeddings specifically for the economic context. This may be due to the computational burden or the necessary amount of text. An alternative to training embeddings from scratch is the use of pre-trained general language models using transfer learning (e.g. Binette and Tchebotarev, 2019; Doh et al., 2020; Istrefi 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 all our embeddings to two such general models as a benchmark: Glove6B and Word2Vec Google News.[38]

In Table 3.2, 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 also different specifications. The hyperparameters we use for each model can be found in the Appendix B.2.

3.4 Evaluation of language models

In this section, we apply the algorithms introduced in the previous section to our corpus and evaluate the corresponding language models. We aim to deter-

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

mine the algorithm that best summarizes the content of our central bank corpus 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.[39]

Since there exists no benchmark for evaluating language models in economics yet, we turn to the fields of computational 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 embeddings indicate associations similar to humans' perceptions. Another task would be the ability to find word analogies that resemble real analogies. We present several intrinsic evaluations in the second part of this section.

3.4.1 Extrinsic evaluation

Extrinsic tasks involve evaluating the embeddings 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 are designed to evaluate embeddings in a broad context, while we are interested in the opposite, their domain specificity. Due to a lack of external evaluation methods, we benchmark the embeddings in the following two steps. First, we test how well the models can predict words and then assess the predictive performance of each model in a monetary policy classification task (Le and Mikolov, 2014). We demonstrate in Section B.3.2 that the presented results are robust to more general tasks.

In the absence of an established procedure, we use an unsupervised approach that takes advantage of Harris's (1954) notion that context defines the meaning. Good language models should be able to predict terms using their adjacent words. Thus, each model is presented a task such as the following: predict the word *substantial* given the bag of words [*outlook, remains, subject, to, uncertainty, also, regarding,*

[39] Whenever we evaluate the word embeddings on document level, we average over all word vectors of a document.

the].[40]

Table 3.3: Evaluation results of word prediction task.

Algorithm	Accuracy	Standard deviation
Word Embeddings		
Doc2Vec Bow	<u>0.846</u>	0.007
Doc2Vec Bow Pre	<u>0.844</u>	0.009
GloVe	0.831	0.008
Doc2Vec PVDM	0.803	0.009
Doc2Vec PVDM Pre	0.800	0.017
Word2Vec Skipgram	0.678	0.007
GloVe 6B	0.646	0.008
Word2Vec GoogleNews	0.546	0.016
Word2Vec Bow	0.502	0.009
LDA	0.064	0.014

Note: The table shows the evaluation results across the different algorithms introduced in the previous section. The accuracy was evaluated as the Number of correct predictions / Total number of predictions. With regards to the specifications: Bow = (Distributed) Bag Of Words; PVDM = Paragraph Vector Distributed Memory; Pre = pretrained embeddings were used as more efficient starting points.

The results are depicted in Table 3.3. There are four noteworthy results. First, the Doc2Vec and GloVe models perform best, correctly predicting more than 80% of the words. Second, the Bag-of-Words models outperform the others in this group, by an additional 5% accuracy. Third, each model's performance does not vary much across folds. In this context, it should be noted that there is no statistically significant difference between the two top models. Finally, the general language models do not fare as well in terms of relative performance, which emphasizes the importance to train on monetary policy documents.

Our second evaluation task concerns the current interest rate level of the ECB and FOMC, which we forecast using the respective central bank's speeches. Since we are primarily concerned with the correct level, we divide the corresponding 3-month interbank rates into quintiles to derive our evaluation target.[41] We are

[40]We train a neural network with a hidden layer. The results presented are simulated out-of-sample predictions with 10-fold cross-validation. More information is provided in Section B.3.1.

[41]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.

interested in the best performance, therefore, we employ a neural network to predict the respective interest rate levels with our embeddings.[42]
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 prediction.

Table 3.4: Evaluation results of algorithms.

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

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 uninformed performance would be $1/6 \approx 0.17$. With regards to the specifications: Bow = (Distributed) Bag Of Words; PVDM = Paragraph Vector Distributed Memory; Pre = pretrained embeddings were used as more efficient starting points.

Table 3.4 summarizes the accuracy of the predictions split by Document- and Word Embedding as well as task. Since there exist several variants in the Word2Vec and Doc2Vec algorithms and we aim for a broad comparison, we es-

[42]We employ a single hidden layer neural network with 64 units and dropout regularization. We tested various specifications, but the performance does not change substantially. The exact parameterization is available upon request.

timate them all. The name in column one starts with the algorithm followed by the variant's abbreviations.

Our evaluation yields some interesting results. First, the federal funds rate level appears to be more challenging to predict across all 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, confirming previous results by outperforming the others on both the document and word levels. [43]

3.4.2 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 interpreted 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:

$$similarity_{a,b} = \frac{a \cdot b}{||a|| \times ||b||} = \frac{\sum_{i=1}^{n} a_i \times b_i}{\sqrt{\sum_{i=1}^{n} a_i^2} \times \sqrt{\sum_{i=1}^{n} b_i^2}} \tag{3.4}$$

In the first evaluation, we select economic concepts in the word embedding space and assess the most similar words to these concepts. The results are presented in Table 3.5, for the words *inflation*, *unemployment*, and *output*.[44]

It is evident that our language model is capable of grouping words with semantically similar meaning. For example, it is reassuring that 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,

[43] The upcoming results are robust across all 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".

[44] On our website (https://sites.google.com/view/whatever-it-takes-bz2021) we provide an interactive tool that allows users to make the same assessment for any word in the entire vocabulary.

Table 3.5: 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 (3.4). The underscore is used to highlight collocations as described in Section 3.3.1.

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. Homonyms arise because their meaning differs in different context. 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 property by estimating the similarity to the term *basel* and compare our results to the general language model GloVe6b and GoogleNews. The results can be found in Table 3.6, 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).[45]

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

[45] In the Appendix, we provide additional examples for the interested reader.

Table 3.6: Intrinsic Evaluation: Similarity to Basel across language models

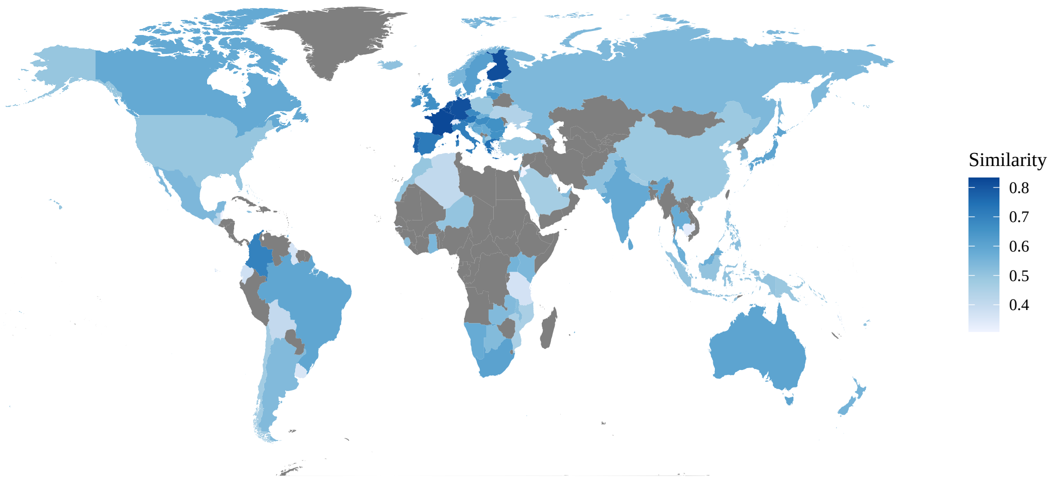
Doc2Vec	GloVe6B	Word2Vec GoogleNews
basel_committee	zurich	abbr
basle	basle	Tst
capital_accord	zürich	iva
basel_accord	bern	tHe
bcbs	switzerland	Neurol
basle_committee	stuttgart	BASLE
basel_ii	hamburg	PARAGRAPH
basel_iii	cologne	tellus
consultative	lausanne	Def.
minimum_capital	schaffhausen	Complementarity

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 (3.4). The underscore is used to highlight collocations as described in Section 3.3.1.

Figure 3.6 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 intuition. This observation is investigated further in our first application in Section 3.5.

To summarize, we used the previously introduced algorithms for quantifying words and documents in this section. We evaluated all methods using out-of-sample predictions and selected the Doc2Vec on the basis of this evaluation. 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 3.6: 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 (3.4).

3.5 Applications

In this section, we will demonstrate how our our Doc2Vec language model can be used to retrieve latent messages, i.e. identifying avenues for $W \rightarrow \theta$. The applications are intended to provide case studies for the use of embeddings via transfer learning. The source code for all applications can be found online.[46] 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.

The first application assesses whether central banks' objectives drive the differences in similarity we reported in the previous section. We find that inflation targeting central banks are indeed 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. Our findings indicate that in times of crisis ECB communication can calm financial markets. In our final application, we investigate prejudices and biases in the technical language of central bankers across the globe.

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

3.5.1 Monetary policy framework classification

In this application, we investigate whether banks' institutional settings can explain the differences in the similarity of communication. Institutional classifications are inherently multidimensional; we only address aspects that are considered relevant for monetary policy. Our analysis relies extensively on Cobham (2021), who uses the IMF's Article IV Consultation Reports to classify de jure monetary policy frameworks, on an annual basis following the end of the Bretton Woods system. A monetary policy framework refers to the *"objectives pursued by the monetary authorities, but also the set of constraints and conventions within which their monetary policy decisions are taken."* (Cobham, 2021, p. 1). David Cobham identifies ten target variables (inflation, money supply, and others) that can be further subdivided into 32 mutually distinct categories, ranging from *loosely structured discretionary targets* to *fully converging inflation targets*. The classification, which covers approximately 150 central banks, is available online. Merging the monetary policy framework with our corpus yields more than 80 central bank classifications and more than 800 country-year observations.[47]

As outlined in Section 3.4, we compute the cosine distance of a central bank's average annual embedding towards a specified institution. The question as to which particular monetary policy institution is used for comparison is ultimately left to the researcher's discretion. Since we focus on monetary policy objectives and inflation targeting is prevalent, we select three institutions that have different histories with respect to this objective. Specifically, we selected the first inflation-targeting central bank (the Reserve Bank of New Zealand (RBNZ)) and two prominent ones in our corpus (the ECB and the FED) because they provide interesting variations given their different institutional settings and objectives, e.g., the FED has a dual mandate, while the ECB has a primary and a secondary objective. In the following, we will refer to those three as benchmark central bank.

Econometrically, we run an OLS regression of the similarities ($S_{i,j,t}$) between a central bank j and the benchmark central bank $i \in \{RBNZ, FED, ECB\}$ on

[47] Members of a currency area are assigned the classifications of the currency area's lead central bank, as opposed to omitting these observations.

the central bank target ($Target_{j,t}$) defined by Cobham (2021). To control for macroeconomic conditions, we take the difference of three macroeconomic indicators (inflation, unemployment, log(GDP)) towards the benchmark central bank, i.e., $\Delta X_{i,j,t} = X_{j,t} - X_{i,t}$. Finally, we control for euro area members ($EA_{j,t}$).[48].

$$S_{i,j,t} = Target_{j,t} + \Delta X_{i,j,t} + EA_{j,t} + \epsilon_i \tag{3.5}$$

In a first step, we examine only the general differences between institutions labelled *inflation targeting* and those otherwise using a dummy variable (ITs) that takes the value of one for inflation targeting central banks. Results are reported in specification (1) in Table 3.7. We find a consistently positive, significant, and economically relevant coefficient in all three benchmarks, suggesting that central banks with inflation targeting communicate more similarly relative to the RBNZ, the FED, and ECB. When accounting for macroeconomic differences and euro area banks, the results persists.

In a second step, we examine the similarities on an annual basis and, moreover, include all regimes. Thus we now exploit the full range of Cobham's (2021) classifications (the footnote of Table 3.7 provides an overview).

The results are shown in model (2). For all three benchmarks, the coefficients on the inflation target increase significantly. Moreover, a more nuanced perspective on the noninflationary institutions emerges. While all three central banks appear to communicate very similarly to the only well-structured (WSD) central bank in our sample (Bank Negara Malaysia), only the ECB is significantly similar to other regimes. A potential explanation may be the ECB's composition of several previously classified central banks as ERT (e.g., Austria and France), MixedT (e.g., Germany and Italy), and LSD (e.g., Greece).

In a final step, we are interested which inflation targeting characteristics influence our results. Therefore, we partition the inflation targeting category further into loose inflation targeting LIT (e.g., euro area, US until 2011, South Africa) and full inflation targeting FIT (e.g., New Zealand, US since 2011, Poland), as well as a converging category for each, representing non-constant targeting over time. The

[48] Neither the choice of macroeconomic variables nor the dummy seems to affect the results. Results are available upon request

Table 3.7: Regression results: Monetary Policy Framework classification

	Dependent Variable: Similarity towards bank i								
$i =$	RBNZ			Federal Reserve			European Central Bank		
	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
ITs	0.07***	0.11***		0.06***	0.09***		0.09***	0.16***	
	(0.02)	(0.03)		(0.02)	(0.02)		(0.02)	(0.03)	
- FIT			0.17***			0.12***			0.12***
			(0.03)			(0.02)			(0.03)
- LIT			0.11***			0.10***			0.18***
			(0.03)			(0.02)			(0.02)
- FCIT			0.05			0.10***			0.14***
			(0.04)			(0.04)			(0.04)
- LCIT			0.06*			0.02			0.09***
			(0.03)			(0.03)			(0.03)
WSD		0.12***	0.14***		0.10***	0.11***		0.17***	0.15***
		(0.04)	(0.04)		(0.03)	(0.03)		(0.04)	(0.04)
LSD		0.05*	0.06**		0.04*	0.04*		0.08***	0.06***
		(0.03)	(0.03)		(0.02)	(0.02)		(0.03)	(0.02)
ERTs		0.04	0.06**		0.001	0.01		0.09***	0.08***
		(0.03)	(0.03)		(0.03)	(0.03)		(0.03)	(0.03)
MixedTs		0.06	0.09*		0.03	0.04		0.14***	0.13***
		(0.05)	(0.05)		(0.04)	(0.04)		(0.05)	(0.05)
NNF		−0.06	−0.04		−0.12	−0.12		0.01	−0.003
		(0.10)	(0.09)		(0.09)	(0.08)		(0.09)	(0.09)
Constant	0.32***	0.31***	0.29***	0.32***	0.43***	0.42***	0.40***	0.41***	0.41***
	(0.02)	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)	(0.02)
Macro. Controls	yes	yes	yes	yes	yes	yes	yes	yes	yes
Observations	84	821	821	84	825	825	83	821	821
R^2	0.15	0.18	0.24	0.20	0.19	0.22	0.28	0.31	0.37
Adjusted R^2	0.11	0.17	0.23	0.16	0.18	0.21	0.24	0.30	0.36

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. We adapt the notations directly from Cobham (2021): ITs = inflation targets; LIT = loose inflation targeting; LCIT = loose converging inflation targeting; FIT = full inflation targeting; FCIT = full converging inflation targeting; WSD = well structured discretion; LSD = loose structured discretion; ERTs = exchange rate targets; MixedTs = mixed targets; NNF = no national framework.

results are interesting both within and across benchmarks, due to the different relative weights depending on the institution being compared. The similarity towards the RBNZ (always FIT) is significantly higher for inflation-targeting institutions only. The FED, which transitioned from LIT to FIT, has nearly equal weight between both, whereas the ECB (which has always been LIT) is closer to LIT institutions.

As a robustness check, we conduct the same regression using the similarity between the word embeddings.[49] 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 the similarity among central banks embeddings is the adoption of a mutual objective and framework. This implies that researchers can use public communication when trying to abstract a central banks monetary policy framework, or, as we will do in application three, its preferences.

3.5.2 Whatever it takes

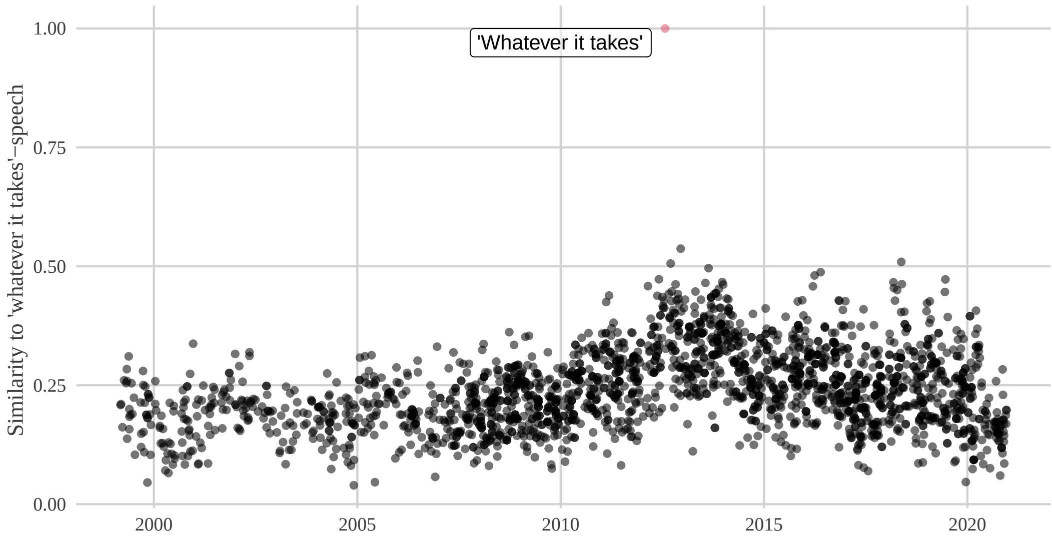
The second application focuses on the question with which this paper opened. We attempt to measure the effect of central bank communication in times of heightened uncertainty, utilizing the document space. There is literature on this topic using word embeddings (Azqueta-Gavaldon et al., 2020), with a focus on measuring uncertainty using news articles and not central bank communication. 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 between the ECB's remaining speeches to this event's embedding, thereby creating a time-series, indicating the central bank's willingness to act as a lender of last resort. Figure 3.7 illustrates that, particularly during the euro area crisis, the

[49] Remember that the jargon used by central banks is very similar, as highlighted in Section 3.3.1.

embeddings of central bank speeches appear more similar to the *whatever it takes*-speech. To investigate whether the similarity to that speech can calm financial

Figure 3.7: Similarity of all ECB 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 (3.4).

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

$$\Delta spread_{10y,t} = wit_{simil,t} + Unc_t + wit_{simil,t} \times Unc_t + X_t + \epsilon_t \tag{3.6}$$

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.[50] 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 uncertainty (UC) and risk aversion (RA) based on Bekaert, Hoerova, and Xu (2021),[51] and finally the ECB's daily CISS index (Hollo et al., 2012). X represents a set of control variables, among them a dummy for the wit speech, Moody's agency ratings for Greek bonds, European and U.S. stock prices, monetary policy surprises based on Altavilla et al. (2019), and a dummy for the ECB's different

[50]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.

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

central bank presidents. Due to considerable risk of autocorrelation, we integrate the first lag of the bond spreads.

Table 3.8: 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($\Delta spread_{10y}$, 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 3.8.[52] Starting with the first benchmark, we 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 levels ($VSTOXX <$

[52] The full table including all controls can be found in Appendix B.3.4.

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, raising spreads.

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. The decomposition of Bekaert, Hoerova, and Xu (2021) highlights that our effect is driven by uncertainty UC_{pd} rather than risk RA_{pd}. Only the configuration with the $CISS$ shows a generally lower level of significance.

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.

3.5.3 Gender Bias

> *"We should mirror the society we serve."*
>
> — Christine Lagarde (ECB, 2020)

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.

Numerous studies have established that women in economics are severely underrepresented (e.g. Ginther and Kahn, 2004; Lundberg, 2017; Auriol et al., 2020) and central banks are no exception (Hospido et al., 2019). This misrepresentation of societal composition might hinder effective monetary policy by eroding legitimacy by their constituencies (see quote above) or communicating less effectivly (D'Acunto et al., 2021). We would like to contribute to the growing literature (1) by providing evidence whether the language used by central bankers is biased and (2) by measuring the extent to which the emphasis on this issue has changed over time.

Our analysis on biased language builds on a fast growing literature that identifies biases in publicly available embeddings (e.g. Caliskan et al., 2017; Garg et al., 2018; Manzini et al., 2019; Sweeney and Najafian, 2019; Badilla 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.[53] As a result, any language model derived from a biased text corpus must inherit these biases as well.

We are following Garg 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 (e.g., childcare) can then be used to calculate the prejudice associated with this apparently 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 word lists v_a and v_b with n dimensions each and a neutral word w, the RND can be calculated by:

$$RND_{a,b} = \sqrt{\sum_{i=1}^{n}(w - v_{a,n})^2} - \sqrt{\sum_{i=1}^{n}(w - v_{b,n})^2} \tag{3.7}$$

In order to test for underlying biases in our embedding, we use academic professions.[54] The most feminine and masculine programs according to this measure as applied to our language model can be found in Table 3.9.[55] With a few exceptions, male pronouns are most closely associated with STEM fields, whereas

[53] It should be noted that our first application provides evidence that central bank objectives can be approximated by analyzing the language of institutions.

[54] We use data from Eurostat on students in Europe enrolled in Bachelors Programs by sex.

[55] We estimate the RND for each study program concerning a set of male and female pronouns as suggested by Garg et al. (2018). 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 3.9: 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 et al. (2018). It illustrates the subset of occupations most associated with gender pronouns.

female pronouns are most closely associated with care-taking and education.

To formally test whether the association could be due to the prominence of a gender in the respective academic occupation, we run a simple OLS regression with the female percentage in that field as the explanatory variable. The result can be found in Table 3.10. The proportion of women in fields more closely associated with female pronouns is indeed significantly higher, and vice versa.

Table 3.10: 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 (3.7). Higher values indicate closer association to female pronouns and lower values closer association with make pronouns. 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.

As central banks have recognized the problematic lack of of gender diversity (e.g., the ECB launched its diversity program in 2010), it may be tempting to assess its development over time. Therefore, we next examine whether the language use of ECB executives has changed concurrently with the implementation of the diversity policy.

In order to empirically test a potential shift in time, we rely on the Hospido et al.'s (2019) findings. Using ECB staff income data from 2003 to 2017, the authors identify a significant pay gap between men and women, with the gap disappearing after 2011 when the ECB announced a series of measures to promote gender diversity.

Figure 3.8: Similarity of all ECB speeches to Sabine Lautenschläger's speech.

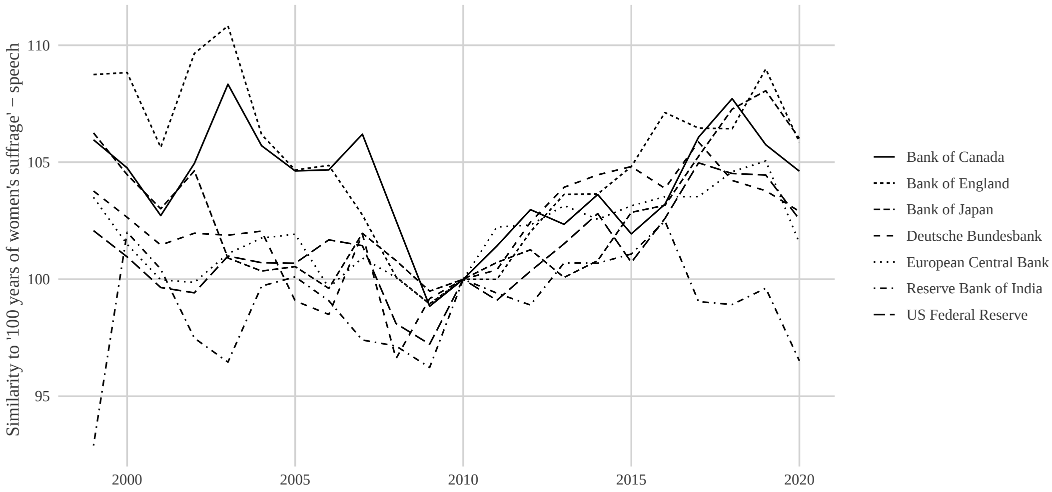
Notes: This graph illustrates the cosine distance between all ECB speeches and Sabine Lautenschläger's "100 years of women's suffrage - equality, freedom and democracy" speech. The cosine distance is defined in Equation (3.4). The solid lines illustrate the mean effect in the corresponding time window.

Adopting the approach from our last application, we filter speeches in our corpus according to gender-specific terms. We choose the speech *"100 years of women's suffrage - equality, freedom and democracy"* by Sabine Lautenschläger (Lautenschläger, 2017) and calculate the similarity of all speeches to this one. This allows us to create a time series index, which we call "gender focus" and which allows time-dependent evaluations.

Next, we identify four events in which the ECB announced new gender-focused policies for its institution in 2010, 2013, 2019, and 2021.[56] Figure 3.8 illustrates

[56]With respect to those dates: 2010: ECB issues an initial statement on the need to promote gender diversity and introduce measures to address gender imbalance. Hospido et al. (2019) findings indicate that gender gaps in promotion subsequently vanished following this announcement. 2013: The Executive Board introduces new gender balance targets (https://www.ecb.europa.eu/press/pr/date/2013/html/pr130829.en.html). 2019: Chris-

Figure 3.9: Similarity of central bank embeddings to Sabine Lautenschläger's speech.



Notes: This graph illustrates the cosine distance between a speech and Sabine Lautenschläger's speech. The cosine distance is defined in Equation (3.4).

the ECB's gender focus index in relation to these four events. There is an increasing trend between these periods. Using a Tukey (1949) test for honestly significant differences, we find that the period before 2010 is indeed significantly different from all following periods. We confirm this in Table B.9 in the Appendix. Note that the effect persists even after we control for the year, indicating that this is not purely a time-trend.

As a next step, we exploit the richness of our corpus to examine whether other central banks show a different trend in terms of gender focus. In Figure 3.9, we use the same index for the six largest banks in our corpus (besides the ECB) to illustrate their focus on gender equality over time, indexed to 2010. The figure provides several interesting observations: First, there does not seem to be a consistent trend prior to 2010; at times institutions were above and at times below 2010 levels in terms of gender focus. However, looking at the post-2010 period, only the Reserve Bank of India does not consistently exceed 2010 levels, indicating increased awareness, albeit with different trajectories. Second, both the 2008-09 financial crisis and the Covid pandemic appear to have lowered the

tine Lagarde succeeds outgoing President Mario Draghi on November 1, 2019, becoming the first woman to head the ECB. In addition, the first scientific findings are published that highlight the gender promotion gap in the institution prior to 2010 (Hospido et al., 2019).

focus on gender issues in general.

In a final step, we test these two assertions statistically. The results are presented in Table 3.11, with the ECB as the reference. There appear to be three interesting findings. First, we find a positive trend over time across all specifications. The trend is significant and economically relevant. Taking the first specification, we find an average change in the cosine distance of 10% over the last 20 years, indicating a substantial adjustment. Second, unsurprisingly, the cosine distance appears to be highest for the ECB speeches, with the exception being the Bundesbank. Third, as indicated in Figure 3.9, the regression suggests heterogeneity in the adoption of gender as an important issue across central banks over time, with the ECB having the strongest and the Reserve Bank of India the weakest adjustment. Nevertheless, once we control for macroeconomic variation (last column), a positive trend is present for all central banks (the year effect dominates the interaction effect). Finally, in contrast to what was previously suspected, economic recessions do not seem to affect the concentration on gender equality. However, macroeconomic factors not listed in the table have a substantial and significant impact on our index. The coefficients of these business cycle variables indicate a reduction in focus on tertiary objectives such as gender whenever the primary and secondary objectives diverge.[57]

In summary, this section provides preliminary evidence of the presence of gender bias in central bankers' language, as well as its (lack of) persistence.

3.6 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-conferences and reports. In this paper, we expanded the research frontier 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

[57]Increases in GDP (unemployment) are associated with a decrease (increase) in gender concentration, whereas changes in the price level appear to have no effect. We find further evidence for this hypothesis in unreported tests when substituting GDP with the output gap, which yields quantitatively and qualitatively similar results.

Table 3.11: Regression results - gender focus

	Dependent variable:			
	Gender focus index			
	(1)	(2)	(3)	(4)
Year	0.003***	0.002***	0.002***	0.007***
	(0.0002)	(0.0002)	(0.0002)	(0.001)
Reserve Bank of India		−0.081***	−0.081***	−0.540***
		(0.008)	(0.008)	(0.054)
US Federal Reserve		−0.090***	−0.090***	−0.043***
		(0.005)	(0.005)	(0.010)
Bank of Canada		−0.061***	−0.061***	−0.034***
		(0.007)	(0.007)	(0.009)
Bank of England		−0.017**	−0.017**	−0.012
		(0.007)	(0.007)	(0.008)
Bank of Japan		−0.078***	−0.078***	−0.073***
		(0.007)	(0.007)	(0.008)
Deutsche Bundesbank		0.051***	0.051***	0.079***
		(0.007)	(0.007)	(0.008)
Year x Reserve Bank of India		−0.001	−0.001	−0.002***
		(0.001)	(0.001)	(0.001)
Year x US Federal Reserve		−0.0003	−0.0003	−0.004***
		(0.0004)	(0.0004)	(0.001)
Year x Bank of Canada		−0.002***	−0.002***	−0.006***
		(0.001)	(0.001)	(0.001)
Year x Bank of England		−0.001***	−0.001***	−0.006***
		(0.001)	(0.001)	(0.001)
Year x Bank of Japan		−0.0002	−0.0002	−0.005***
		(0.001)	(0.001)	(0.001)
Year x Deutsche Bundesbank		−0.0001	−0.0001	−0.006***
		(0.001)	(0.001)	(0.001)
Recession			0.004	0.004
			(0.004)	(0.004)
Constant	0.057***	0.109***	0.109***	1.796***
	(0.002)	(0.004)	(0.004)	(0.194)
Macro-controls	no	no	no	yes
Observations	6,715	6,715	6,715	6,715
R^2	0.042	0.402	0.402	0.416
Adjusted R^2	0.042	0.401	0.401	0.414
Note:				*p<0.1; **p<0.05; ***p<0.01

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 text. 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 three prominent examples in the fields of central bank institutions, financial uncertainty, and gender bias. For example, we illustrate that our language model is able to approximate central bank objectives effectively. Throughout our applications, we emphasize several techniques for extracting the abundance of information contained within embeddings. We found that similarities — euclidean and cosine — are a suitable metric for integrating textual information into economic models, investigating them as dependent and independent variables. 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.

4 Financial markets and ECB monetary policy communication - A second QE surprise[*]

Martin Baumgärtner

Abstract

This paper shows that a different communication style of the ECB affects stock prices differently. A break in the ECB's communication from 2016 onwards makes it necessary to adjust the identification of monetary policy surprises in the euro area. By modifying the high-frequency identification of monetary policy shocks in the euro area, I can show that two quantitative easing shocks occur per decision: One during the release and one during the press conference. Although the impact on policy rates is identical, the release window shock seems to have a more pronounced effect on stock prices.

Keywords: Unconventional Monetary Policy, High-Frequency Data, ECB, Communication

JEL classification: E44, E52, E58

[*]The Paper is currently under review at the International Journal of Central Banking.

4.1 Introduction

Monetary policy surprises are typically measured during central bank decisions. Even though the measures taken by central banks are similar, the form in which they are announced differs. Therefore, a rich literature develops various methodologies to construct suitable surprises for different central banks. In this paper, I show, using the ECB as an example, that not only the differences between central banks are relevant for the identification of shocks, but that the identification must also be adjusted when the form of communication within a central bank changes.

Unlike other central banks, the ECB's monetary decisions are published in two steps. First, every six[58] weeks on Thursdays at 13:45, the decision is published in written form on the ECB's website and via news agencies. Until 2016, this *release* included only a brief statement on changes in the ECB's primary interest rates. In a second step, the *Press Conference* at 14:30, the measures taken will be explained by the president and journalists will be allowed to ask questions. In this window, the most unconventional measures were also announced. While the focus in the release window was on short-term interest rates, the second window mainly was about measures that had the longer end of the yield curve in view. However, since 2016 information on purchase programs or other supportive activities have also been integrated into the first written report so that there is no longer a clear separation of the two windows. So far, this change in detail has not yet been reflected in the literature.

In order to do so, I show in this paper that the change in OIS in the release window is driven by more than one significant latent factor. I replicate previous studies and show that financial market reactions change with new data. By adjusting the methodology, I demonstrate how this effect can be integrated, and the shocks can be correctly identified. It turns out that the structure of the additional shock is similar to the previously known quantitative easing surprise.

[58]Until 2015, the frequency of the meetings was every four weeks.

However, the effects depend on the window the shock occurs. In the release window, the reaction of stock prices is much more pronounced.

This paper builds on the literature based on Gürkaynak et al. (2005) to estimate monetary policy surprises using high-frequency data and factor models. The authors use a narrow time frame around Federal Open Market Committee releases to estimate two latent factors via factor decomposition. These *target* and *path* factors explain a large part of the variation in OIS. While the former puts the highest weight on the short term, the *path* factor impacts longer-term rates. The *path* factor was mostly associated with forward guidance, in which, later also partly LSAP fell. To separate the two effects, Swanson (2021) varies the rotation. He shows that for 1991 to 2019, not two factors have influence, but three. The author introduces the identification assumption that the influence of this factor should be minimal in the pre-QE period and creates a third (LSAP) factor, which is orthogonal to the previous two.

Due to the unique structure of the publication of ECB decisions, an adapted strategy is needed in the euro area. Brand et al. (2010) are the first to use the ECB structure to separate the shocks in detail. They find a target and path factor and a timing factor, interpreted as a kind of short-term forward guidance. The methodologies of Gürkaynak et al. (2005), Brand et al. (2010), and Swanson (2021) are combined into one framework by Altavilla et al. (2019). The authors construct a total of four shocks: a target shock from the release window and timing, forward guidance and QE shock for the press conference window. In addition, the authors provide the EA-MPD, which captures the changes in various financial market variables during the two windows.

The remainder of this paper is structured as follows. In Section 4.2 I show that the current identification in the literature omits meaningful central bank surprises in the release window and modify the identification. Then I evaluate the new resulting surprise in comparison to the literature in Section 4.3. The final section concludes this paper.

4.2 Methodology

For the years from 1999 to 2015, the press release following an ECB governing council meeting consisted of two sentences: One on future interest rates and a notice that the president would further explain these measures in the press conference. This changes as of December 2015, as the central bank, also comments on its LSAP. This announcement is limited to the raw facts. Neither program details are elaborated nor why the central bank considers the measures necessary.

The main question is, of course, how this effect can be measured and how financial market participants interpret this information. To investigate this, I use the EA-MPD published by Altavilla et al. (2019). The dataset contains the change in various financial market variables 30 minutes around both time windows for data for each ECB decision. The idea is that markets should have inserted all known information into the market price by then. Any new news, expansionary and restrictive, should therefore be measurable in this time window. Since OIS rates[59] have a strong link to central bank policy, I use them to measure surprises in financial markets concerning monetary policy. Per factor decomposition, the observed responses are attributed to several latent factors. Similar to previous studies, I consider the change of OIS rates with maturities of 1, 3, 6 months, 1, 2, 5, and 10 years. The data spans from January 2002 to March 2021, with 205 observations.

4.2.1 Number of relevant surprises

The first question is to examine how many factors are relevant and whether the change in central bank communication has altered this number. A common way to determine the number of factors (k) is to test the rank of the matrix using the method developed by Cragg and Donald (1997). The null hypothesis is that the matrix has the rank k.

For the release window, I use three different periods: first the window before the communication switchover (January 2002 to November 2015), second the window

[59] OIS allow for securing an interest rate linked to the Eonia in the future. Thus, the product has a direct link to central bank policy.

Table 4.1: Example of the changes in the release note

Meeting date	**22 October 2015**	**02 June 2016**
Interest rate	*At today's meeting, which was held in Malta, the Governing Council of the ECB decided that the interest rate on the main refinancing operations and the interest rates on the marginal lending facility and the deposit facility will remain unchanged at 0.05%, 0.30% and -0.20% respectively.*	*At today's meeting, which was held in Vienna, the Governing Council of the ECB decided that the interest rate on the main refinancing operations and the interest rates on the marginal lending facility and the deposit facility will remain unchanged at 0.00%, 0.25% and -0.40% respectively.*
QE		*Regarding non-standard monetary policy measures, on 8 June the Eurosystem will start making purchases under its corporate sector purchase programme (CSPP). Moreover, starting on 22 June, it will conduct the first operation in its new series of targeted longer-term refinancing operations. Further information on implementation aspects of the CSPP will be released after the press conference on the ECB's website.*
Press conference	*The President of the ECB will comment on the considerations underlying these decisions at a press conference starting at 14:30 CET today.*	*The President of the ECB will comment on the considerations underlying these decisions at a press conference starting at 14:30 CET today.*

after the switchover (December 2015 to March 2021), thus dividing the data set into a pre-change and post-change sample, and third the whole sample. Table 4.2 shows the results for all windows.

Table 4.2: Ranktest

	Release Window		
	Pre 2002-2015	Post 2016-2021	Full 2002-2021
$H_0 : k = 0$	46.26 (0.00)	46.51 (0.00)	53.90 (0.00)
$H_0 : k = 1$	18.94 (0.17)	25.85 (0.03)	25.77 (0.03)
$H_0 : k = 2$		14.95 (0.06)	13.54 (0.09)

Note: The table shows the Wald statistic of the Cragg and Donald (1997) rank test for the release window. The hypotheses $H_0 = k$ is evaluated against $H_0 < k$. The resulting p-values are in parentheses.

There is a clear difference between the samples. In the period up to 2015, the null hypothesis for one factor cannot be rejected. It follows, in agreement with Brand et al. (2010) and Altavilla et al. (2019), that one factor is relevant. However, the later period shows a structural break: the hypothesis $k = 1$ is rejected, assuming that two factors are relevant here. This is also evident in the full sample, which is influenced by two factors. Integrating the unconventional measures into the press release added information. Therefore, two factors are relevant.[60]

4.2.2 Factor Model

To adequately present the central bank's policy, it is necessary to adapt the identification of monetary policy surprises to this change in communication. In doing so, I adapt the dominant approach in the literature so that the modification can be easily interpreted and compared. Let us assume a factor model:

[60] For the conference window, I can replicate the results of Altavilla et al. (2019). Two and three factors are relevant here.

$$X = F\Lambda + \epsilon \tag{4.1}$$

where X contains the change in OIS rates, F is a corresponding matrix with the factors, and Λ is the loading matrix. After decomposition, the factors cannot be interpreted structurally since they usually influence each component X. To distinguish the factors from each other and to be able to interpret them, the factors must be rotated accordingly. Therefore, one rotates the model by introducing a matrix U, where $UU' = I$ and I corresponds to the identity matrix. This results in:

$$X = \bar{F}\bar{\Lambda} + \epsilon \tag{4.2}$$

with $\bar{F} = FU$ and $\bar{\Lambda} = U'\Lambda$. Introducing restrictions in U makes it possible to identify orthogonal factors. Furthermore, the monetary policy surprises can be identified by using the following restrictions for the release window:

1. The second factor does not load on the one-month OIS rates.
2. The second factor has a minimum variance before December 2015.

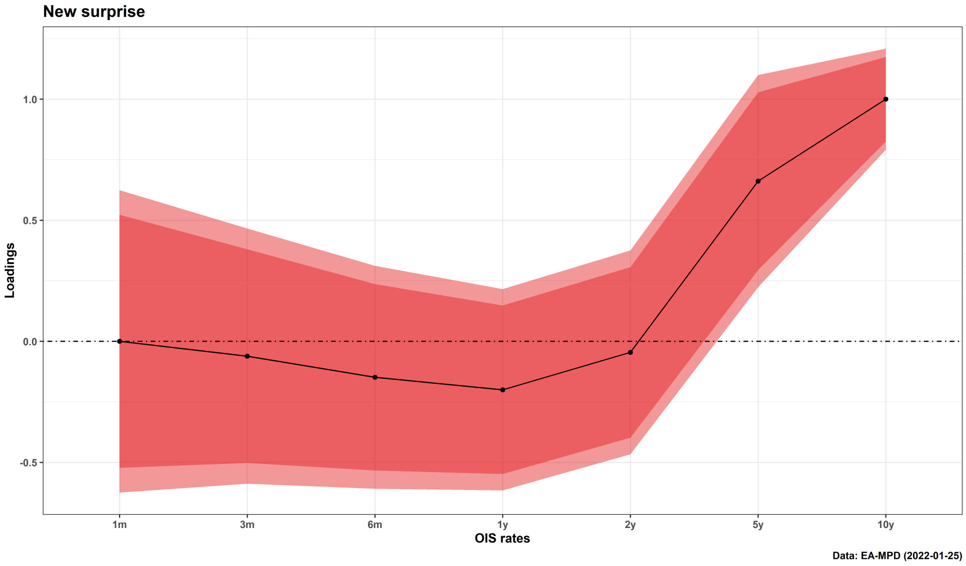
The first restriction is based on Gürkaynak et al. (2005) and separates the additional factors from the short-term oriented measures. The second restriction is inspired by the approach used by Swanson (2021) that if a factor does not exist, it should have a minimal variance.[61] Accordingly, before the change in press releases, the factor should have no impact.

Figure 4.1 shows the loading of the new second factor on the OIS rates.[62] The identified factor loads exclusively on the long term, five and ten-year swaps. Therefore, the new factor is reminiscent of a QE factor. This is also consistent with the expected outcome, as the new information in the releases concentrates

[61] Swanson (2021) thus identifies QE surprises. Since no QE shocks are expected before the financial crisis, the variance for this period should be minimal.

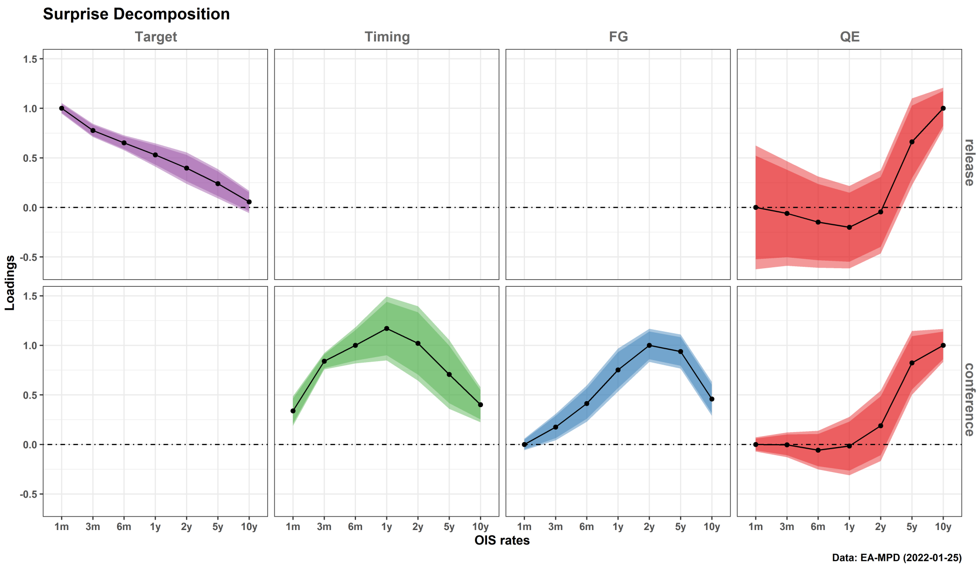
[62] It should be noted that, except for the one-month swap rate, the influences are not forced by restrictions but are estimated and thus can be interpreted.

Figure 4.1: Loading structure



Notes: This graph depicts the loadings of the new factor in the release window for different OIS horizons based on the rotation. The factor is scaled that it has an unit effect on the ten-year OIS rates. The shaded area shows the upper and lower bands of the 90% and 95% of the confidence intervals.

Figure 4.2: Loading structure (All surprises)



Notes: This graph depicts the loadings of all factors for different OIS horizons based on the rotation based on the EA-MPD. The first row shows the factors identified in the release window and the second row in the conference window. The target factor is scaled to have a unit effect on the one-month OIS rate, the timing factor on the six-month OIS rate, forward guidance on two years, and the QE surprises are normalized to ten-year OIS rates. The shaded areas show the upper and lower bands of the 90% and 95% of the confidence intervals.

on LSAP. Also, the comparison with the conference window shows strong similarities between the second factor in the release window and the third surprise in the conference window found by Altavilla et al. (2019) (See Figure 4.2).
So, starting in December 2015, there are two QE surprises per central bank meeting, one at the time of the ECB press release and one at the press conference. These are not different from each other in terms of the loading structure. Still, they are different in terms of the information they convey: While relatively little and condensed information is published in the release window, these measures are explained more during the press conference. In addition, it is possible for the public to ask questions and thus better understand the measures and the intention behind them.
The next step is to ask how the two factors differ in their effect. It would be possible to conclude which communication style is better suited to produce the desired result.

4.3 Results

In the process of an ECB announcement, based on the preceding analysis, two QE shocks occur, one in the conference window (starting in October 2014) and one in the release window (starting in December 2015). The differences between the two surprises in other high-frequency variables will be examined below. To ensure that the different starting times do not distort the results, I only consider the period starting December 2015. Both surprises are normalized to have a unit effect on the 10-year OIS rates[63] in the respective window. To study the impact of different windows and thus different communication of QE, I estimate the following equation:

$$\Delta x_{w,t} = QE_{w,t} + D_{release} + QE_{w,t} \times D_{release} + C_{w,t} + \epsilon_{w,t} \tag{4.3}$$

where Δx_t denotes the change of different financial variables during the monetary

[63] 10-year rates are most affected by a QE shock. (See Figure 4.2)

announcement t in window w. As variables for $\Delta x_{w,t}$ I use changes in OIS rates, the STOXX50 and the EURO STOXX Banks (SX7E) index[64] available in the EA-MPD.

Table 4.3: Regression results: OIS rates

	OIS_{6m}		OIS_{2y}		OIS_{10y}	
	(1)	(2)	(3)	(4)	(5)	(6)
QE	-0.08^{**}	-0.10^{***}	0.18^{**}	0.17^{**}	1.00^{***}	0.99^{***}
	(0.04)	(0.04)	(0.07)	(0.07)	(0.05)	(0.05)
$D_{release}$	0.35^{***}	0.33^{***}	0.14	0.16	-0.04	0.01
	(0.10)	(0.11)	(0.18)	(0.21)	(0.13)	(0.15)
$QE \times D_{release}$	0.04	0.05	-0.13	-0.13	0.00	0.01
	(0.06)	(0.06)	(0.11)	(0.11)	(0.08)	(0.08)
Target	0.78^{***}	0.76^{***}	0.61^{***}	0.58^{***}	0.26^{***}	0.23^{***}
	(0.03)	(0.03)	(0.06)	(0.06)	(0.04)	(0.05)
Timing	0.95^{***}	0.89^{***}	1.12^{***}	0.98^{***}	0.51^{***}	0.38^{***}
	(0.10)	(0.10)	(0.19)	(0.19)	(0.14)	(0.14)
FG	0.52^{***}	0.50^{***}	1.02^{***}	0.95^{***}	0.56^{***}	0.49^{***}
	(0.05)	(0.05)	(0.10)	(0.10)	(0.07)	(0.07)
jobless claims		-0.00		0.00		0.00
		(0.00)		(0.00)		(0.00)
$D_{Lagarde}$		0.31		0.27		0.16
		(0.31)		(0.59)		(0.43)
Constant	-0.22^{***}	0.80^{**}	-0.20	1.47^{**}	-0.13	1.32^{***}
	(0.07)	(0.34)	(0.13)	(0.65)	(0.10)	(0.47)
D_{years}	No	Yes	No	Yes	No	Yes
Adj. R^2	0.91	0.91	0.79	0.80	0.93	0.94
Num. obs.	86	86	86	86	86	86
F statistic	139.74	64.56	53.06	24.62	191.76	89.42

Regression of OIS intraday changes on monetary surprises per window. The odd model numbers show the basic model and the even ones extend it with various control variables: U.S. unemployment claims, and two dummies for the ECB president and years. Standard errors are displayed in parentheses. ***, **, * indicate significance at the 1, 5, and 10 per cent level, respectively.

$QE_{w,t}$ stands for the QE surprise at time t in window w and $D_{release}$ is a dummy which indicates whether it is the conference window (0) or the release window (1). $C_{w,t}$ includes several control variables: First, all other monetary surprises known from Altavilla et al. (2019), the weekly seasonally adjusted U.S. jobless claims published during the ECB press conference, and a dummy controlling for the ECB president. The estimation in one equation allows separating the effect

[64] The EURO STOXX Bank index focuses on banks and financial services providers from the STOXX 600.

of the release window on the intercept and the interaction of the dummy with the QE Surprise. If the two windows produce similar effects in the variables, then the *window* coefficient and the interaction term should be insignificant. A significant dummy by itself suggests a general difference between the two windows. On the other hand, a significant interaction term would show that QE has different effects depending on the relevant period. Otherwise, these elements would indicate how exactly the effect differs. The results can be found in Table 4.3 and (4.4).

For the variables most closely related to the central bank, OIS rates, there seems to be no deviating influence of the QE factor in the release window. Still, there are indications of a generally different effect between the windows. Looking at the coefficient of the six-month OIS rates for the release window, $D_{release}$, we find a significant positive effect. The reactions are stronger in the release window than during the press conference. Simultaneously, there is no evidence of a divergent effect between the QE shocks in the different windows. In general, positive (restrictive) surprises increase OIS rates. However, the effect varies according to the construction of the factors. Target surprises have the most substantial impact in the short term, and QE factors have the most decisive influence in the long term. This changes when looking at stock prices, STOXX 50 and SX7E. Target, timing and forward guidance show the expected signs in each specification but vary in significance. Especially the effects on the SX7E are less clear. Altavilla et al. (2019) find similar results in their evaluation. They attribute this to a possible existence of information shocks. That is, the central bank's interest rate decision also reveals information about the economy, which has the opposite effect on stock prices (Campbell et al., 2012; Miranda-Agrippino and Ricco, 2021). Other studies find similar effects, but point to other possible explanations, such as delayed information processing by financial market participants and uncertainty in the announcement (Bauer and Swanson, 2020; Baumgärtner, 2020). A complicated picture emerges for QE and $D_{release}$, the variables of interest. For the STOXX50, I find no effect different from zero but a significant negative interaction coefficient with the release window. Thus, in the release window, the effect of a QE shock is significantly stronger than in the press conference window. For SX7E, however, the QE coefficient is positive and thus contradicts economic intuition. The inter-

Table 4.4: Regression results: Stock prices

	STOXX50		SX7E	
	(1)	(2)	(3)	(4)
QE	0.03	0.05	0.35***	0.37***
	(0.05)	(0.05)	(0.11)	(0.11)
$D_{release}$	0.14	0.16	0.42	0.53
	(0.12)	(0.14)	(0.27)	(0.32)
$QE \times D_{release}$	−0.20***	−0.20***	−0.32*	−0.34*
	(0.07)	(0.08)	(0.16)	(0.17)
Target	−0.15***	−0.11**	−0.09	−0.02
	(0.04)	(0.04)	(0.09)	(0.10)
Timing	−0.23*	−0.14	−0.45	−0.37
	(0.13)	(0.13)	(0.28)	(0.30)
FG	−0.27***	−0.23***	−0.42***	−0.39**
	(0.06)	(0.07)	(0.14)	(0.15)
jobless claims		0.00		0.00
		(0.00)		(0.00)
$D_{Lagarde}$		0.17		0.74
		(0.40)		(0.92)
Constant	−0.07	−1.22***	−0.24	−1.36
	(0.09)	(0.44)	(0.20)	(1.00)
D_{years}	No	Yes	No	Yes
Adj. R^2	0.30	0.31	0.14	0.14
Num. obs.	86	86	86	86
F statistic	7.09	3.68	3.29	1.97

Regression of stock prices on monetary surprises per window. The odd model numbers show the basic model and the even ones extend it with various control variables: U.S. unemployment claims, and two dummies for the ECB president and years. Standard errors are displayed in parentheses. ***, **, * indicate significance at the 1, 5, and 10 per cent level, respectively.

action term has the appropriate negative sign and is significant at the 10% level, but even then, the sum of both coefficients would be zero, so there would be no clear relationship between QE and SX7E.

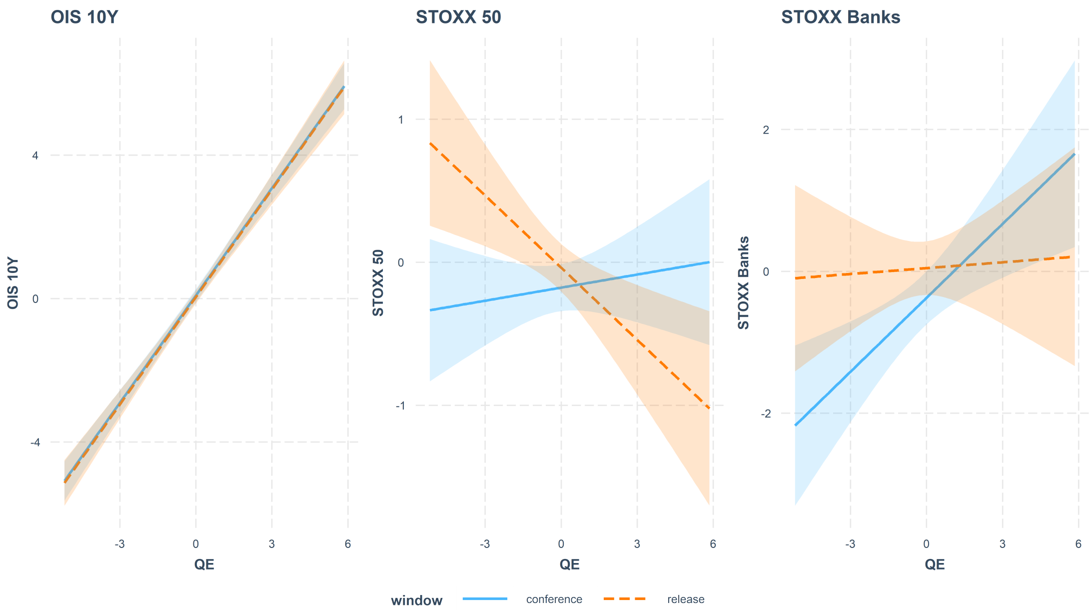
Figure 4.3 illustrates the observed effects by plotting the magnitude of the overall coefficient depending on the QE shock and the policy window. For the ten-year OIS rates[65], the interaction term does not play a role, so the two straight lines are almost synchronous. For the STOXX50, however, the difference becomes apparent. While a QE shock in the press conference window shows no significant correlation, the effect in the release window is pronounced. A restrictive (expansionary) QE shock lowers (raises) stock prices. The graph for SX7E shows that the direction of the effect in the press conference window is more similar to the OIS rates than the STOXX50. In the release window, the slope of the straight line becomes flatter, but still shows no negative correlation. The SX7E's reaction can presumably be explained by the fact that the index focuses on banks for which QE has a potentially negative side effect. An expansion of QE keeps the yield outlook in the interest rate environment lower for longer and, therefore, lowers the banks' stock prices.

Possible explanations for the difference between the two policy windows in the STOXX50 could be related to the content of the release and follow the findings of Smales and Apergis (2017a), Smales and Apergis (2017b), and Hayo, Henseler, et al. (2020). The authors show that press conferences have become more linguistically complex with the introduction of unconventional monetary policy, which leads to a change in trading activity during the press conference. This could explain the results: First, the two surprises differ in their content: While the release window is very focused on the actual central bank policy, the press conference explains the background and motivation of the central bank in much more detail. The possibility of follow-up questions requires the president to communicate quickly and consistently. If this does not succeed, it is conceivable that the central bank's signal will be more restrained compared to the release.

Second, the form of the release is initially different. The release window is always

[65]Other OIS rates show a very similar pattern, but have been omitted here for the sake of clarity.

Figure 4.3: Interaction effects



Notes: This graph illustrates the level of the coefficient for OIS_{10Y}, STOXX50 and STOXX Banks from Table 4.4 for the release window (orange) and the press conference window (blue). Overlapping lines indicate no interaction effect whereas crossed lines indicate a relevant interaction term. The shaded areas show the upper and lower bands of the 95% of the confidence intervals.

in text form, whether through the central bank's website or news outlets. On the other hand, the press conference is initially only audio-visual, i.e., a video stream. This can mean that it becomes more challenging to process the incoming information, as the amount of information that can be evaluated increases[66], and at the same time, it becomes technically more demanding to evaluate the content, as automation solutions are usually based on plain text.

4.4 Conclusion

This paper sharpens the identification of central bank shocks in the euro area. The analysis of high-frequency data during the ECB release shows that a new relevant factor appears in the data with the integration of QE in the ECB press release. This demonstrates that when central bank communication changes, the identification of monetary policy surprises also needs to be examined.

[66]Thus, in addition to the language of the central bank, there are attempts to include the appearance of central bankers during the pronouncement in the analysis. (Gorodnichenko et al., 2021)

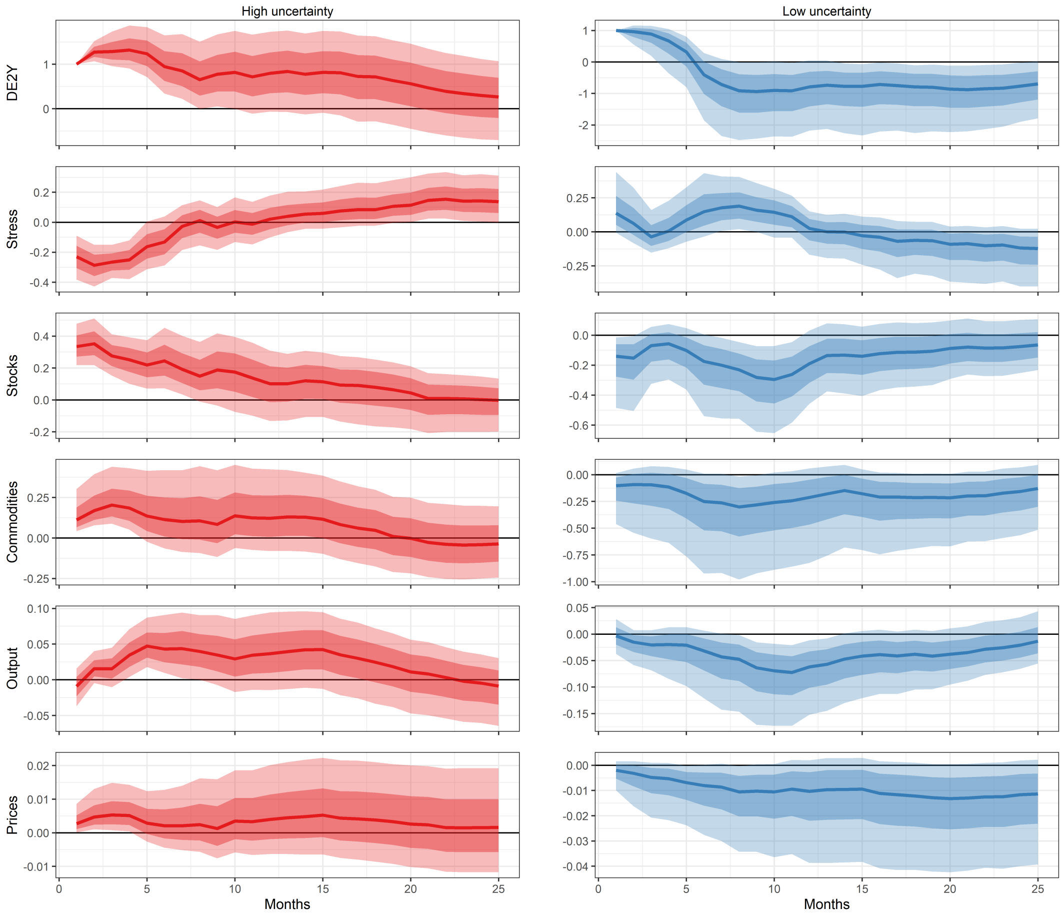
My analysis shows that this additional factor is indeed a QE shock. The effects on the different OIS maturities are almost identical, so comparing the two shocks in their impact is possible. There is a significantly different effect between the release and the press conference window for stock prices. One explanation for this reaction would be that compressed information from the central bank is more easily captured by financial markets, thus generating less uncertainty.

Although future research should focus on the specific link between complexity and stock prices, an important policy conclusion can be drawn. In addition to choosing the right policy instrument, central banks should pay more attention to how they announce them. Short and clear texts have a more substantial effect than more complex press conferences.

A Appendix: Information shocks and Uncertainty

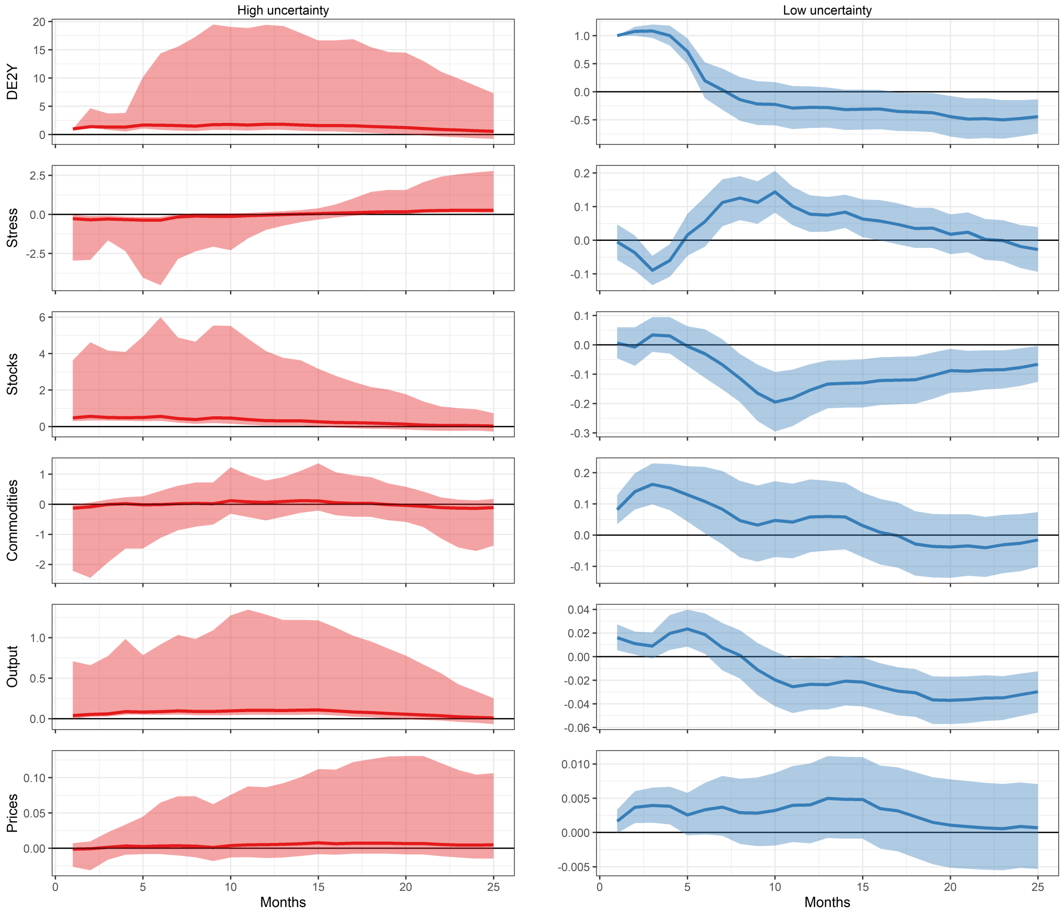
A.1 VAR results with Baekart et al. 2021 UC component

Figure A.1: Impulse responses of timing shocks decomposed by UC component



Notes: The shaded areas show the upper and lower bands of the 68% and 90% of the confidence intervals. The intervals shown are robust for weak instruments (Montiel Olea et al., 2020).

Figure A.2: Impulse responses of forward guidance shocks decomposed by UC component

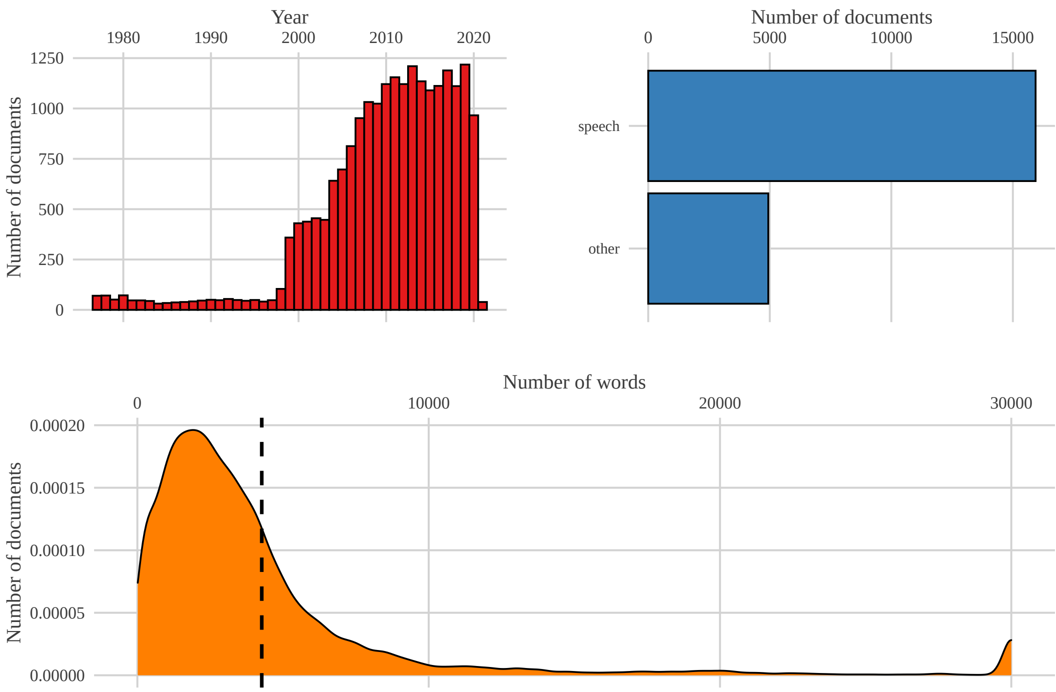


Notes: The shaded areas show the upper and lower bands of the 68% of the confidence intervals. The intervals shown are robust for weak instruments (Montiel Olea et al., 2020). The Wald statistic for the covariance between the insturment and shock variable is too small for the shock under high uncertainty to compute valid weak-IV robust confidence set at the 0.9% level, which is why these are not shown.

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

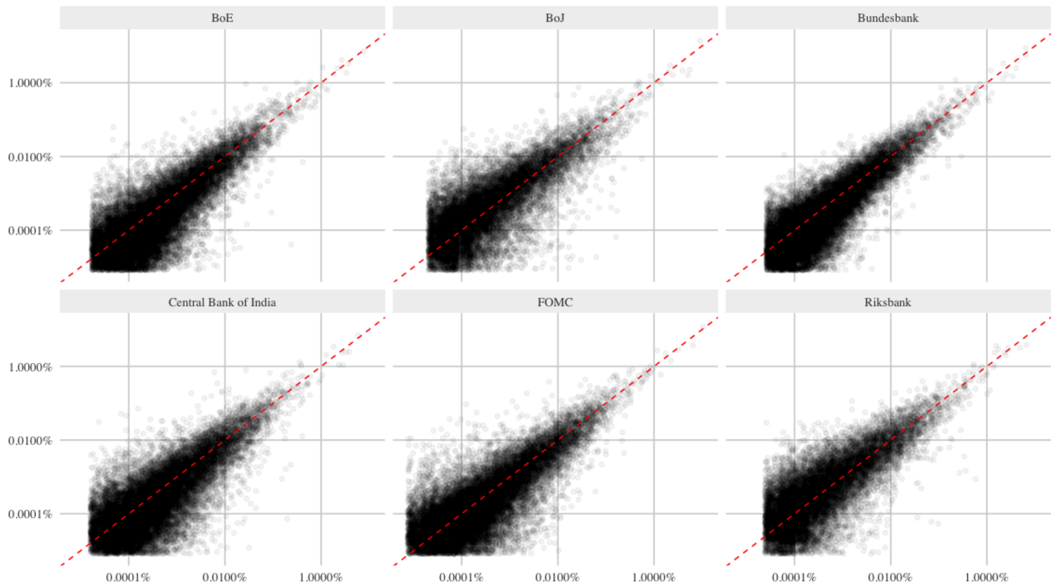
B.1 Graphical illustrations of text corpus

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



B.2 Language Model specifications

We use the hyperparameters for our models. 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 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).[67] The hyperparameters are summarized in the following table:

[67]For the Gibbs sampling draws we chose a burn-in rate of 1000, sampled 2000 iterations and returned every fifth iteration.

Table B.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

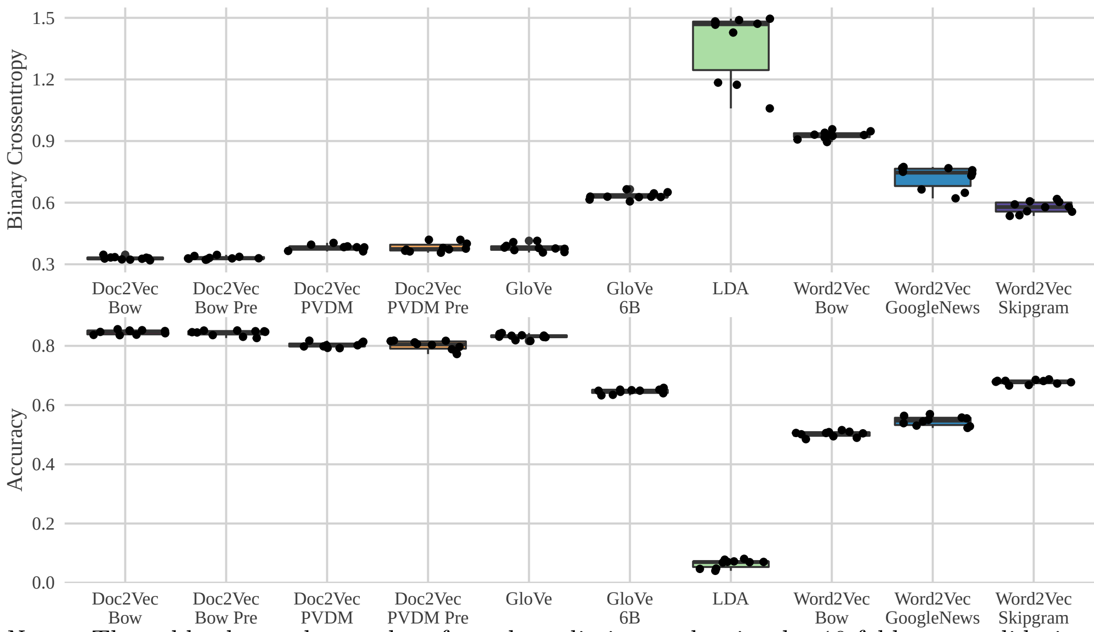
B.3 Additional evaluation

B.3.1 External evaluation I

The neural network is based on the Word2Vec skip-gram algorithm. Starting from a central word, the model is to predict the context, the surrounding words. We use a neural network with two embedding matrices. The first is the (word) embedding matrix of the language models mentioned in Figure 3.2. The second matrix, which represents the context, is first randomly initalized. Both elements are combined using the dot product and a sigmoid layer. We predict which other word is most likely to be nearby for each word. The critical difference to the Word2Vec skip-gram structure is that the first matrix is kept constant throughout training. This ensures that the word embeddings are evaluated even after training rather than an adapted version. We simulate out-of-sample prediction using 10-fold cross validation to ensure a fair comparison between embeddings. Each model is, in each fold, first trained on 90% of the observations.[68] Then, the performance is checked using the remaining 10%. The average overall ten out-of-sample predictions are used as a benchmark for evaluating our embeddings. Figure (B.3) shows the detailed results for the individual folds per model. The mean values of the folds correspond to the values in Table 3.3.

[68] We train each model with a window size of 1 and 10 negative examples. During backpropagation, the weights of the target matrix are not adjusted. In total, each model is initially trained for 20 epochs.

Figure B.3: Evaluation Results Word Prediction



Notes: The table shows the results of word prediction evaluation by 10-fold cross-validation. Each point corresponds to one test result. The boxplots summarize these results per evaluated model. The measurement on the y-axis is binary cross-entropy and accuracy. For the former, low values indicate good performance and for the latter, high values.

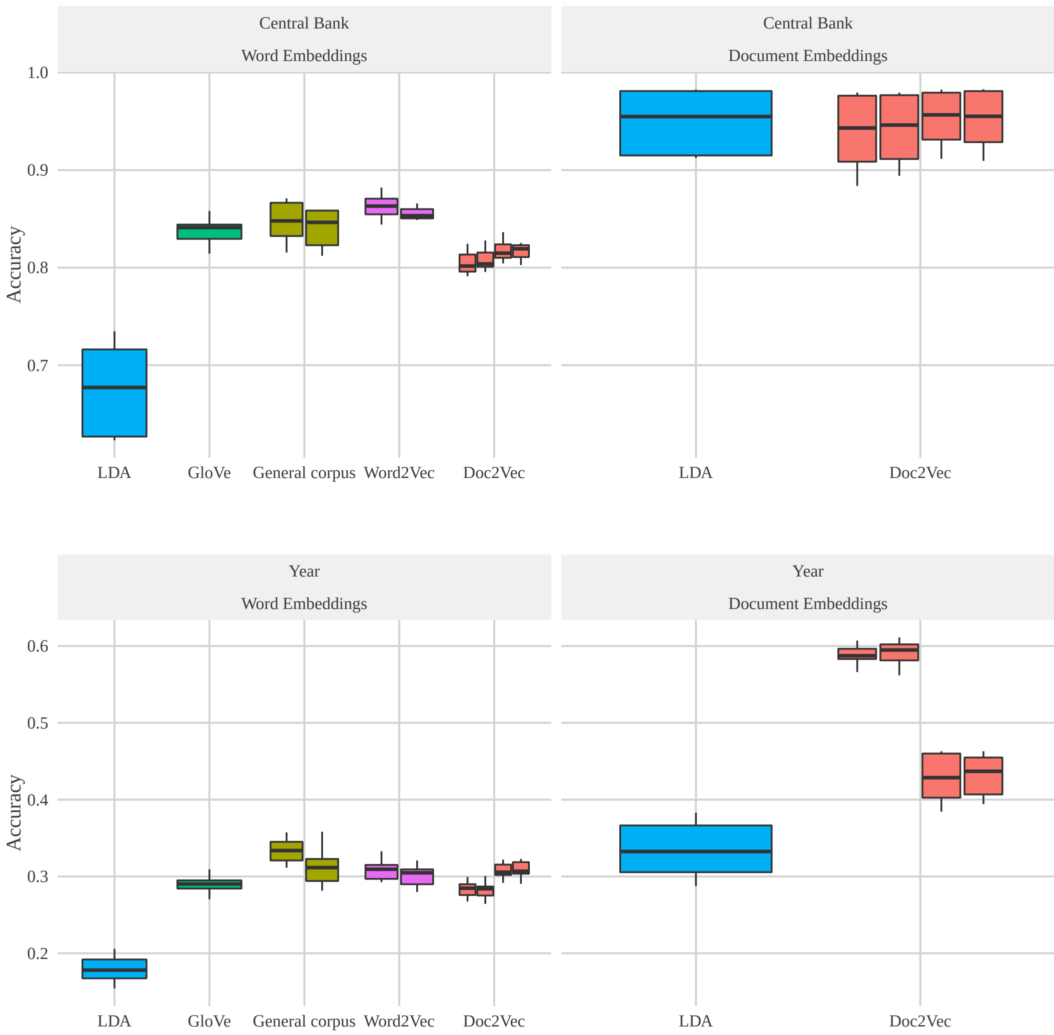
B.3.2 External evaluation II

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 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 B.4, 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 B.4: 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 $(true\ positive + true\ negative)/(number\ of\ observation)$.

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

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

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 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.

B.3.3 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 B.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 (3.4). The underscore is used to highlight collocations as described in Section 3.3.1.

B.3.4 Applications - Robustness checks

Table B.3: Regression results: Monetary Policy Framework classification (Word Embeddings)

	Similarity								
	New Zealand			Federal Reserve			European Central Bank		
	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
ITs	0.005***	0.004***		0.004***	0.003***		0.004***	0.01***	
	(0.001)	(0.001)		(0.001)	(0.001)		(0.001)	(0.001)	
- FIT			0.01***			0.004***			0.01***
			(0.001)			(0.001)			(0.001)
- LIT			0.004***			0.004***			0.01***
			(0.001)			(0.001)			(0.001)
- FCIT			0.002			0.001			0.005***
			(0.001)			(0.001)			(0.001)
- LCIT			0.003***			0.002**			0.004***
			(0.001)			(0.001)			(0.001)
ERTs		0.002**	0.003***		0.001	0.002**		0.003***	0.003***
		(0.001)	(0.001)		(0.001)	(0.001)		(0.001)	(0.001)
LSD		0.002**	0.003***		0.002**	0.002***		0.003***	0.003***
		(0.001)	(0.001)		(0.001)	(0.001)		(0.001)	(0.001)
MixedTs		0.001	0.002		−0.001	−0.0003		0.004***	0.004***
		(0.002)	(0.002)		(0.001)	(0.001)		(0.001)	(0.001)
nonat		−0.01***	−0.01***		−0.01**	−0.01**		0.001	0.001
		(0.003)	(0.003)		(0.003)	(0.003)		(0.003)	(0.003)
WSD		0.002*	0.003**		0.002	0.002*		0.004***	0.004***
		(0.001)	(0.001)		(0.001)	(0.001)		(0.001)	(0.001)
Constant	0.99***	0.99***	0.99***	0.99***	0.99***	0.99***	0.99***	0.99***	0.99***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Macro-controls	yes	yes	yes	yes	yes	yes	yes	yes	yes
Observations	84	811	811	84	811	811	83	811	811
R^2	0.21	0.25	0.27	0.20	0.21	0.23	0.21	0.26	0.26
Adjusted R^2	0.17	0.24	0.26	0.16	0.20	0.22	0.17	0.25	0.25
Note:									*p<0.1; **p<0.05; ***p<0.01

Monetary policy framework classification

Table B.4: Monetary Policy Framework classification: New Zealand results

	Similarity to New Zealand							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ITs	0.06***	0.07***	0.11***	0.11***				
	(0.02)	(0.02)	(0.02)	(0.03)				
- FIT					0.15***	0.17***	0.11***	0.15***
					(0.02)	(0.03)	(0.02)	(0.04)
- LIT					0.10***	0.11***	0.05***	0.16***
					(0.02)	(0.03)	(0.02)	(0.04)
- FCIT					0.04	0.05		0.06*
					(0.04)	(0.04)		(0.04)
- LCIT					0.03	0.06*	−0.01	
					(0.03)	(0.03)	(0.02)	
ERTs			0.04	0.04	0.04	0.06**		0.08*
			(0.03)	(0.03)	(0.03)	(0.03)		(0.04)
LSD			0.03	0.05*	0.03	0.06**	0.004	0.06**
			(0.02)	(0.03)	(0.02)	(0.03)	(0.02)	(0.03)
MixedTs			0.07	0.06	0.07	0.09*	0.02	
			(0.05)	(0.05)	(0.05)	(0.05)	(0.04)	
nonat			−0.07	−0.06	−0.07	−0.04		−0.03
			(0.10)	(0.10)	(0.09)	(0.09)		(0.08)
WSD			0.11***	0.12***	0.11***	0.14***	0.08**	
			(0.04)	(0.04)	(0.03)	(0.04)	(0.03)	
Constant	0.33***	0.32***	0.31***	0.31***	0.31***	0.29***	0.35***	0.26***
	(0.01)	(0.02)	(0.02)	(0.03)	(0.02)	(0.03)	(0.02)	(0.03)
Macro-controls	no	yes	no	yes	no	yes	yes	yes
Countries	all	all	all	all	all	all	initial	rest
Observations	84	84	821	821	821	821	649	172
R^2	0.11	0.15	0.15	0.18	0.22	0.24	0.18	0.15
Adjusted R^2	0.09	0.11	0.15	0.17	0.21	0.23	0.17	0.10

Note: *p<0.1; **p<0.05; ***p<0.01

Table B.5: Monetary Policy Framework classification: US results

	Similarity to United States							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ITs	0.06***	0.06***	0.12***	0.09***				
	(0.01)	(0.02)	(0.02)	(0.02)				
- FIT					0.14***	0.12***	0.11***	0.15***
					(0.02)	(0.02)	(0.01)	(0.04)
- LIT					0.12***	0.10***	0.09***	0.09**
					(0.02)	(0.02)	(0.01)	(0.04)
- FCIT					0.10***	0.10***		0.10***
					(0.04)	(0.04)		(0.04)
- LCIT					0.03	0.02	0.01	
					(0.03)	(0.03)	(0.02)	
LSD			0.05**	0.04*	0.05**	0.04*	0.06***	0.02
			(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)
ERTs			0.03	0.001	0.03	0.01		0.01
			(0.02)	(0.03)	(0.02)	(0.03)		(0.04)
MixedTs			0.06	0.03	0.06	0.04	0.04	
			(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	
NNF			−0.11	−0.12	−0.11	−0.12		−0.12
			(0.09)	(0.09)	(0.08)	(0.08)		(0.08)
WSD			0.13***	0.10***	0.13***	0.11***	0.10***	
			(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	
Constant	0.32***	0.32***	0.40***	0.43***	0.40***	0.42***	0.43***	0.41***
	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.03)
Macro-controls	no	yes	no	yes	no	yes	nyeso	yes
Countries	all	all	all	all	all	all	initial	rest
Observations	84	84	825	825	825	825	653	172
R^2	0.19	0.20	0.17	0.19	0.22	0.22	0.15	0.21
Adjusted R^2	0.18	0.16	0.17	0.18	0.21	0.21	0.14	0.17

Note: *p<0.1; **p<0.05; ***p<0.01

Table B.6: Monetary Policy Framework classification: Euro Area results

	Similarity to Euro area							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ITs	0.10***	0.09***	0.15***	0.16***				
	(0.02)	(0.02)	(0.02)	(0.03)				
- FIT					0.12***	0.12***	0.03	0.22***
					(0.02)	(0.03)	(0.02)	(0.04)
- LIT					0.18***	0.18***	0.09***	0.08*
					(0.02)	(0.02)	(0.02)	(0.04)
- FCIT					0.13***	0.14***		0.13***
					(0.04)	(0.04)		(0.03)
- LCIT					0.06**	0.09***	0.005	
					(0.03)	(0.03)	(0.02)	
LSD			0.04	0.08***	0.04	0.06***	−0.02	0.03
			(0.02)	(0.03)	(0.02)	(0.02)	(0.02)	(0.03)
ERTs			0.06**	0.09***	0.06**	0.08***		0.005
			(0.03)	(0.03)	(0.03)	(0.03)		(0.04)
MixedTs			0.13***	0.14***	0.13***	0.13***	0.03	
			(0.05)	(0.05)	(0.04)	(0.05)	(0.04)	
NNF			−0.03	0.01	−0.03	−0.003		−0.04
			(0.10)	(0.09)	(0.09)	(0.09)		(0.08)
WSD			0.13***	0.17***	0.13***	0.15***	0.07**	
			(0.04)	(0.04)	(0.03)	(0.04)	(0.03)	
Constant	0.38***	0.40***	0.41***	0.41***	0.41***	0.41***	0.51***	0.42***
	(0.01)	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)	(0.02)	(0.03)
Macro-controls	no	yes	no	yes	no	yes	nyeso	yes
Countries	all	all	all	all	all	all	initial	rest
Observations	83	83	821	821	821	821	649	172
R^2	0.21	0.28	0.26	0.31	0.34	0.37	0.32	0.33
Adjusted R^2	0.20	0.24	0.26	0.30	0.33	0.36	0.31	0.29
Note:						*p<0.1; **p<0.05; ***p<0.01		

Table B.7: Document Embedding - Monetary Degree - on yearly basis

	Dependent variable:					
	similarity					
	NZ		US		EA	
	(1)	(2)	(3)	(4)	(5)	(6)
monetary_policy_degreeintens	0.2**	0.2*	0.2***	0.2**	0.1*	0.1
	(0.1)	(0.1)	(0.1)	(0.1)	(0.1)	(0.1)
monetary_policy_degreeinter	0.1	0.03	0.1	0.1	0.03	0.01
	(0.1)	(0.1)	(0.1)	(0.1)	(0.1)	(0.1)
monetary_policy_degreesubst	0.1	0.1	0.2**	0.2**	0.1	0.1
	(0.1)	(0.1)	(0.1)	(0.1)	(0.1)	(0.1)
ea_membermember	0.03***	−0.01	0.1***	0.1***	0.2***	0.1***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
inflation_dif_nz		0.000				
		(0.001)				
unemployment_rate_dif_nz		0.003***				
		(0.001)				
gdp_dif_nz		0.02***				
		(0.004)				
inflation_dif_us				−0.001		
				(0.001)		
unemployment_rate_dif_us				−0.000		
				(0.001)		
gdp_dif_us				0.01**		
				(0.003)		
inflation_dif_ea						−0.001**
						(0.001)
unemployment_rate_dif_ea						0.001**
						(0.001)
gdp_dif_ea						0.01***
						(0.003)
Constant	0.2**	0.3***	0.3***	0.3***	0.4***	0.4***
	(0.1)	(0.1)	(0.1)	(0.1)	(0.1)	(0.1)
Observations	821	821	825	825	821	821
R^2	0.1	0.2	0.1	0.1	0.4	0.4
Adjusted R^2	0.1	0.2	0.1	0.1	0.4	0.4
Note:				*p<0.1; **p<0.05; ***p<0.01		

Whatever it takes

Table B.8 shows table 3.7 with all control variables.

Table B.8: Application 2: Whatever it takes - Full table

	Dependent variable: $\Delta\text{spread}_{10y}$		
$Unc_t =$	$VSTOXX_{pd,t}$	$CISS_{pd}$	UC_{pd}
$\text{wit}_{\text{simil}}$	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)
L(sp500,1)	−0.0001 (0.0001)	−0.0001 (0.0001)	−0.0001 (0.0001)
L(stoxx,1)	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)
$\text{L}(\Delta spread_{10y}, 1)$	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)
R^2	0.116	0.113	0.116
F Statistic	10.529***	10.153***	10.101***

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 B.9: Regression results - gender focus

	Dependent variable:	
	gender_focus	
	(1)	(2)
2010 - ECB diversity strategy	0.01***	0.01**
	(0.004)	(0.01)
2013 - ECB employment	0.03***	0.04***
	(0.004)	(0.01)
2019 - ECB women scholarship	0.02***	0.03**
	(0.01)	(0.01)
Year		−0.000
		(0.001)
Constant	0.1***	0.1***
	(0.003)	(0.01)
Observations	2,183	2,183
R^2	0.04	0.04
Adjusted R^2	0.04	0.04
Note:	*p<0.1; **p<0.05; ***p<0.01	

References

Aastveit, Knut, Gisle James James Natvik, and Sergio Sola (2013). *Economic Uncertainty and the Effectiveness of Monetary Policy*. SSRN Scholarly Paper ID 2353008. Rochester, NY: Social Science Research Network. DOI: 10.2139/ssrn.2353008.

Acosta, Miguel and Ellen E Meade (2015). *Hanging on every word: Semantic analysis of the FOMC's postmeeting statement*. Tech. rep., p. 30. DOI: 10.17016/2380-7172.1580.

Altavilla, Carlo, Luca Brugnolini, Refet S. Gürkaynak, Roberto Motto, and Giuseppe Ragusa (2019). "Measuring Euro Area Monetary Policy". en. In: *Journal of Monetary Economics* 108, pp. 162–179. ISSN: 0304-3932. DOI: 10.1016/j.jmoneco.2019.08.016.

Amaya, Diego and Jean-Yves Filbien (2015). "The Similarity of ECB's Communication". In: *Finance Research Letters* 13, pp. 234–242. DOI: 10.1016/j.frl.2014.12.006.

Andrade, Philippe and Filippo Ferroni (2021). "Delphic and odyssean monetary policy shocks: Evidence from the euro area". In: *Journal of Monetary Economics* 117, pp. 816–832. ISSN: 0304-3932. DOI: 10.1016/j.jmoneco.2020.06.002.

Apel, Mikael, Marianna Blix Grimaldi, and Isaiah Hull (2019). "How much information do monetary policy committees disclose? Evidence from the FOMC's minutes and transcripts". In: *Journal of Money, Credit and Banking*. DOI: 10.1111/jmcb.12885.

Apel, Mikael and Marianna Blix Grimaldi (2014). "How informative are central bank minutes?" In: *Review of Economics* 65.1, pp. 53–76. DOI: 10.1515/roe-2014-0104.

Arias, Jonas E., Juan F. Rubio-Ramírez, and Daniel F. Waggoner (2018). "Inference Based on Structural Vector Autoregressions Identified With Sign and Zero Restrictions: Theory and Applications". In: *Econometrica* 86.2, pp. 685–720. ISSN: 1468-0262. DOI: 10.3982/ECTA14468.

Athey, Susan (2019). "21. The Impact of Machine Learning on Economics". In: *The Economics of Artificial Intelligence: An Agenda*. Ed. by Ajay Agrawal, Joshua Gans, and Avi Goldfarb. University of Chicago Press, pp. 507–552. DOI: 10.7208/9780226613475-023.

Auriol, Emmanuelle, Guido Friebel, and Sascha Wilhelm (2020). "Women in European economics". In: *Women in Economics*. Ed. by Shelly Lundberg. CEPR Press, p. 26. URL: https://voxeu.org/system/files/epublication/Women_in_Economics_.pdf.

Azqueta-Gavaldon, Andres, Dominik Hirschbühl, Luca Onorante, and Lorena Saiz (2020). *Economic policy uncertainty in the euro area: an unsupervised machine learning approach*. Tech. rep. 2359. European Central Bank. URL: https://ideas.repec.org/p/ecb/ecbwps/20202359.html.

Badilla, Pablo, Felipe Bravo-Marquez, and Jorge Pérez (2020). "WEFE: The Word Embeddings Fairness Evaluation Framework". In: *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence*. International Joint Conferences on Artificial Intelligence Organization, pp. 430–436. ISBN: 978-0-9992411-6-5. DOI: 10.24963/ijcai.2020/60.

Baker, Scott R., Nicholas Bloom, and Steven J. Davis (2016). "Measuring Economic Policy Uncertainty". In: *The Quarterly Journal of Economics* 131.4, pp. 1593–1636. ISSN: 0033-5533. DOI: 10.1093/qje/qjw024.

Bauer, Daniel J. and Patrick J. Curran (2005). "Probing Interactions in Fixed and Multilevel Regression: Inferential and Graphical Techniques". In: *Multivariate Behavioral Research* 40.3, pp. 373–400. ISSN: 0027-3171. DOI: 10.1207/s15327906mbr4003_5.

Bauer, Michael D and Eric T. Swanson (2020). *An Alternative Explanation for the "Fed Information Effect"*. Tech. rep. DOI: 10.3386/w27013.

Baumgärtner, Martin (2020). *Information shocks and Uncertainty*. Tech. rep. DOI: 10.13140/RG.2.2.22504.19201.

Baumgärtner, Martin and Jens Klose (2021). "Why central banks announcing liquidity injections is more effective than forward guidance". In: *International Finance* 24.2, pp. 236–256. DOI: 10.1111/infi.12389.

Bekaert, Geert, Marie Hoerova, and Marco Lo Duca (2013). "Risk, uncertainty and monetary policy". In: *Journal of Monetary Economics* 60.7, pp. 771–788. ISSN: 0304-3932. DOI: 10.1016/j.jmoneco.2013.06.003.

Bekaert, Geert, Marie Hoerova, and Nancy R. Xu (2021). *Risk, Monetary Policy and Asset Prices in a Global World*. SSRN Scholarly Paper ID 3599583. Rochester, NY: Social Science Research Network. DOI: 10.2139/ssrn.3599583. URL: https://papers.ssrn.com/abstract=3599583.

Bengio, Yoshua, Réjean Ducharme, Pascal Vincent, and Christian Janvin (2003). "A neural probabilistic language model". In: *The Journal of Machine Learning Research* 3, pp. 1137–1155. DOI: 10.5555/944919.944966.

Bernanke, Ben S. and Kenneth N. Kuttner (2005). "What Explains the Stock Market's Reaction to Federal Reserve Policy?" In: *The Journal of Finance* 60.3, pp. 1221–1257. ISSN: 1540-6261. DOI: 10.1111/j.1540-6261.2005.00760.x.

Bholat, David, Stephen Hans, Pedro Santos, and Cheryl Schonhardt-Bailey (2015). *Text mining for central banks*. Handbooks 33. Centre for Central Banking Studies, Bank of England. URL: https://ideas.repec.org/b/ccb/hbooks/33.html.

Binette, André and Dmitri Tchebotarev (2019). *Canada's Monetary Policy Report: If Text Could Speak, What Would It Say?* Staff Analytical Notes 2019-5. Bank of Canada. URL: https://ideas.repec.org/p/bca/bocsan/19-5.html.

Blaheta, Don and Mark Johnson (2001). “Unsupervised learning of multi-word verbs”. In: *Association for Computational Linguistics Workshop on Collocation (2001)*. Association for Computational Linguistics (ACL), pp. 54–60. URL: http://web.science.mq.edu.au/~mjohnson/papers/2001/dpb-colloc01.pdf.

Blei, David M. and John D. Lafferty (2009). “Topic models”. In: *Text mining*. Chapman and Hall/CRC, pp. 101–124.

Blei, David M., Andrew Y. Ng, and Michael I. Jordan (2003). “Latent dirichlet allocation”. In: *The Journal of Machine Learning Research* 3, pp. 993–1022. URL: https://www.jmlr.org/papers/volume3/blei03a/blei03a.pdf.

Blinder, Alan S., Michael Ehrmann, Marcel Fratzscher, Jakob De Haan, and David-Jan Jansen (2008). “Central bank communication and monetary policy: A survey of theory and evidence”. In: *Journal of Economic Literature* 46.4, pp. 910–45. DOI: 10.1257/jel.46.4.910.

Bloom, Nicholas (2009). “The Impact of Uncertainty Shocks”. In: *Econometrica* 77.3, pp. 623–685. ISSN: 1468-0262. DOI: 10.3982/ECTA6248.

Boeckx, Jef, Maarten Dossche, Alessandro Galesi, Boris Hofmann, and Gert Peersman (2019). *Do SVARs with sign restrictions not identify unconventional monetary policy shocks?* Tech. rep. 1926. Banco de España. URL: https://ideas.repec.org/p/bde/wpaper/1926.html.

Boeckx, Jef, Maarten Dossche, and Gert Peersman (2017). “Effectiveness and Transmission of the ECB’s Balance Sheet Policies”. In: *International Journal of Central Banking* 13.1, pp. 297–333. URL: https://ideas.repec.org/a/ijc/ijcjou/y2017q0a8.html.

Bonam, Dennis, Gabriele Galati, Irma Hindrayanto, Marco Hoeberichts, Anna Samarina, and Irina Stanga (2019). *Inflation in the euro area since the Global Financial Crisis*. 1703. Netherlands Central Bank, Research Department. URL: https://ideas.repec.org/p/dnb/dnbocs/1703.html.

Borrallo Egea, Fructuoso and Luis Ángel Hierro (2019). “Transmission of monetary policy in the US and EU in times of expansion and crisis”. en. In: *Journal of Policy Modeling* 41.4, pp. 763–783. ISSN: 0161-8938. DOI: 10.1016/j.jpolmod.2019.02.012.

Brand, Claus, Daniel Buncic, and Jarkko Turunen (2010). “The Impact of ECB Monetary Policy Decisions and Communication on the Yield Curve”. In: *Journal of the European Economic Association* 8.6, pp. 1266–1298. DOI: 10.1111/j.1542-4774.2010.tb00555.x.

Burriel, Pablo and Alessandro Galesi (2018). “Uncovering the heterogeneous effects of ECB unconventional monetary policies across euro area countries”. In: *European Economic Review* 101 (C), pp. 210–229. DOI: 10.1016/j.euroecorev.2017.10.007.

Caliskan, Aylin, Joanna J. Bryson, and Arvind Narayanan (2017). “Semantics derived automatically from language corpora contain human-like biases”. In: *Science* 356.6334, pp. 183–186. ISSN: 0036-8075. DOI: 10.1126/science.aal4230.

Campbell, Jeffrey R., Charles L. Evans, Jonas D. M. Fisher, and Alejandro Justiniano (2012). “Macroeconomic Effects of Federal Reserve Forward Guidance”. In: *Brookings Papers on Economic Activity* 43.1, pp. 1–80. URL: https://ideas.repec.org/a/bin/bpeajo/v43y2012i2012-01p1-80.html.

Carriero, Andrea, Todd E. Clark, and Massimiliano Marcellino (2018). *Endogenous Uncertainty*. Tech. rep. 1805. Federal Reserve Bank of Cleveland. URL: https://ideas.repec.org/p/fip/fedcwp/1805.html.

Castelnuovo, Efrem and Giovanni Pellegrino (2018). “Uncertainty-dependent effects of monetary policy shocks: A new-Keynesian interpretation”. In: *Journal of Economic Dynamics and Control*. Monetary and Fiscal Policy Stabilization amid a Debt Crisis 93, pp. 277–296. ISSN: 0165-1889. DOI: 10.1016/j.jedc.2018.01.034.

Chakraborty, Chiranjit and Andreas Joseph (2017). *Machine learning at central banks*. Tech. rep. Bank of England. URL: https://www.bankofengland.co.uk/-/media/boe/files/working-paper/2017/machine-learning-at-central-banks.pdf?la=en&hash=EF5C4AC6E7D7BDC1D68A4BD865EEF3D7EE5D7806.

Cieslak, Anna and Andreas Schrimpf (2019). “Non-monetary news in central bank communication”. In: *Journal of International Economics* 118. ISSN: 0022-1996. DOI: 10.1016/j.jinteco.2019.01.012.

Cobham, David (2021). “A comprehensive classification of monetary policy frameworks in advanced and emerging economies”. In: *Oxford Economic Papers* 73.1, pp. 2–26. DOI: 10.1093/oep/gpz056.

Collobert, Ronan and Jason Weston (2008). “A unified architecture for natural language processing: deep neural networks with multitask learning”. In: *Proceedings of the 25th international conference on Machine learning*. ICML ’08. New York, NY, USA: Association for Computing Machinery, pp. 160–167. ISBN: 978-1-60558-205-4. DOI: 10.1145/1390156.1390177.

Correa, Ricardo, Keshav Garud, Juan M. Londono, and Nathan Mislang (2021). “Sentiment in central banks’ financial stability reports”. In: *Review of Finance* 25.1, pp. 85–120. DOI: 10.1093/rof/rfaa014.

Cragg, John G. and Stephen G. Donald (1997). “Inferring the rank of a matrix”. In: *Journal of Econometrics* 76.1, pp. 223–250. ISSN: 0304-4076. DOI: 10.1016/0304-4076(95)01790-9.

D’Agostino, Antonello and Karl Whelan (2008). “Federal Reserve Information during the Great Moderation”. In: *Journal of the European Economic Association* 6.2, pp. 609–620. ISSN: 1542-4766. URL: https://www.jstor.org/stable/40282670.

D'Acunto, Francesco, Andreas Fuster, and Michael Weber (2021). *Diverse Policy Committees Can Reach Underrepresented Groups*. Tech. rep. 29275. National Bureau of Economic Research, Inc. URL: https://ideas.repec.org/p/nbr/nberwo/29275.html.

Dixit, Avinash K. and Robert S. Pindyck (1994). *Investment Under Uncertainty*. Princeton University Press. 484 pp. ISBN: 978-0-691-03410-2.

Doh, Taeyoung, Dongho Song, Shu-Kuei Yang, et al. (2020). *Deciphering federal reserve communication via text analysis of alternative fomc statements*. Tech. rep. URL: https://papers.ssrn.com/sol3/Delivery.cfm/SSRN_ID4135736_code694895.pdf?abstractid=4135736&mirid=1.

Dominguez-Torres, Helena and Luis Ángel Hierro (2020). "Are there monetary clusters in the Eurozone? The impact of ECB policy". en. In: *Journal of Policy Modeling* 42.1, pp. 56–76. ISSN: 0161-8938. DOI: 10.1016/j.jpolmod.2019.11.001.

ECB (2020). *ECB announces new measures to increase share of female staff members*. URL: https://www.ecb.europa.eu/press/pr/date/2020/html/ecb.pr200514~94dbb7c109.en.html.

Egami, Naoki, Christian J Fong, Justin Grimmer, Margaret E Roberts, and Brandon M Stewart (2018). *How to make causal inferences using texts*. Tech. rep. URL: https://arxiv.org/pdf/1802.02163.pdf.

Ehrmann, Michael and Marcel Fratzscher (2007). "Communication by Central Bank Committee Members: Different Strategies, Same Effectiveness?" en. In: *Journal of Money, Credit and Banking* 39.2-3, pp. 509–541. ISSN: 1538-4616. DOI: 10.1111/j.0022-2879.2007.00034.x.

Ehrmann, Michael and Jonathan Talmi (2020). "Starting from a blank page? Semantic similarity in central bank communication and market volatility". In: *Journal of Monetary Economics* 111, pp. 48–62. DOI: 10.1016/j.jmoneco.2019.01.028.

Eickmeier, Sandra, Norbert Metiu, and Esteban Prieto (2016). *Time-varying Volatility, Financial Intermediation and Monetary Policy*. Tech. rep. 19/2016. Publication Title: IWH Discussion Papers. Halle Institute for Economic Research (IWH). URL: https://ideas.repec.org/p/zbw/iwhdps/iwh-19-16.html.

Elbourne, Adam (2019). *SVARs, the central bank balance sheet and the effects of unconventional monetary policy in the euro area*. Tech. rep. 407. CPB Netherlands Bureau for Economic Policy Analysis. URL: https://ideas.repec.org/p/cpb/discus/407.html.

Elbourne, Adam and Kan Ji (2019). *Do zero and sign restricted SVARs identify unconventional monetary policy shocks in the euro area?* Tech. rep. 391. CPB Netherlands Bureau for Economic Policy Analysis. URL: https://ideas.repec.org/p/cpb/discus/391.html.

Faust, Jon, Eric T Swanson, and Jonathan H Wright (2004). "Do Federal Reserve Policy Surprises Reveal Superior Information about the Economy?" In: *Contributions in Macroeconomics* 4.1. ISSN: 1534-6005. DOI: 10.2202/1534-6005.1246.

Ferrari, Massimo and Helena Le Mezo (2021). *Text-based recession probabilities*. Working Paper Series 2516. European Central Bank. URL: https://ideas.repec.org/p/ecb/ecbwps/20212516.html.

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

Fraccaroli, Nicolò, Alessandro Giovannini, and Jean-François Jamet (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*. Tech. rep. 2442. URL: https://www.ecb.europa.eu/pub/pdf/scpwps/ecb.wp2442~e78be127c0.en.pdf.

Fujiwara, Ippei and Yuichiro Waki (2019). *Private news and monetary policy - Forward guidance as Bayesian persuasion*. CAMA Working Papers 2019-91. Centre for Applied Macroeconomic Analysis, Crawford School of Public Policy, The Australian National University. URL: https://ideas.repec.org/p/een/camaaa/2019-91.html.

Gambacorta, Leonardo, Boris Hofmann, and Gert Peersman (2014). "The Effectiveness of Unconventional Monetary Policy at the Zero Lower Bound: A Cross-Country Analysis". In: *Journal of Money, Credit and Banking* 46.4, pp. 615–642. ISSN: 1538-4616. DOI: 10.1111/jmcb.12119.

Garg, Nikhil, Londa Schiebinger, Dan Jurafsky, and James Zou (2018). "Word embeddings quantify 100 years of gender and ethnic stereotypes". In: *Proceedings of the National Academy of Sciences* 115.16, 3635E3644. DOI: 10.1073/pnas.1720347115.

Gentzkow, Matthew, Bryan Kelly, and Matt Taddy (2019). "Text as Data". In: *Journal of Economic Literature* 57.3, pp. 535–574. DOI: 10.1257/jel.20181020.

Gentzkow, Matthew and Jesse M. Shapiro (2010). "What drives media slant? Evidence from US daily newspapers". In: *Econometrica* 78.1, pp. 35–71. DOI: 10.3982/ECTA7195.

Gertler, Mark and Peter Karadi (2015). "Monetary Policy Surprises, Credit Costs, and Economic Activity". In: *American Economic Journal: Macroeconomics* 7.1, pp. 44–76. ISSN: 1945-7707. DOI: 10.1257/mac.20130329.

Ginther, Donna K and Shulamit Kahn (2004). "Women in economics: moving up or falling off the academic career ladder?" In: *Journal of Economic perspectives* 18.3, pp. 193–214. DOI: 10.1257/0895330042162386.

Gorodnichenko, Yuriy, Tho Pham, and Oleksandr Talavera (2021). *The Voice of Monetary Policy*. Working Paper 28592. National Bureau of Economic Research. DOI: 10.3386/w28592.

Gürkaynak, Refet S., Brian Sack, and Eric Swanson (2005). "Do Actions Speak Louder Than Words? The Response of Asset Prices to Monetary Policy Actions and Statements". In: *International Journal of Central Banking* 1.1, pp. 55–93. URL: https://ideas.repec.org/a/ijc/ijcjou/y2005q2a2.html.

Haitsma, Reinder, Deren Unalmis, and Jakob de Haan (2016). "The impact of the ECB's conventional and unconventional monetary policies on stock markets". en. In: *Journal of Macroeconomics* 48, pp. 101–116. ISSN: 0164-0704. DOI: 10.1016/j.jmacro.2016.02.004.

Handlan, Amy (2020). *Text shocks and monetary surprises: Text analysis of FOMC statements with machine learning*. Tech. rep. URL: https://www.terry.uga.edu/sites/default/files/inline-files/Handlan-%20Amy_Paper.pdf.

Hansen, Stephen and Michael McMahon (2016). "Shocking language: Understanding the macroeconomic effects of central bank communication". In: *Journal of International Economics* 99, S114–S133. DOI: 10.1016/j.jinteco.2015.12.008.

Hansen, Stephen, Michael McMahon, and Andrea Prat (2018). "Transparency and Deliberation Within the FOMC: A Computational Linguistics Approach". In: *The Quarterly Journal of Economics* 133.2, pp. 801–870. DOI: 10.1093/qje/qjx045.

Hansen, Stephen, Michael McMahon, and Matthew Tong (2019). "The long-run information effect of central bank communication". In: *Journal of Monetary Economics* 108, pp. 185–202. DOI: 10.1016/j.jmoneco.2019.09.002.

Harris, Zellig S (1954). "Distributional structure". In: *Word* 10.2-3, pp. 146–162.

Hayo, Bernd, Kai Henseler, Marc Steffen Rapp, and Johannes Zahner (2020). *Complexity of ECB Communication and Financial Market Trading*. MAGKS Joint Discussion Paper Series in Economics 201919. Philipps-University Marburg. URL: https://ideas.repec.org/p/mar/magkse/201919.html.

Hayo, Bernd and Matthias Neuenkirch (2013). "Do Federal Reserve Presidents communicate with a Regional bias?" In: *Journal of Macroeconomics* 35, pp. 62–72. DOI: 10.1016/j.jmacro.2012.10.002.

Hoesch, Lukas, Barbara Rossi, and Tatevik Sekhposyan (2020). *Has the information channel of monetary policy disappeared? Revisiting the empirical evidence*. Tech. rep. Department of Economics and Business, Universitat Pompeu Fabra. URL: https://econpapers.repec.org/paper/upfupfgen/1701.htm.

Hollo, Daniel, Manfred Kremer, and Marco Lo Duca (2012). *CISS-a composite indicator of systemic stress in the financial system*. Tech. rep. ECB. URL: https://www.ecb.europa.eu//pub/pdf/scpwps/ecbwp1426.pdf.

Hornik, Kurt and Bettina Grün (2011). "topicmodels: An R package for fitting topic models". In: *Journal of Statistical Software* 40.13, pp. 1–30.

Hospido, Laura, Luc Laeven, and Ana Lamo (2019). "The gender promotion gap: evidence from central banking". In: *The Review of Economics and Statistics*, pp. 1–45. DOI: 10.1162/rest_a_00988.

Hu, Nan and Zexi Sun (2021). *Uncertain Talking at Central Bank's Press Conference: News or Noise?* Tech. rep. HKIMR Working Paper. URL: https://papers.ssrn.com/sol3/Delivery.cfm/SSRN_ID3796816_code849351.pdf?abstractid=3796816&mirid=1.

Istrefi, Klodiana, Florens Odendahl, and Giulia Sestieri (2020). *Fed communication on financial stability concerns and monetary policy decisions: revelations from speeches.* Tech. rep. URL: https://publications.banque-france.fr/sites/default/files/medias/documents/wp-779.pdf.

Jäger, Jannik and Theocharis Grigoriadis (2017). "The effectiveness of the ECB's unconventional monetary policy: Comparative evidence from crisis and non-crisis Euro-area countries". en. In: *Journal of International Money and Finance* 78, pp. 21–43. ISSN: 0261-5606. DOI: 10.1016/j.jimonfin.2017.07.021.

Jarociński, Marek and Peter Karadi (2020). "Deconstructing Monetary Policy Surprises—The Role of Information Shocks". In: *American Economic Journal: Macroeconomics* 12.2, pp. 1–43. ISSN: 1945-7707. DOI: 10.1257/mac.20180090.

Johnson, Palmer O. and Leo C. Fay (1950). "The Johnson-Neyman technique, its theory and application". In: *Psychometrika* 15.4, pp. 349–367. ISSN: 1860-0980. DOI: 10.1007/BF02288864.

Kalamara, Eleni, Arthur Turrell, Chris Redl, George Kapetanios, and Sujit Kapadia (2020). *Making text count: economic forecasting using newspaper text.* Tech. rep. 865. Bank of England. DOI: 10.2139/ssrn.3610770.

Kerssenfischer, Mark (2019). *Information effects of euro area monetary policy: New evidence from high-frequency futures data.* Tech. rep. 07/2019. Deutsche Bundesbank. URL: https://ideas.repec.org/p/zbw/bubdps/072019.html.

Kincaid, J. Peter, Robert P. Fishburne Jr., Richard L. Rogers, and Brad S. Chissom (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. URL: https://stars.library.ucf.edu/cgi/viewcontent.cgi?article=1055&context=istlibrary.

Knight, Frank Hyneman (1921). *Risk, uncertainty and profit.* Vol. 31. Houghton Mifflin.

Krippner, Leo (2013). "Measuring the stance of monetary policy in zero lower bound environments". In: *Economics Letters* 118.1, pp. 135–138. ISSN: 0165-1765. DOI: 10.1016/j.econlet.2012.10.011.

Krippner, Leo (2020). "A Note of Caution on Shadow Rate Estimates". In: *Journal of Money, Credit and Banking* 52.4, pp. 951–962. DOI: 10.1111/jmcb.12613.

Kuttner, Kenneth N (2001). "Monetary policy surprises and interest rates: Evidence from the Fed funds futures market". In: *Journal of Monetary Economics* 47.3, pp. 523–544. ISSN: 0304-3932. DOI: 10.1016/S0304-3932(01)00055-1.

Lau, Jey Han and Timothy Baldwin (2016). "An Empirical Evaluation of doc2vec with Practical Insights into Document Embedding Generation". In: *Proceedings of the 1st Workshop on Representation Learning for NLP*, pp. 78–86. DOI: 10.48550/arXiv.1607.05368.

Lautenschläger, Sabine (2017). *100 years of women's suffrage – gender equality, freedom and democracy*. URL: https://www.ecb.europa.eu/press/key/date/2017/html/ecb.sp170726.en.html.

Lazer, David, Ryan Kennedy, Gary King, and Alessandro Vespignani (2014). "The parable of Google Flu: traps in big data analysis". In: *science* 343.6176, pp. 1203–1205. DOI: 10.1126/science.1248506.

Le, Quoc and Tomas Mikolov (2014). "Distributed representations of sentences and documents". In: *International conference on machine learning*. PMLR, pp. 1188–1196. DOI: 10.48550/arXiv.1405.4053.

Loughran, Tim and Bill McDonald (2011). "When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks". In: *The Journal of Finance* 66.1, pp. 35–65. DOI: 10.1111/j.1540-6261.2010.01625.x.

Lowe, William (2021). *Text as Data: Summer School*.

Ludvigson, Sydney, Sai Ma, and Serena Ng (2021). "Uncertainty and Business Cycles: Exogenous Impulse or Endogenous Response?" In: *American Economic Journal: Macroeconomics*. ISSN: 1945-7707. DOI: 10.1257/mac.20190171.

Lundberg, Shelly (2017). "Report: Committee on the status of women in the Economics profession (CSWEP)". In: *American Economic Review* 107.5, pp. 759–76. DOI: 10.1257/pandp.108.704.

Manzini, Thomas, Yao Chong Lim, Yulia Tsvetkov, and Alan W. Black (2019). *Black is to Criminal as Caucasian is to Police: Detecting and Removing Multiclass Bias in Word Embeddings*. Tech. rep. URL: http://arxiv.org/abs/1904.04047.

Masciandaro, Donato, Davide Romelli, and Gaia Rubera (2020). *Does It Fit? Tweeting on Monetary Policy and Central Bank Communication*. Tech. rep. SUERF. URL: https://www.suerf.org/docx/f_e3c92f539bc17efbcc0c2e5229efa268_11243_suerf.pdf.

McKay, Alisdair, Emi Nakamura, and Jón Steinsson (2016). "The Power of Forward Guidance Revisited". In: *American Economic Review* 106.10, pp. 3133–3158. ISSN: 0002-8282. DOI: 10.1257/aer.20150063.

Mertens, Karel and Morten O. Ravn (2013). "The Dynamic Effects of Personal and Corporate Income Tax Changes in the United States". In: *American Economic Review* 103.4, pp. 1212–1247. ISSN: 0002-8282. DOI: 10.1257/aer.103.4.1212.

Mikolov, Tomas, Kai Chen, Greg S. Corrado, and Jeffrey Dean (2013). *Efficient estimation of word representations in vector space*. Tech. rep. URL: http://arxiv.org/abs/1301.3781.

Mikolov, Tomas, Ilya Sutskever, Kai Chen, Greg Corrado, and Jeffrey Dean (2013). *Distributed representations of words and phrases and their compositionality*. Tech. rep. DOI: 10.48550/arXiv.1310.4546.

Mikolov, Tomas, Wen-tau Yih, and Geoffrey Zweig (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*, pp. 746–751. URL: https://aclanthology.org/N13-1090.pdf.

Miranda-Agrippino, Silvia and Giovanni Ricco (2021). "The Transmission of Monetary Policy Shocks". In: *American Economic Journal: Macroeconomics*. ISSN: 1945-7707. DOI: 10.1257/mac.20180124.

Montiel Olea, José L., James H. Stock, and Mark W. Watson (2020). "Inference in Structural Vector Autoregressions identified with an external instrument". In: *Journal of Econometrics*. ISSN: 0304-4076. DOI: 10.1016/j.jeconom.2020.05.014.

Nakamura, Emi and Jón Steinsson (2018). "High-Frequency Identification of Monetary Non-Neutrality: The Information Effect". In: *The Quarterly Journal of Economics* 133.3, pp. 1283–1330. DOI: 10.1093/qje/qjy004.

Peek, Joe, Eric S Rosengren, and Geoffrey Tootell (2016). *Does Fed policy reveal a Ternary Mandate?* Tech. rep. Federal Reserve Bank of Boston. URL: https://www.econstor.eu/bitstream/10419/171765/1/86922896X.pdf.

Pellegrino, Giovanni (2018). "Uncertainty and the real effects of monetary policy shocks in the Euro area". In: *Economics Letters* 162, pp. 177–181. ISSN: 0165-1765. DOI: 10.1016/j.econlet.2017.10.006.

Pellegrino, Giovanni (2020). *Uncertainty and Monetary Policy in the US: A Journey into Non-Linear Territory*. Tech. rep. 2020-05. Department of Economics and Business Economics, Aarhus University. URL: https://ideas.repec.org/p/aah/aarhec/2020-05.html.

Pennington, Jeffrey, Richard Socher, and Christopher D Manning (2014). "Glove: Global vectors for word representation". In: *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pp. 1532–1543. DOI: 10.3115/v1/D14-1162.

Picault, Matthieu and Thomas Renault (2017). "Words are not all created equal: A new measure of ECB communication". In: *Journal of International Money and Finance* 79, pp. 136–156. DOI: 10.1016/j.jimonfin.2017.09.005.

Rehurek, Radim and Petr Sojka (2011). *Gensim–python framework for vector space modelling*. Tech. rep. 2.

Romer, David H. and Christina D. Romer (2000). “Federal Reserve Information and the Behavior of Interest Rates”. In: *American Economic Review* 90.3, pp. 429–457. URL: https://ideas.repec.org/a/aea/aecrev/v90y2000i3p429-457.html.

Rossi, Barbara (2020). “Identifying and estimating the effects of unconventional monetary policy: How to do it and what have we learned?” In: *The Econometrics Journal* 24.1, pp. C1–C32. ISSN: 1368-4221. DOI: 10.1093/ectj/utaa020.

Rossi, Barbara and Tatevik Sekhposyan (2016). “Forecast Rationality Tests in the Presence of Instabilities, with Applications to Federal Reserve and Survey Forecasts”. In: *Journal of Applied Econometrics* 31.3, pp. 507–532. ISSN: 1099-1255. DOI: 10.1002/jae.2440.

Schmeling, Maik and Christian Wagner (2019). *Does Central Bank Tone Move Asset Prices?* CEPR Discussion Papers 13490. URL: https://EconPapers.repec.org/RePEc:cpr:ceprdp:13490.

Shapiro, Adam H., Moritz Sudhof, and Daniel J. Wilson (2020). “Measuring news sentiment”. In: *Journal of Econometrics*. DOI: 10.1016/j.jeconom.2020.07.053.

Shapiro, Adam H. and Daniel J. Wilson (2019). *Taking the Fed at its Word: A New Approach to Estimating Central Bank Objectives Using Text Analysis*. Tech. rep. Federal Reserve Bank of San Francisco, pp. 01–74. DOI: 10.24148/wp2019-02.

Smales, L.A. and N. Apergis (2017a). “Does More complex Language in FOMC Decisions impact Financial Markets?” In: *Journal of International Financial Markets, Institutions and Money* 51, pp. 171–189. DOI: 10.1016/j.intfin.2017.08.003.

Smales, LA and Nicholas Apergis (2017b). “Understanding the impact of monetary policy announcements: The importance of language and surprises”. In: *Journal of Banking & Finance* 80, pp. 33–50. DOI: 10.1016/j.jbankfin.2017.03.017.

Stock, James H. and Mark W. Watson (2012). “Disentangling the Channels of the 2007–09 Recession”. In: *Brookings Papers on Economic Activity* 2012.1, pp. 81–135. ISSN: 1533-4465. DOI: 10.1353/eca.2012.0005.

Stock, James H. and Mark W. Watson (2018). “Identification and Estimation of Dynamic Causal Effects in Macroeconomics Using External Instruments”. In: *The Economic Journal* 128.610, pp. 917–948. ISSN: 1468-0297. DOI: 10.1111/ecoj.12593.

Stock, James H. and Motohiro Yogo (2001). “Testing for weak instruments in linear IV regression.” In: *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*. Ed. by Donald W. K. Andrews and James H. Stock. Vol. 1; 1. February. Cambridge: Cambridge University Press, pp. 80–108. URL: https://eml.berkeley.edu/symposia/nsf01/stock_p.pdf.

Swanson, Eric T. (2021). “Measuring the Effects of Federal Reserve Forward Guidance and Asset Purchases on Financial Markets”. en. In: *Journal of Monetary Economics* 118, pp. 32–53. ISSN: 0304-3932. DOI: 10.1016/j.jmoneco.2020.09.003.

Swanson, Eric T. and John C. Williams (2014). "Measuring the Effect of the Zero Lower Bound on Medium- and Longer-Term Interest Rates". In: *American Economic Review* 104.10, pp. 3154–3185. DOI: 10.1257/aer.104.10.3154.

Sweeney, Chris and Maryam Najafian (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*. Association for Computational Linguistics, pp. 1662–1667. DOI: 10.18653/v1/P19-1162.

Tadle, Raul Cruz (2022). "FOMC minutes sentiments and their impact on financial markets". In: *Journal of Economics and Business* 118, p. 106021. ISSN: 0148-6195. DOI: 10.1016/j.jeconbus.2021.106021.

Tetlock, Paul C (2007). "Giving content to investor sentiment: The role of media in the stock market". In: *The Journal of Finance* 62.3, pp. 1139–1168. DOI: 10.1111/j.1540-6261.2007.01232.x.

Tillmann, Peter (2021). "Financial markets and dissent in the ECB's Governing Council". In: *European Economic Review* 139, p. 103848. ISSN: 0014-2921. DOI: 10.1016/j.euroecorev.2021.103848.

Tillmann, Peter (forthcoming). "Macroeconomic surprises and the demand for information about monetary policy". In: *International Journal of Central Banking*. URL: https://drive.google.com/file/d/17NeXzG3qcbmo0FQCIFz_3h5jSABQNLZi/view?usp=sharing.

Tobback, Ellen, Stefano Nardelli, and David Martens (2017). *Between Hawks and Doves: Measuring Central Bank Communication*. Tech. rep. ECB. URL: https://www.ecb.europa.eu/pub/pdf/scpwps/ecb.wp2085.en.pdf.

Tukey, John W. (1949). "Comparing Individual Means in the Analysis of Variance". In: *Biometrics* 5.2, pp. 99–114. ISSN: 0006341X, 15410420. URL: http://www.jstor.org/stable/3001913.

Turian, Joseph, Lev Ratinov, and Yoshua Bengio (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*. ACL '10. USA: Association for Computational Linguistics, pp. 384–394. URL: https://aclanthology.org/P10-1040.pdf.

Wischnewsky, Arina, David-Jan Jansen, and Matthias Neuenkirch (2021). "Financial stability and the Fed: Evidence from congressional hearings". In: *Economic Inquiry* 59.3, pp. 1192–1214. DOI: 10.1111/ecin.12977.

Wittgenstein, Ludwig (1958). *Philosophische Untersuchungen (Zweite Auflage). English edition.*

Wu, Jing Cynthia and Fan Dora Xia (2016). "Measuring the Macroeconomic Impact of Monetary Policy at the Zero Lower Bound". In: *Journal of Money, Credit and*

Banking 2-3.48, pp. 253–291. ISSN: 0022-2879, 1538-4616. DOI: 10.1111/jmcb.12300.

Zahner, Johannes (2020). *Above, but close to two percent. Evidence on the ECB's inflation target using text mining.* MAGKS Joint Discussion Paper Series in Economics. Philipps-University Marburg. URL: https://www.uni-marburg.de/en/fb02/research/institutes/macie/activities/macie-papers/2020/2020_05_zahner.pdf.

Affidavit

Ich erkläre hiermit, dass ich die vorgelegten und nachfolgend aufgelisteten Aufsätze selbstständig und nur mit den Hilfen angefertigt habe, die im jeweiligen Aufsatz angegeben oder zusätzlich in der nachfolgenden Liste aufgeführt sind. In der Zusammenarbeit mit den angeführten Koautoren war ich mindestens anteilig beteiligt. Bei den von mir durchgeführten und in den Aufsätzen erwähnten Untersuchungen habe ich die Grundsätze guter wissenschaftlicher Praxis, wie sie in der Satzung der Justus-Liebig-Universität Gießen zur Sicherung guter wissenschaftlicher Praxis niedergelegt sind, eingehalten.

Martin Baumgärtner | Bonn, der 02. Juli 2022