



LUND UNIVERSITY

Text-Based Identification of Monetary Policy Shocks

Almgren, Pelle

2026

Document Version:
Other version

[Link to publication](#)

Citation for published version (APA):

Almgren, P. (2026). *Text-Based Identification of Monetary Policy Shocks*. (Working Papers; No. 2026:10).

Total number of authors:

1

General rights

Unless other specific re-use rights are stated the following general rights apply:

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Read more about Creative commons licenses: <https://creativecommons.org/licenses/>

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

LUND UNIVERSITY

PO Box 117
221 00 Lund
+46 46-222 00 00

Working Paper 2026:10

Department of Economics
School of Economics and Management

Text-Based Identification of Monetary Policy Shocks

Pelle Almgren

September 2026



LUND
UNIVERSITY

Text-Based Identification of Monetary Policy Shocks*

Pelle Almgren[†]

September, 2026

Abstract

This paper identifies monetary policy shocks for Sweden. Based on the written language in documents prepared by staff ahead of Riksbank policy meetings, I construct text-based regressors that are shown to systematically predict shocks identified with standard methods such as the Romer and Romer (2004) Greenbook approach and high-frequency identification. Defining text-based monetary policy shocks as the part of the standard shock that is uncorrelated with this rich nonstandard source of information, I study the effect of monetary policy on key monetary aggregates using local projections. The dynamic responses to my text-based policy shocks are in line with theoretical expectations for a majority of variables at both quarterly and annual frequency. Text-based Romer and Romer shocks are more in line with 'textbook' responses compared to high-frequency based shocks. All results are robust to alternative text-analytic implementations, model specifications and lag length selection in local projections.

Keywords: Monetary Policy Shocks, Textual Analysis, Riksbank, BERT, Local Projections.

JEL Classification: C22, C38, C45, E32, E52, E58.

*I thank Anders Vilhelmsson, Joakim Westerlund and Andreas Ek for support, discussions and constructive comments.

[†]Department of Economics, Lund University, Box 7082, 220 07 Lund, Sweden. E-mail address: pelle.almgren@nek.lu.se

1 Introduction

Macroeconomists typically think of changes in central bank policy rates as a mix of systematic responses to economic conditions and non-systematic variation unrelated to the real economy. Since the systematic component of policy rates is endogenous to economic variables, causal study of the effect of monetary policy on the economy requires isolating the non-systematic variation. This quantity is referred to as monetary policy shocks and originate from e.g. changes in central bank preferences, under- or overreactions to economic events, or intended policy shocks (Coibion, 2012; Cloyne and Hürtgen, 2016). A vast literature has developed empirical methods for identifying monetary policy shocks (see Ramey, 2016 for a review).

In a recent publication, Aruoba and Drechsel (2026) show that nonstandard data such as the natural language used in central bank documents prepared ahead of each policy meeting provides an important source of information when identifying monetary policy shocks. In this paper, I build on their approach and develop a method that relies on modern deep learning and machine learning tools for textual analysis to identify monetary policy shocks for Sweden. The Swedish context is appealing because such documents are published at the same time the policy decision is announced. The main idea is that the language used in central bank documents paints a rich picture about what information is available to the central bank at each policy meeting, and how the central bank react to current and changing economic conditions. Such information is valuable when separating systematic from non-systematic variation in the central bank policy rate. The idea is not new, but is in fact the principal argument behind the Romer and Romer (1989) 'narrative' method for identifying shocks. The same idea informed Jarociński and Karadi (2020) how to disentangle central bank information shocks from monetary surprises and motivated Romer and Romer (2023) to revisit their 1989 approach. What is new in this paper is the tools with which I quantitatively model narrative information, and how the information is used to identify policy shocks.

Based on the natural language in Inflation Reports, Monetary Policy Reports and Monetary Policy Updates, I construct time series that summarize the large set of information available to the Riksbank Executive Board at each policy meeting. I use these time series in the two most popular approaches to identify monetary policy shocks, the Romer and Romer (2004) (R&R) Greenbook approach, and the High-Frequency (HF) identification strategy developed by Kuttner (2001) and

Gertler and Karadi (2015) and others. For both types of monetary policy shocks, I show that text-based regressors explain a significant part of the variation in standard monetary policy shocks. The rich information available in Riksbank documents is thus crucial for identifying shocks.

I estimate text-based monetary policy shocks by including text-based regressors in the R&R proxy for the central bank information set, and by constructing a HF instrument for monetary surprises that is orthogonal to these regressors. To study the effect of my policy shocks on macroeconomic aggregates, I estimate local projections (Jordà, 2005). For most variables, the findings are in line with theoretical expectations as output and prices decrease, and unemployment increase following a contractionary shock. Text-based R&R shocks alleviate a price puzzle present in the response to standard R&R shocks, this is not the case for HF-type shocks.

This paper contributes to two strands of the literature. First, to the literature that identify monetary policy shocks with the R&R method (see e.g. Cloyne and Hürtgen, 2016; Wieland and Yang, 2020; Holm et al., 2021) and the HF method (see e.g. Amberg et al., 2022; Acosta et al., 2025; Kekre and Lenel, 2025). In this broad literature, the current paper relates in particular to the new branch that suggest dimensions along which the identification of monetary policy shocks can be improved (see e.g. Jarociński and Karadi, 2020; Bauer and Swanson, 2023a; Aruoba and Drechsel, 2026). For the R&R method, Aruoba and Drechsel (2026) show that text-based regressors, modeled using manually created groups of words and dictionary methods, are crucial to identify a proxy for the Federal Reserve (Fed) reaction function. I use modern, and conceptually more appealing, methods for constructing such text-based regressors. My results confirm those of Aruoba and Drechsel (2026) in that this type of information is important for identifying a proxy for the central bank reaction function, and therefore for the identification of monetary policy shocks, however in the context of a central bank operating in a small open economy.

When identifying HF shocks, it is important to disentangle policy shocks from central bank information shocks in HF monetary surprises (see Nakamura and Steinsson, 2018; Jarociński and Karadi, 2020). In parallel, Miranda-Agrippino and Ricco (2021) and Bauer and Swanson (2023a) show that HF monetary surprises are often predictable with information that predate policy announcements such as Greenbook forecasts and developments on labor and equity markets. These authors recommend orthogonalizing the policy instrument to such information. My results show that text-based information predict monetary policy surprises defined with the poor man's sign re-

striction (PMSR), based on which I construct an instrument orthogonal to the textual information.

A second branch of the literature the current paper contribute to is the literature that uses textual analysis on central bank communication and documents. Hansen et al. (2017) use Federal Open Market Committee (FOMC) communications to study central bank transparency and Shapiro and Wilson (2022) analyze Fed objectives on similar data. Closely related to the economic policy uncertainty index introduced by Baker et al. (2016), Cieslak et al. (2023) construct text-based measures of policy maker uncertainty. Different from these papers, similar to Bertsch et al. (2025) and Silva et al. (2025), I base my textual analysis of central bank communications on modern deep learning methods. Hansen and McMahon (2016) study the effect of central bank communication, particularly about forward guidance and current conditions, on the macro economy. Gardner et al. (2022) study the response of equity prices to Fed statements and Cieslak and Vissing-Jorgensen (2021) study the reverse relation. While these papers study the response in monetary aggregates and equity prices to shocks in sentiment indices directly, I use sentiment indices to identify monetary policy shocks in a first step and study responses to those shocks in a second step.

The rest of the paper is structured as follows. Section 2 introduces two standard identification strategies and estimate monetary policy shocks using Swedish data. Section 3 describes my method to quantitatively model narrative information in central bank documents, and estimates text-based monetary policy shocks. Section 4 present responses of macroeconomic variables to monetary policy shocks. Section 5 verifies the robustness of my results. Section 6 concludes.

2 Standard Identification Strategies

Variation in a central banks' policy tool is commonly decomposed into systematic and non-systematic responses to current and expected future economic conditions. Systematic variation captures the endogenous relation between monetary policy and economic variables while the non-systematic component are referred to as monetary policy shocks (Coibion, 2012). Monetary policy shocks are typically thought of as variation due to shifts in central bank preferences, private information of central bankers, updated views on how policy should be conducted, or intended policy shocks (Cloyne and Hürtgen, 2016). An influential literature has identified the causal effect of monetary policy by isolating changes in the policy rate that are orthogonal to the systematic re-

sponses of central banks (see e.g. Romer and Romer, 2004; Bernanke et al., 2005; Gertler and Karadi, 2015 for alternative approaches, and Ramey, 2016 for an overview). In general, this idea can be described by the following model

$$\Delta s_t = f(\Omega_t) + \varepsilon_t, \quad (1)$$

where Δs_t is the change in policy rate decided at time t and Ω_t represents the information set available to the central bank at t . The central bank reaction function is captured by $f(\cdot)$ which describe the mapping between the information set and the policy instrument and is thought of as capturing the systematic component of monetary policy. The non-systematic component of monetary policy, the monetary policy shock, is captured in ε_t .

The three most common approaches to identifying monetary policy shocks are i) by estimating versions of R&Rs linear regression of Greenbook forecasts on policy rate changes, ii) by using HF movements of market prices around policy announcements and iii) by recovering reduced form innovations from a VAR model. While there exist extensive work on identifying monetary policy shocks in VAR models (see e.g. Blanchard and Perotti, 2002; Bernanke et al., 2005; Uhlig, 2005; Antolín-Díaz and Rubio-Ramírez, 2018; Fusari, 2025), it is increasingly common to include externally identified monetary policy shocks as instruments in VARs and use the parameters in the VAR model to study the dynamic effect of such shocks. Examples of this include Mertens and Ravn (2014), Caldara and Herbst (2019), and Miranda-Agrippino and Ricco (2021) for the US economy and Laséen (2020), Amberg et al. (2022), and Almerud et al. (2024) for Sweden. Such external shocks are most often identified using either the HF or the R&R approach, which are the methods I will focus on in this paper.

2.1 R&R Identification

The R&R approach to identify monetary policy shocks is based on defining an empirical proxy for the central bank reaction function which is used in an empirical counterpart to Equation (1) to isolate exogenous changes in in policy rate. Formally, the R&R shocks are the residuals in the least squares regression

$$\Delta s_t = \alpha + \beta s_{t-1} + \gamma \mathbb{X}_t + \varepsilon_t^{RR}, \quad (2)$$

where Δs_t is the change in the Riksbank policy rate decided at meeting t and s_{t-1} the level of the policy rate at the previous meeting included to capture tendencies for mean reversion in the policy rate (Romer and Romer, 2004). $\mathbb{X}_t$ contains economic forecasts of output growth and inflation available to the central bank at time t . $\mathbb{X}_t$ traditionally also include a variable capturing unemployment. By using forecasts, which are constructed in separate models that generally use a large set of variables, it is possible to condition on a large information set using only a few variables in Equation (2). Additionally, since monetary policy operates with a lag, using forecasts and forecast updates allows proxying future expected states of the economy.

New forecasts are currently produced for four of the eight annual meetings of the Executive board, no new forecasts are available for the remaining four meetings. Before each meeting, the Riksbank staff prepare documents that contain forecasts, data and analyses which forms the basis for the policy decision (Sveriges Riksbank, 2026b). Documents with new forecasts are currently called Monetary Policy Reports (MPRs) while those without are called Monetary Policy Updates (MPUs), see Section 2.3 for more details on Riksbank communication and report contents¹. It is important to note that all content in both the MPRs and MPUs are produced before each policy meeting, and the Riksbank staff learn about the policy decision when the decision, along with the current MPR or MPU, are made available to the public. In Sweden, the policy rate is announced the day after the Executive Board made their decision (Sveriges Riksbank, 2026b). This procedure eliminates reverse causality from policy decision to Greenbook forecasts in Equation (2).

In context of the theoretical model in Equation (1), the R&R framework use $\widehat{f(\Omega_t)} = \hat{\alpha} + \hat{\beta}i_{t-1} + \hat{\gamma}\mathbb{X}_t$ to proxy the systematic component of monetary policy. Assuming a linear decomposition, following Andersson and Kilman (2021), this estimation recovers

$$\hat{\varepsilon}_t^{RR} = \varepsilon_t + \nu_t, \quad (3)$$

where ε_t is as before and ν_t is the error associated with estimating $f(\Omega_t)$ with $\widehat{f(\Omega_t)}$. The often

¹The publishing scheme for these documents, and their names, has varied over the sample. Importantly, the content in these documents has always been a summary of the basis for the monetary policy decision at each meeting. What differs is that up until 2007, all documents were called *Inflation Reports*, after this point the Riksbank distinguish between reports that publish new forecasts of key economic variables (*Monetary Policy Reports*) and those that do not (*Monetary Policy Updates*). In detail: from June 1996 to October 2006, they were called *Inflation Reports*, published quarterly. During 2007 and 2008, a *Monetary Policy Report* was published for the first three quarters and a *Monetary Policy Update* was published in the fourth quarter. Between 2009 and 2014, the Riksbank published six reports per year, *Monetary Policy Reports* in February, July and October and *Monetary Policy Updates* in April, September and December. In 2015-2019, the Riksbank published six *Monetary Policy Reports*, one each in February, April, July, September, October and December. For details on the contents of these reports, see Section D

implicit assumption in the literature is that $\nu_t \stackrel{iid}{\sim} (0, \sigma_\nu)$ under which R&R identifies monetary policy shocks.

Using a dataset consisting of Riksbank forecasts of output growth, inflation and unemployment for the period 1996-2019 recently compiled in Kilman (2022), I estimate the following model²

$$\begin{aligned} \Delta s_t = & \alpha + \beta s_{t-1} + \sum_{h=0}^1 \gamma_h^y \tilde{y}_{t,h} + \sum_{h=0}^1 \gamma_h^\pi \tilde{\pi}_{t,h} + \sum_{h=0}^1 \gamma_h^u \tilde{u}_{t,h} \\ & + \sum_{h=0}^1 \theta_h^y \Delta \tilde{y}_{t,h} + \sum_{h=0}^1 \theta_h^\pi \Delta \tilde{\pi}_{t,h} + \sum_{h=0}^1 \theta_h^u \Delta \tilde{u}_{t,h} + \varepsilon_t^{RR}, \end{aligned} \quad (4)$$

where Δs_t and s_{t-1} are as before. Forecasts of output growth, CPI inflation and unemployment for the current and upcoming year, $h \in \{0, 1\}$, available to the Riksbank at meeting t are captured with $\tilde{y}_{t,h}$, $\tilde{\pi}_{t,h}$ and $\tilde{u}_{t,h}$, respectively. Forecast updates, which capture revisions in the forecasts relative to the previous round of forecasts, are captured with $\Delta \tilde{y}_{t,h}$, $\Delta \tilde{\pi}_{t,h}$ and $\Delta \tilde{u}_{t,h}$. For observations where no new forecasts are available, in line with Cloyne and Hürtgen (2016) and Kilman (2022), I use the latest available forecast.

Compared to Romer and Romer (2004), my empirical approach deviates in three ways. First, since quarterly forecasts are not available for Sweden before 2007, I use annual forecasts. Second, also due to data limitations, I do not use real-time data for past periods to compute forecast updates. Third, I treat unemployment similar to inflation and the GDP growth rate by including current and next period forecasts and forecast revisions instead of just the contemporaneous forecast.

Table 1 presents the results from estimating Equation (4). All significant coefficients have the expected sign and suggest, in line with e.g. Cloyne and Hürtgen (2016) and Holm et al. (2021), a counter-cyclical behavior. A one percentage point higher policy rate at the previous meeting results in a 0.053 percentage point lower policy rate at the current meeting. If the GDP growth forecast for next year increases with one percentage point, the policy rate significantly increases with 0.138 percentage points. Finally, if the change in the inflation forecast for the current year increases with one percentage point, the policy rate increases with 0.130 percentage points.

²In this sample, there was a methodological change in how the forecasts were produced. Forecasts from before 2005 were based on an assumption of a constant policy rate, forecasts between 2005-2007 on the assumption that the policy rate follows market prices of interest rates, forecast since 2007 are based on the Riksbanks' own policy rate forecast (Riksbanken, 2018). This methodological change is not a threat for identification as the Riksbank staff always produce the forecasts before the policy decision is taken. The change does not matter for the effects of the identified shocks, see Kilman (2022) for analysis on shocks identified in a sample between 1996-2006 or Section 5 in this paper for a study of shocks identified between 2007-2019.

The results presented in Table 1 indicates that the systematic component of monetary policy, as proxied by Equation (4), explains 32.9% of the total variation in the Riksbank policy rate. This is comparable to recent estimates for Sweden (Kilman, 2022), Norway (Holm et al., 2021), the UK (Cloyne and Hürtgen, 2016) and the original R&R study. This means that approximately 67% of the variation in the Riksbank policy rate is due to monetary policy shocks.

Table 1: Estimation of Romer & Romer Monetary Policy Shocks

Variable	Estimate
s_{t-1}	−0.0392* (0.0201)
$\tilde{y}_{t,0}$	0.00346 (0.0271)
$\tilde{y}_{t,1}$	0.138** (0.0556)
$\Delta\tilde{y}_{t,0}$	0.0396 (0.0344)
$\Delta\tilde{y}_{t,1}$	−0.0834 (0.0943)
$\tilde{\pi}_{t,0}$	0.0118 (0.0379)
$\tilde{\pi}_{t,1}$	0.0728 (0.0698)
$\Delta\tilde{\pi}_{t,0}$	0.130*** (0.0441)
$\Delta\tilde{\pi}_{t,1}$	−0.0002 (0.0540)
$\tilde{u}_{t,0}$	−0.0637 (0.0978)
$\tilde{u}_{t,1}$	0.0281 (0.100)
$\Delta\tilde{u}_{t,0}$	0.0425 (0.107)
$\Delta\tilde{u}_{t,1}$	−0.0756 (0.101)
α	−0.230 (0.148)
Observations	167
F-statistic	9.60
R^2_{adj}	0.329

Notes: The dependent variable is the change in the policy rate. Time progress at meeting frequency. Newey–West standard errors are reported in parentheses. ***: $p < 0.01$, **: $p < 0.05$, *: $p < 0.1$.

I use the standard practice of converting the shocks estimated at meeting frequency to monthly, quarterly and annual frequency by summing up shocks in a period if there are multiple meetings in that period. For periods without meetings, I set the shock to zero. The R&R shocks are plotted at quarterly frequency in Figure 1, for monthly and annual shocks, see Figure 7 in Appendix A.

2.2 HF Identification

Following the work of Kuttner (2001), Gertler and Karadi (2015), and Nakamura and Steinsson (2018) and others, another methodology for identifying monetary policy shocks uses HF movements on financial markets in a narrow window around monetary policy announcements. Federal funds futures or treasury bills with short maturities are typically used as they are close proxies for forecasts of monetary policy. This approach is appealing in the sense that for policy decisions taken sufficiently long before the policy announcement, surprise changes in asset prices in a narrow window around the announcement is clearly caused by the announcement, and not vice versa (Bauer and Swanson, 2023a). For efficient markets, changes in asset prices capture policy surprises that can be used to study the causal effect of monetary policy. If, on the other hand, the policy decision was expected, prices on financial markets should not move systematically at the announcement.

A large literature uses this type of shock to study the causal effect of monetary policy on both asset prices (see e.g. Hanson and Stein, 2015; Swanson, 2021; Kekre and Lenel, 2025; Acosta et al., 2025) and macroeconomic variables (see e.g. Stock and Watson, 2018; Jarociński and Karadi, 2020; Bauer and Swanson, 2023a). Swedish examples include Sandström (2018), Amberg et al. (2022), and Almerud et al. (2024).

Recent papers suggest that such HF surprises contain two components, a monetary policy surprise and a central bank information surprise (Nakamura and Steinsson, 2018; Jarociński and Karadi, 2020). This is largely based on observed positive co-movement between market rates and broad stock indices on many announcement days while the standard transmission mechanism of monetary policy would suggest negative co-movement (Jarociński and Karadi, 2020). Theoretically, a policy tightening mechanically increases the discount rate and decreases expected future dividends via deteriorating financial conditions, both leading to a lower valued stock market. Positive co-movement are thought of as caused by central bank information about current or future economic conditions (Jarociński and Karadi, 2020)³.

Therefore, to study the causal effect of monetary policy, it is crucial to disentangle monetary policy surprises from central bank information surprises in the HF price movements around an-

³Formally, when the central bank announces policy decision s_t which depends on central bank information, $s_t = f(\Omega_t^{CB})$, the financial market participants observe this signal and update their information set, $s_t \in \Omega_{t+\tau}^{FM}$ where τ is some small number. In doing so, they trade which moves prices on both interest and equity markets, $\Delta P_{t+\tau} = g(\Omega_{t+\tau}^{FM})$ and $\Delta i_{t+\tau} = h(\Omega_{t+\tau}^{FM})$ where P is a broad stock index, i short market rates, and g and h are functions that describe how financial market participants set prices as a function of their current information. A monetary policy surprise is, by this definition, identified on days where $sign(\Delta P_{t+\tau}) \neq sign(\Delta i_{t+\tau})$.

nouncements. One common approach for doing this is to impose Jarociński and Karadi (2020)'s Poor Man's Sign Restriction (PMSR) which decompose HF surprises into pure monetary policy surprises and pure central bank information surprises⁴. The HF-PMSR identification has recently been used in e.g. Sweden (Amberg et al., 2022), Spain (Buda et al., 2025) and the Euro Area (Delis et al., 2025) to study different aspects of the transmission of monetary policy⁵.

In general, these authors estimate the following model

$$\begin{aligned}\Delta s_t &= \alpha + \beta \times \Delta i_t^{PMSR} + \omega_t, \\ \Delta i_t^{PMSR} &= \Delta i_t \times D_t, \\ D_t &= \begin{cases} 1 & \text{if } \text{sign}(\Delta P_t) \neq \text{sign}(\Delta i_t) \\ 0 & \text{otherwise,} \end{cases} \end{aligned} \quad (5)$$

where Δi_t is the change in short market rates and ΔP_t is the change on a broad shock index such as the S&P500, the EURO STOXX 50 or the Stockholm All Shares Index (OMXSPI). All changes are computed in a window around the policy announcement at day t . From this definition of D_t , it is clear that only HF surprises on days where the equity and interest markets move in opposite directions are taken as monetary policy surprises. The fitted values from Equation (5), $\widehat{\Delta s_t}$, are taken as HF monetary policy shocks. In contrast to the R&R approach which revolved around identifying a functional form for the central banks reaction function, $f(\Omega_t)$, with which the shock $\hat{\varepsilon}_t^{RR}$ was taken as least squares residuals, the HF method aims to identify the non-systematic part of monetary policy, ε_t in Equation (1), directly as the part of the variation in Δs_t explained by the unpredictable monetary policy surprise, Δi_t^{PMSR} . That is, $\widehat{\Delta s_t} = \hat{\varepsilon}_t^{HF}$.

Using the replication package accompanying Amberg et al. (2022)⁶ I estimate Equation (5). The available dataset contain daily changes in the Riksbank policy rate, Δs_t , 1-month Swedish Treasury Bills, Δi_t , and the OMXSPI, P_t , at policy announcement days from February 1999 to December 2018. Results from this estimation are presented in Table 2.

While it is uncommon to interpret the coefficient on Δi_t^{PMSR} , it is expected to be positive given

⁴Jarociński and Karadi (2020) also present a more sophisticated set of sign restrictions that allow, in a VAR setting, a HF surprise to be partly monetary policy surprise and partly central bank information shock.

⁵Fusari (2025) is a recent paper that uses PMSR in VAR model in the US setting, however with a focus of making such shocks also condition on central bank information as proxied by Greenbook forecasts.

⁶Available at: <https://www.aeaweb.org/articles?id=10.1257/aeri.20210262>. Accessed: 2025-10-23.

the close relation between the policy rate and market interest rates. The estimation results suggest that the non-systematic component of monetary policy, the monetary policy shock, explain 17.55% of the variation in the policy rate. This is in stark contrast to the R&R approach which identified 67.1% of the variation in Δs_t as a shock.

Table 2: Estimation of High-Frequency Monetary Policy Shocks

Variable	Estimate
Δi_t^{PMSR}	2.131*** (0.385)
α	-0.006 (0.0197)
Observations	140
R_{adj}^2	0.176

Notes: The dependent variable is the change in the policy rate. Standard errors are reported in parentheses. Time progress at meeting frequency. The explanatory variable Δi_t^{PMSR} denotes the high-frequency monetary policy surprise around policy announcements, filtered with the PMSR filter. ***: $p < 0.01$, **: $p < 0.05$, *: $p < 0.1$.

After estimating the HF-PMSR shocks at the announcement frequency, I convert the shocks to monthly, quarterly and annual frequency. The quarterly shocks, plotted against the R&R shocks, are presented in Figure 1 while the HF-PMSR shocks at monthly and yearly frequency are shown in Figure 8 in Appendix A.

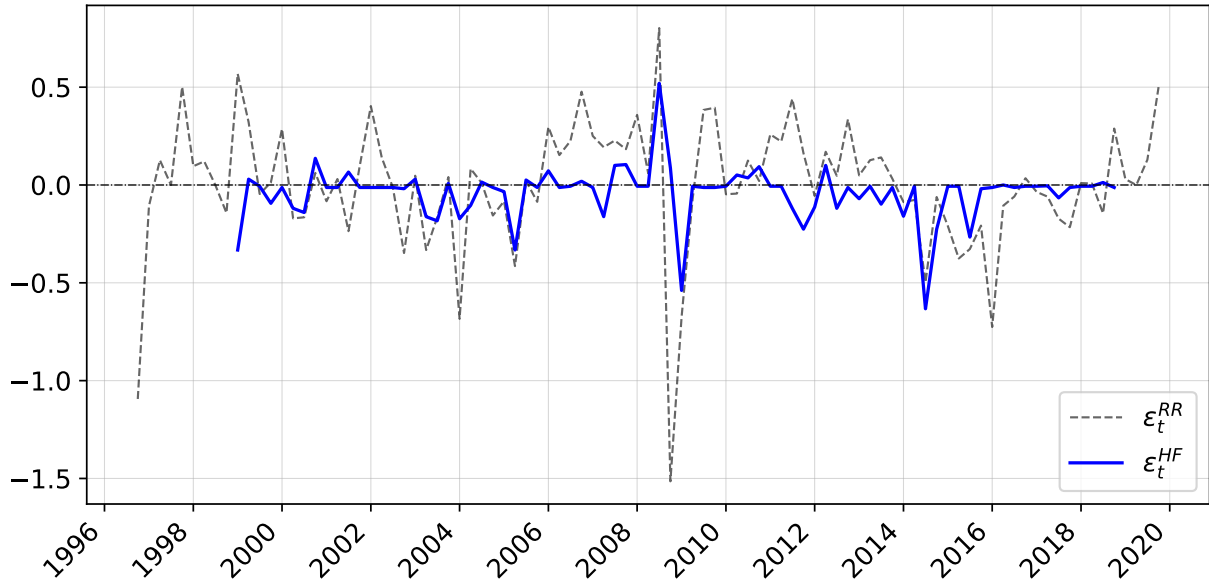


Figure 1: R&R and HF Shocks at Quarterly Frequency

The R&R approach identify more monetary policy shocks and shocks that display larger volatil-

ity over the sample compared to the HF approach. While both types of shocks captures many important events in the Swedish economy, the HF-PMSR shocks are smaller in magnitude and less frequent in contrast to the R&R shocks. The correlation between the shocks is 0.42 and both methods identify large shocks during the global financial crisis in 2008 and the loosening of policy rates and quantitative easing implemented between 2015-2019. See Andersson and Jonung (2018) and Kilman (2022) for a more detailed discussion on the origin of Swedish policy shocks.

The R&R approach also clearly identify shocks in the periods of fear for spiraling household debt levels that led to tighter than expected policy before and after the global financial crisis. This was not the case for HF shock which did not identify any, or only very small, shocks in this period. While there exist observations where the two shocks move in completely different directions, it is more often the case that R&R identifies a shock while HF finds no shock, as in the period around the financial crisis. This happens for example in the beginning of 2002, 2007 and after 2016. For such observations, it means that financial market participants already had adjusted to the news provided in the policy announcement before it happened. This is expected to happen when the Riksbank informs the public about e.g. quantitative easing and policy decreases during the European debt crisis before the formal policy decision was taken. As long as that type of information does not end up in Greenbook forecasts, the R&R approach will continue to identify shocks.

This points to an important distinction. What constitutes a shock in the different methodological approaches are fundamental different. While R&R attempts to proxy the central bank's information set, the HF approach relies on the financial market's information set to identify a shock. A shock in the R&R sense is a change in the policy rate which cannot be explained by the public information used to proxy the central bank's information set. A HF shock, on the other hand, is a change in the policy rate that financial market participants had not anticipated using their information set. While Greenbook forecasts capture important information for the central bank, and causal study of monetary policy on a certain macroeconomic variable require shocks to be orthogonal to all central bank information for that *particular* variable (see e.g. Cochrane's (2004) discussion of Romer and Romer (2004)), it is unlikely that such forecasts alone could proxy the full information set of the central bank. Compared to the financial market's information set which reflect all available information, the 14 variables in the standard R&R formulation is not likely able to capture all dimensions that go into the policy decision. Conditioning on a larger information set would therefore

likely be preferred when modeling the Riksbank reaction function.

2.3 Institutional Background and Riksbank Communication

Monetary policy in Sweden is conducted by the Riksbank. Since 1995, its main objective has been price stability, currently defined as a yearly inflation rate of around 2% (Sveriges Riksbank, 2026c)⁷. To achieve its policy mandate, the Executive Board of the Riksbank mainly decides on a policy rate which is implemented by steering overnight money-market rates through the Riksbank's operational framework and payment system. For more details about policy implementation, see Sveriges Riksbank (2025). While the policy rate remain the main policy tool, the Riksbank may also use unconventional measures such as engaging in direct trading of government bonds on the secondary market, and by providing liquidity to the financial system (Sveriges Riksbank, 2026a).

Policy decisions are reached by majority vote at scheduled monetary policy meetings, and the decision is typically announced the day after the meeting (Almerud et al., 2024). Board members are allowed to have deviating opinions and may express individual considerations, made public in meeting minutes published approximately five days after the policy announcement (Sveriges Riksbank, 2023). On the announcement day, at a pre-announced time, the Riksbank simultaneously releases the monetary policy decision, a press release and a report containing the data, analyses and forecasts that formed the basis for the decision (Sveriges Riksbank, 2026d). After the announcement, the Governor holds a press conference explaining the decision and answering questions from journalists (Sveriges Riksbank, 2023).

During my sample, the Riksbank communication framework has evolved and become more formalized and standardized over time. Up until 2007, the quarterly inflation reports were the main analytical communication tool which included forecasts, economic developments, inflation assessments and policy deliberations. Over the years they were used, these reports became more elaborate and communication gradually expanded (Svensson, 2010). Meeting minutes were published for every meeting after the 1999 reform (Svensson, 2010). Following the major 2007 reform, monetary policy reports and updates (MPRs and MPUs) replaced inflation reports, and press con-

⁷The inflation targeting was announced in 1993 and implemented in 1995. In 1993, inflation targeting was defined in terms of CPI. Because of its short term effects on household mortgage rates, the Riksbank has throughout the sample allowed the development of other price indices than the CPI to influence interest rate decisions. In 2017, the Executive Board officially changed the target variable for monetary policy to CPIF (CPI with a fixed interest rate). This did not reflect a change in how monetary policy was conducted, it rather clarified how the Riksbank interpreted their task of maintaining price stability (Sveriges Riksbank, 2017).

ferences were held after each policy meeting (Meyersson and Mikiver, 2018). With every press conference, the Riksbank publish a press release. Compared to inflation reports, MPRs contained more in-depth analyses for main and alternative policy scenarios. MPRs reported the forecasts included in the inflation reports but also a forecast of the Riksbanks own policy rate (Meyersson and Mikiver, 2018). The MPUs mainly focus on assessments of how new information affected the prospects for economic activity, inflation and monetary policy relative to scenarios presented in the previous MPR (Sveriges Riksbank, 2026d).

While the Greenbook forecasts used in the paper were published in inflation reports and MPRs, common for all three types of documents is that they are produced by Riksbank staff in collaboration with the Executive Board and provides the data, analyses and forecasts that form the basis for the monetary policy decision (Sveriges Riksbank, 2026d). In the rest of the paper, I refer to these documents jointly as *Riksbank documents*, or imply as *documents*.

3 Text Identification

The recent literature on shock identification has proposed several improvements on the standard approaches introduced in Section 2. For example, in order to proxy the systematic component of monetary policy in the R&R framework, researchers must use a larger set of variables than previously thought. Aruoba and Drechsel (2026) show that the narrative around numerical forecasts in Fed documents carry information about the systematic component which is crucial for the identification of monetary policy shocks. For the HF approach, it has been shown that Greenbook and private sector forecasts that pre-date announcements have predictive power on monetary policy surprises (Miranda-Agrippino and Ricco, 2021; Bauer and Swanson, 2023b). Given that Greenbook forecasts capture anticipated future movements of the economy and that the accompanying text describes central bank information (Aruoba and Drechsel, 2026), the narrative information around numerical forecasts is expected to be important for both proxying the systematic component of monetary policy and understanding movements on financial markets around policy announcements. This is especially true for Sweden as Riksbank documents are released together with the policy announcement (Sveriges Riksbank, 2026b).

For R&R and HF shocks, the remedy proposed to account for the documented predictability

differs. When Aruoba and Drechsel (2026) extract sentiment time series from Fed documents and include them as regressors in the R&R regression, Bauer and Swanson (2023a) simply regress HF monetary policy shocks on a set of variables with predictive power, taking the residual as a shock.

While keeping their conceptual structure intact, I extend the approach of Aruoba and Drechsel (2026) by incorporating deep learning and machine learning methods in the analysis of textual data to construct a set of regressors that capture the high-quality narrative information available in the Riksbank documents. This type of modern methods are preferred as they allow for a modeling of linguistic nuances that older tools are unable to capture. This new set of variables, together with standard Greenbook forecasts, are used to identify monetary policy shocks in both the R&R and the HF method. To the best of my understanding, this is the first paper to use deep learning methods to construct text-based regressors used for identifying R&R shocks and the first paper to use textual information to identify HF shocks.

3.1 Text-based R&R Shocks

To augment the R&R framework with text-based regressors, a corpus of relevant textual data must be defined. As the R&R regression is estimated at announcement day frequency, see Section 2.1, candidate sources of Riksbank text must be released on the day of the policy announcement. Riksbank documents and press releases meet that criteria, see Section 2.3. While the Riksbank document contains the information that formed the basis of the policy decision, press releases are essentially a very compressed version of the document (Sveriges Riksbank, 2026b; Sveriges Riksbank, 2026d). Recall that press releases are not available before the Riksbank 2007 reform. Based on this, the corpus of Riksbank text contains all Riksbank documents released over the sample.

The idea that these documents contain relevant information to the decision made by the Riksbank means, in the context of Equation (1), that the standard R&R shocks can be described by

$$\varepsilon_t^{RR} = \theta \mathbb{Z}_t + \varepsilon_t, \quad (6)$$

where θ is a coefficient vector, $\mathbb{Z}_t$ capture the high-quality textual information available in the Riksbank documents and ε_t is the non-systematic component of monetary policy. Since the Executive Board has access to the information in these documents at each meeting it can be shown that the

standard R&R shocks are predictable with $\mathbb{Z}$ at time t , see Section 5. This means that ε_t^{RR} are not orthogonal to the central bank information set (Ω_t), thus not true monetary policy shocks.

To identify monetary policy shocks, the information contained in the Riksbank documents must be projected from ε_t^{RR} . One way to do this is to include $\mathbb{Z}_t$ on the right hand side in the standard R&R Equation (2). This is the main idea in Aruoba and Drechsel (2026). In that paper, the authors parameterize $\mathbb{Z}_t$ by constructing time series of the sentiment used in discussion of manually identified economically meaningful concepts. Sentiment is there computed based on an extended version of Loughran and McDonald’s (2011) dictionary, see Aruoba and Drechsel (2026) for details.

I extend this idea by incorporating modern deep learning and machine learning methods in the analysis of the textual data to construct topic-sentiment time series. Different from Aruoba and Drechsel (2026), I identify topics, my analog of their economically meaningful concepts, with BERTopic which essentially is a data-driven clustering algorithm (Grootendorst, 2022). To quantify the sentiment within topics, I use CentralBankRoBERTa, a large language model (LLM) designed to classify the meaning of central bank language (Pfeifer and Marohl, 2023). Since written language is structured in sentences, I compute topics and sentiment at the sentence level.

One advantage of using models from the BERT-family in the analysis of textual data is that the starting point of such models is to produce a context-aware representation of input text (Gorodnichenko et al., 2023; Bertsch et al., 2025). This means that BERT models can differentiate the meaning of the word *not* in the two sentences $s_1 = \{\text{this is not good}\}$ and $s_2 = \{\text{this is not bad}\}$. This is possible by allowing the numeric representation of *not*, the embedding of *not*, to be different in the two sentences. The same fundamental principle is shared with more complex models such as OpenAI’s ChatGPT. Dictionary methods, and other bag-of-word models, are inherently unable to make this differentiation. As noted in Hansen et al. (2017), ”accounting for context with dictionary methods is practically very difficult”. For details on embeddings in BERT, see Appendix I.

Based on embeddings of the Riksbank documents, BERTopic identifies topics by clustering the embedding vectors while CentralBankRoBERTa predicts whether the embeddings are associated with positive or negative sentiment. The embeddings are the only input in the respective task. That is, no information about topic allocation enter the sentiment classification task, or vice versa. If this was not the case and, say, information about the classified sentiment of a sentence enter BERTopic as an input, the topic of that sentence would be identified, partly, based on classified sentiment

rather than solely on the content of the sentence.

3.1.1 Sentiment Classification

To determine the sentiment of each Riksbank sentence, I use CentralBankRoBERTa, a LLM classifier fine-tuned on central bank relevant labels (Pfeifer and Marohl, 2023). This model is preferred over dictionary methods since LLMs consistently outperform dictionary methods in sentiment classification across various domains (Frankel et al., 2022; Huang et al., 2023) and because CentralBankRoBERTa has been shown to outperform other popular LLM and machine learning models in central bank sentiment classification (Pfeifer and Marohl, 2023).

Compared to dictionary-based approaches for sentiment analysis, a LLM classifier offer a treatment of text that is fundamentally different in two main dimensions. First, the two different classes of models make assessments of sentiment based on different inputs. LLMs such as BERT models allow for a context-aware representation of text that dictionary methods do not. Since the meaning of words depend on the context in which they appear, the former representation better capture the linguistic nuances of written language. Second, based on the inputs, the two classes of models make assessments of sentiment based on different methods. For dictionary models, sentiment analysis is an exercise of counting the number of times words in pre-defined *positive* and *negative* dictionaries appear in a sentence (Loughran and McDonald, 2011). With LLMs, sentiment analysis is generally phrased as a classification problem.

In CentralBankRoBERTa, the model is fine-tuned to classify whether sentences are positive or negative, given a manually labeled training dataset (Pfeifer and Marohl, 2023). That is, a binary classification problem. Let $p_1 = p_{\theta^*}(y = +1|s_i)$ be the probability that sentence s_i is positive and $p_2 = p_{\theta^*}(y = -1|s_i)$ the probability that a sentence is negative under the optimal set of parameters, θ^* . Then, once trained, the CentralBankRoBERTa sentiment classification of an arbitrary sentence s_i is $\text{sentiment}(s_i) = 1 \cdot I(p_1 > p_2) + (-1) \cdot I(p_1 < p_2)$ where $I(\cdot)$ is an indicator function. Since p_1 and p_2 are continuous probabilities, the probability of the event $p_1 = p_2$ is zero. With this definition, $\text{sentiment}(s_i) \in \{-1, 1\}$. For details about the training procedure, see Appendix III.c.

The uncertainty of a prediction may carry important information about the meaning of a sentence as high uncertainty associated with a binary prediction may indicate the true class is neither positive nor negative. I therefore define a probability-weighted sentiment score by multiplying the

predicted label of each sentence with the associated model uncertainty:

$$sent_i = sentiment_i \cdot \mathbb{P}_i, \quad \mathbb{P}_i = \max\{p_1, p_2\}. \quad (7)$$

While CentralBankRoBERTa classifies the sentiment of a sentence conditional on a context-aware representation of the sentence and all model parameters, dictionary methods count words in pre-defined dictionaries. As such, sentiment classification in the latter method relies fully on the exhaustiveness and relevance of the chosen dictionaries. This means that the use of alternating phrases and the use of synonyms may distort sentiment classification. BERT-models do not share this dependency (see Leippold, 2023 for an illustration on financial sentiment classification).

3.1.2 Topic Analysis

Central bank documents naturally cover multiple different themes, ranging from monetary policy deliberations and inflation measures to discussions on financial stability and global economic outlooks. In this sense, a Riksbank document can be seen as an unstructured mixture of different themes. To extract more information from the textual data, it is common to attempt to uncover these underlying themes (Ash and Hansen, 2023). That is, separate from the question of how to measure sentiment of text is that of how to *organize* text into analytically meaningful representations of economic content. In the field of Natural Language Processing (NLP), such themes are called *topics* and models that uncover them are referred to as topic models.

Previous studies using central bank communication has identified topics by using either probabilistic models such as LDA (Hansen and McMahon, 2016; Hansen et al., 2017) and LSA (Boukous and Rosenberg, 2006), or by identifying the most frequent n -grams and manually selecting those that correspond to meaningful economic concepts (Aruoba and Drechsel, 2026). LDA and LSA both rely on a bag-of-words assumption, which means that the order of words in sentences does not matter (see Blei et al., 2003; Dumais, 2004 for more details). That is, regarding the impact of context, LDA and LSA treat text similar to dictionary methods.

While identifying topics based on n -gram frequency can highlight salient phrases it too ignores the semantic context in which words appear. This means that distinct uses of the same phrase across different discussions will be identified as the same topic. In contrast, different phrases that discuss

the same concept in different lexical forms will be put in different topics. An example of this are the bi-grams "price pressures" and "inflation dynamics" which relate to the same theme while being expressed differently. Additionally, such a method does also not scale well across domains. Therefore, similar to the recent paper of Silva et al. (2025), I use a flexible LLM approach for uncovering the topics in the textual data. Given its documented ability to uncover meaningful topics, Grootendorst's (2022) BERTopic is preferred.

Similar to CentralBankRoBERTa, BERTopic takes embeddings as inputs. While the embeddings used in CentralBankRoBERTa are learned based on the classification task, BERTopic uses a sentence embedding model that is learned separately from any down-stream task such as topic modeling. Importantly, these embeddings represent similar text with vectors that are close in vector space, why clustering on the embeddings are able to identify meaningful topics (Grootendorst, 2022). See Appendix III.d for details on the particular embedding model.

Once sentence embeddings has been computed for the Riksbank documents, each sentence is represented by a dense 384-dimensional vector. Since traditional distance, or similarity, measures are less meaningful in a high-dimensional spaces (Kriegel et al., 2009), BERTopic use McInnes et al.'s (2018) UMAP to project the embedding vectors to a lower-dimensional space. In this lower-dimensional space HDBSCAN, a general model that finds the optimal number of clusters (Campello et al., 2013), cluster the data. Since embeddings of similar sentences are 'close' in vector space, and since UMAP preserves much high-dimensional structure in a lower-dimensional representation, clusters in the lower-dimensional space are thought of as topics (Grootendorst, 2022).

Of particular importance in HDBSCAN is a parameter that controls the number of observations required to form a cluster. This makes the model a 'soft' clustering algorithm, in contrast to e.g. the popular k-means, in the sense that it does not force topic assignment. This allows treating sentences that cannot easily be allocated a cluster as outliers (Campello et al., 2013). I exclude sentences not assigned to a cluster from the subsequent analysis. The parameter controlling this threshold will prove a key component for defining optimal topic-sentiment series later on.

3.1.3 Estimation Approach

With information regarding topics assignment and sentiment classification of every Riksbank sentence, I compute topic-sentiment time series:

$$topic_sentiment_{i,t} = \begin{cases} \frac{1}{N_{i,t}} \sum_{j=1}^{N_{i,t}} sent_{j,t} & \text{if } N_{i,t} > 0, \\ 0 & \text{otherwise.} \end{cases} \quad (8)$$

Here, $sent_{j,t}$ is the sentiment score computed as in Equation (7) of a sentence j that appears in a Riksbank document at time t and $N_{i,t}$ are the number of sentences identified as part of topic i at time t . That is, the value of a topic-sentiment series at time t is the average sentiment among the sentences that appear in the particular topic at time t . Since there is no time-dimension in the topic modeling, nor the sentiment classification task, it means that $N_{i,t}$ varies freely over time. Together with the varying nature of language used within topics, this induce time variation in the topic-sentiment series stemming from both $N_{i,t}$ and $sent_{j,t}$.

To improve the empirical proxy for the systematic component of monetary policy, the topic-sentiment series are included as regressors in a text-based R&R regression model:

$$\Delta s = \mathbf{W}\beta + \epsilon^{RR-Text}, \quad \mathbf{W} = [\mathbf{X} \ \mathbf{Z}], \quad (9)$$

where $\mathbf{X} = [\mathbf{1} \ s_{t-1} \ \mathbb{X}] \in \mathbb{R}^{T \times p_x}$ where, in the baseline specification, $\mathbb{X}$ consists of the 12 variables described in Section 2.1. That is, $p_x = 14$. Analogously to the case with missing Greenbook forecasts, see Cloyne and Hürtgen (2016), for policy decisions taken without the release of a new Riksbank document I use the most recent value for each topic-sentiment series. Following from Equation (8), if a topic is not discussed in a document, the corresponding topic-sentiment series takes value zero. As a robustness check I consider alternative specifications of $\mathbb{X}$, see Section 5.

Based on the results presented in Aruoba and Drechsel (2026), there are reasons to challenge the assumption of linearity in the central bank reaction function. Following this, I include p_z linear and p_z squared topic-sentiment series in each $\mathbb{Z}_t \in \mathbf{Z} = [\mathbb{Z}_1, \mathbb{Z}_2, \dots, \mathbb{Z}_T]^\top \in \mathbb{R}^{T \times 2p_z}$ where p_z denotes the number of topics identified with BERTopic. The residual from estimating Equation (9), $\epsilon^{RR-Text}$, is taken as the text-based R&R monetary policy shock.

Recall the notion of noise in the clustering algorithm in BERTopic. A parameter in HDBSCAN determines the threshold for the smallest allowed cluster size, groups consisting of fewer vectors than this value are considered noise (Campello et al., 2013). Denote this parameter v . The larger v , the fewer topics are found which means that $p_z = p_z(v)$, $\mathbf{Z} = \mathbf{Z}(v)$ and in turn $\mathbf{W} = [\mathbf{X} \ \mathbf{Z}(v)] =$

$\mathbf{W}(v)$ in Equation (9) are all functions of v . Recall also that the goal with this analysis is to capture as much relevant information for the central bank reaction function from the Riksbank documents as possible. In this regard, what matters is explaining variation in $\Delta \mathbf{s}$ as well as possible. Therefore, the optimal value for the threshold parameter, v^* , can be defined as the value that generates topics for which the topic-sentiment series maximized the explanatory power in the R&R framework. I use the empirical setting to formulate a data-driven approach for finding v^* :

$$v^* = \arg \max_{v \in \mathcal{V}} R_{adj}^2 = \arg \max_{v \in \mathcal{V}} \frac{RSS(v)/(n - df)}{TSS/(n - 1)}, \quad (10)$$

where $RSS(v) = \|\Delta \mathbf{s} - \widehat{\Delta \mathbf{s}}(v)\|^2$ and $\widehat{\Delta \mathbf{s}}(v)$ are fitted values from Equation (9) that depend on v through $\mathbf{W}(v)$. That is, for each threshold value $v \in \mathcal{V}$ i) find $p_z(v)$ topics with BERTopic, ii) construct $\mathbf{Z}$ with the resulting $p_z(v)$ linear and $p_z(v)$ squared topic-sentiment series, iii) estimate the text-based R&R regression in Equation (9) and iv) compute R_{adj}^2 with the resulting estimates.

To have a large pool of potential values, I take $v \in \mathcal{V} = \{50, \dots, 150\}$ as candidate threshold values. Estimating BERTopic with $v = 50$ and $v = 150$ results in 348 and 59 topics, respectively, which suggest that an optimal number of topic-sentiment series to use in Equation (9), in a $\Delta \mathbf{s}$ -predictive sense, is $2p_z(v) \in \{118, 119, \dots, 796\}$. With the Greenbook forecasts, this means that $p_w(v) = p_x + 2p_z(v) \in \{132, 133, \dots, 810\}$ regressors will be used in search of v^* .

Since the sample spans $T = 167$ observations, I use shrinkage methods to estimate the text-based R&R regression. In this setting, a ridge penalty is preferred over alternatives such as the LASSO or Elastic Net. Compared to the LASSO, ridge favor a dense rather than a sparse model (Hastie et al., 2009). Given the goal of capturing as much relevant information for the Riksbank reaction function as possible, ridge, which can extract information from a large set of predictors, has clear advantages (see Giannone et al., 2021 for similar reasoning). In a similar application, Aruoba and Drechsel (2026) find that the more general Elastic Net often collapses to a pure ridge.

Formally, for Model (9), ridge solves the following problem for each threshold value v :

$$\hat{\beta}_v^{\text{ridge}} = [\mathbf{W}(v)^\top \mathbf{W}(v) + \lambda \mathbf{I}_{p_w(v)}]^{-1} \mathbf{W}(v)^\top \Delta \mathbf{s}, \quad (11)$$

where the optimal value of the shrinkage parameter, λ , is found using Leave-One-Out Cross-Validation (LOOCV). All variables in $\mathbf{W}(v)$ are standardized. From Equation (11), the projec-

tion matrix is defined as $\mathbf{H}_v = \mathbf{W}(v)[\mathbf{W}(v)^\top \mathbf{W}(v) + \lambda \mathbf{I}_{p_w(v)}]^{-1} \mathbf{W}(v)^\top$ and the fitted values as $\widehat{\Delta \mathbf{s}}(v) = \mathbf{H}_v \Delta \mathbf{s}$ which gives $RSS(v)$ in Equation (10). Total sum of squares is $TSS = \|\Delta \mathbf{s} - \overline{\Delta \mathbf{s}} \mathbf{1}\|^2$. Similar to the case with standard linear regression, the degrees of freedom can be computed by taking the trace of the projection matrix, $df = tr(\mathbf{H}_v)$. In penalized models like ridge, this quantity is called the *effective* degrees of freedom, because shrinkage reduces the functional flexibility of the model even as the number of parameters, $p_w(v)$, grow (Harrell, 2025). This means that $df < p_w(v)$ which, when used in Equation (10), produces an adjusted R^2 that accounts for the actual model complexity.

The value $v \in \mathcal{V}$ that generate topic-sentiment series which maximizes R_{adj}^2 is taken as the optimal value, v^* . From this iteration of the optimization algorithm, the optimal projection matrix $\mathbf{H}_{v^*}$, is used to compute fitted values $\widehat{\Delta \mathbf{s}}(v^*)$ following which text-based Romer & Romer monetary policy shocks are identified as $\hat{\varepsilon}^{RR-Text} = \Delta \mathbf{s} - \widehat{\Delta \mathbf{s}}(v^*)$.

3.1.4 Estimation Results

Figure 2 plots the curve $R_{adj}^2(v) \forall v \in V$. While $R_{adj}^2(v)$ is non-smooth in v , it is clear that it decreases in its argument. The reason is rather straight-forward. The higher the threshold v , the more likely it is that adjacent groups of sentences, which for lower values of v formed their own topic, be pooled together. That is, the larger the v , the more likely is the formation of larger groups of sentences that discuss *almost-similar* things. A direct consequence of this grouping of *almost-similar* sentences is that the intra-topic heterogeneity increases. This leads to topic-sentiment series that less clearly capture time variation in distinctly defined topics. As a result, the text-based regressors are jointly less informative about the behavior the Riksbank, which is reflected in the decreasing trend of $R_{adj}^2(v)$ over v .

The optimization algorithm selects $v^* = 59$ as the optimal threshold parameter. At this value, BERTopic identifies 234 topics, implying that the text-augmented specification is estimated using 468 topic-sentiment series in addition to the 14 Greenbook variables. Table 3 report results from the ridge estimation based on the optimal topic-sentiment series.

First, it can be noted that the ridge regularization applies considerable shrinkage, substantially limiting the effective model complexity. While the number of regressors increase by a factor of more than 30 relative to the Greenbook specification, the effective degrees of freedom increase

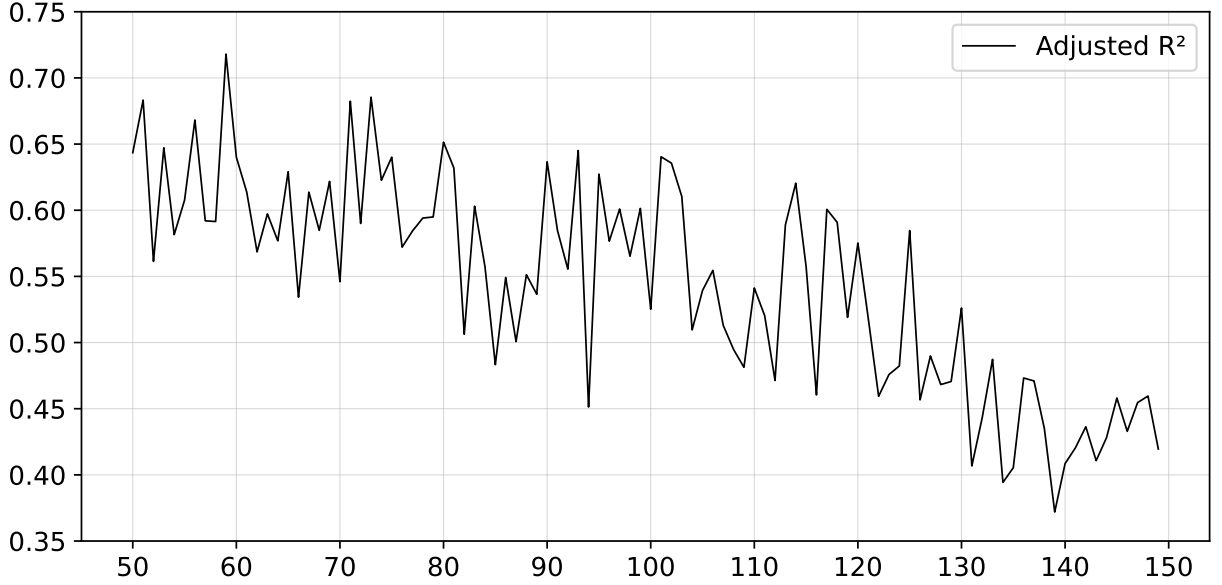


Figure 2: $R_{adj}^2(v)$ for candidate values $v \in \mathcal{V}$

slightly more than fourfold, from 14 to 57.6. Since the ridge regression is no model selection algorithm that shrinks coefficients to exactly zero, this means that the coefficients of many of the 482 variables in the model are shrunk to almost-zero. By investigating the topics for the highest loading topic-sentiment series, I verify that the most important topics for explaining variation in Δs have a clear economic interpretation and are of key relevance to monetary policy. The top 10 keywords, two representative sentences and topic interpretation for the 20 highest-loading topics used in this regression are presented in Table 12 and 13 in Appendix B.1.

My text-based specification provides substantial improvement in explanatory power over the standard Greenbook specification as adjusted R^2 increases from 0.329 to 0.718. This is also visible in a reduction in training mean squared error by approximately 30%. While improvements of in-sample fit is expected when additional regressors are included, the magnitude of improvement together with the considerable shrinkage imposed by the ridge penalty suggest that the written text in Riksbank documents contain information about the systematic component of monetary policy not fully reflected in the Greenbook forecasts, and the topic-sentiment series capture meaningful variation in these texts. In other words, in line with Aruoba and Drechsel (2026), the natural language in Riksbank documents provide information about the systematic component of monetary policy that goes beyond what is contained in numerical forecasts. The additional explanatory power

Table 3: Text-Based R&R Shock Estimation

Panel A. Text Representation		
Embedding dimension (v^*)		59
Number of "linear" text features		234
Number of squared text features		234
Number of R&R parameters		14
Total regressors		482
Panel B. Estimation Results		
	Standard R&R	Text-based R&R
Observations	167	167
Number of regressors	14	482
Effective degrees of freedom	14.0	57.6
Optimal ridge penalty (λ^*)	—	91.6
Adjusted R^2	0.329	0.718
Training MSE	0.0439	0.0302
LOOCV MSE	0.0595	0.0516
Reduction in LOOCV MSE (%)	—	13.28

Notes: Standard R&R refers to the specification in Section 2.1 where policy rate changes are regressed on Greenbook forecasts, $\mathbf{X}_t$. The text-augmented specification is based on the optimal set of topic-sentiment series and standard Greenbook forecasts, $W(v^*) = [X \ Z(v^*)]$. The ridge penalty parameter λ^* is selected by minimizing the leave-one-out cross-validation mean squared error (LOOCV MSE). Effective degrees of freedom are computed as $\text{tr}(X(X'X + \lambda I)^{-1}X')$.

of $\mathbf{W}$ over $\mathbf{X}$ is robust across different monetary policy regimes, and the highest-loading topics in each regime are all of key relevance to monetary policy, see Appendix C for a sub-sample analysis.

The improvements in R_{adj}^2 and training MSE reflects in-sample performance, but it is clear from a comparison of LOOCV MSE that the additional predictive power provided by the text-based regressors survive on out-of-sample (OOS) observations. While LOOCV MSE is larger than training MSE for both models, the differences do not suggest that the explanatory power of $\mathbf{W}$ over $\mathbf{X}$ is purely driven by overfitting the data. By comparing between models, it is evident that the inclusion of text-based regressors results in a decline from 0.0595 to 0.0516. While it is not massive, it does corresponds to an OOS improvement of approximately 13%. This suggest that a meaningful proportion of the information extracted from the Riksbank documents generalize beyond the estimation sample.

Since the goal with the estimation described in this section, similar to Aruoba and Drechsel (2026), is to explain as much in-sample variation in Δs as possible and not predicting future values of the policy rate given historical data, the results presented in Table 3 suggests that the resulting

residuals are informative as monetary policy shocks. These text-based monetary policy shocks, aggregated to quarterly frequency, are plotted together with $\hat{\varepsilon}_t^{RR}$ in Figure 3. Such shocks aggregated to monthly and yearly frequency are found in Figure 7 in Appendix A.3.

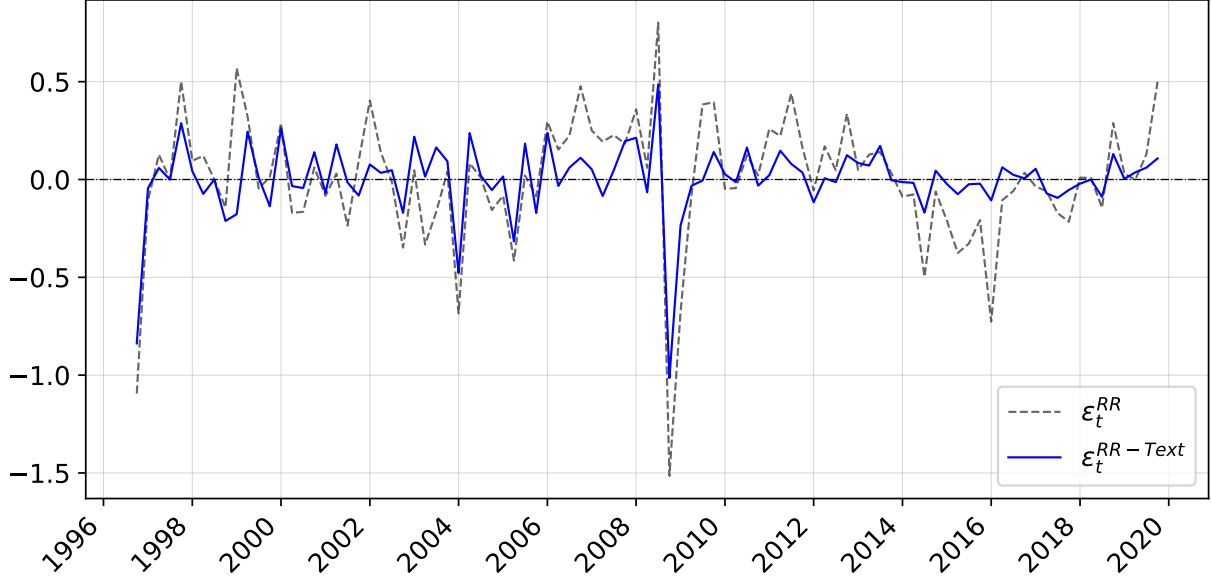


Figure 3: Standard and Text-based R&R Shocks at Quarterly Frequency

Since the text-based model explain more than twice the variation in Δs compared to the baseline R&R regression, the text-based shocks are smaller in comparison. This is consistent with the idea that movements in the policy instrument is mostly driven by the systematic component of monetary policy (Leeper et al., 1996; Ramey, 2016), and that the narrative in central bank documents contain valuable information about the central bank information set (Aruoba and Drechsel, 2026). As the differences between the shocks appear slightly larger in the period after 2014, it suggests that communication and forward guidance in this period was particularly able to explain policy decisions. This is verified by a sub-sample analysis where my model explains 87.2% of the variation in Δs in the post-2014 sample. The highest-loading topics in this regression capture topics discussing, among others, the unconventional monetary policy measures enforced by the Riksbank in this period. See Table 16 and 20 in Appendix C for more details. Further, the improvements in explanatory power documented for the main sample are of comparable magnitude to those reported for the United States, see Aruoba and Drechsel (2026).

However, $\hat{\varepsilon}_t^{RR-Text}$ is simply not a scaled-down version of $\hat{\varepsilon}_t^{RR}$. While the dashed line often

is larger in absolute value, there are many occasions where the two shock measures go in different directions. See for example in early 1999 and in the beginning of 2007 where $\hat{\varepsilon}_t^{RR}$ identifies a contractionary shock and the text-based shock instead identifies an expansionary shock. That is, conditional on the narrative information in the Riksbank documents the policy decision was more dovish than expected. In early 2003 and in the end of 2015 the opposite happens, $\hat{\varepsilon}_t^{RR}$ identifies an expansionary shock while $\hat{\varepsilon}_t^{RR-Text}$ instead finds contractionary shocks.

3.2 Text-based HF Shocks

A fundamental aspect of the HF identification is that financial markets are assumed efficient. This means that the market price of both short interest rates and broad stock indices at the end of the HF measurement window has incorporated the monetary policy decision. Since Riksbank documents are released at the same time as the monetary policy decision, the same market prices must also reflect the information available in these documents. For a measurement window of size $\pm\tau$ it means that $i_{t+\tau} = g(\Omega_t^{doc})$ and $P_{t+\tau} = h(\Omega_t^{doc})$ where $g(\cdot)$ and $h(\cdot)$ are functions that represent how financial markets incorporate the information available in the Riksbank document at announcement date t , Ω_t^{doc} . Since both Greenbook forecasts and the narrative information around them are shown to be important when explaining Δs , it suggest that both the independent and dependent variables in Equation (5) depend on $\mathbf{W}$ while it is omitted from the regression. A direct effect of this is that the coefficient on Δi_t^{PMSR} will suffer from omitted variable bias. Shocks identified as $\widehat{\Delta s}_t = \hat{\alpha} + \hat{\beta}\Delta i_t^{PMSR}$, may therefore fail to capture the 'true' monetary policy shock as the extent to which this bias is a problem is unknown ex ante.

Using the methodology for handling Riksbank documents developed in Section 3.1, I test the hypothesis of predictability of the information available in the Riksbank corpus⁸. In line with the literature on the predictability of monetary policy surprises, I also use the most recently available Riksbank document that pre-dates the current announcement. While focus in the paper has been on monetary policy surprises purged with the PMSR filter, I will also test the predictability of pure monetary policy surprises, following Miranda-Agrippino and Ricco (2021), and Bauer and Swanson (2023a). This is done by estimating different versions of the following model:

⁸Here I use the Riksbank documents, the same corpus as in Section 3.1. See Section 2.3 for more details on Riksbank communication.

$$\mathbf{Y} = \mathbf{A}(v)\boldsymbol{\beta} + \boldsymbol{\eta}, \quad (12)$$

where $\mathbf{Y}$ is either $\Delta \mathbf{i}$ or $\Delta \mathbf{i}^{PMSR}$ and $\mathbf{A}(v)$ contain either Greenbook forecasts ($\mathbf{X}$), topic-sentiment series ($\mathbf{Z}(v)$), or both. When the Greenbook forecasts are the only regressors, I estimate the model with OLS. In cases where topic-sentiment time series are included as regressors, the model is fit with a ridge penalty on the coefficients and the optimal topic-model threshold parameter is found via the optimization algorithm described in Equation (10). Table 4 report the adjusted R^2 from the different regressions.

Table 4: Predictability of Monetary Surprises

Variables	$\Delta \mathbf{i}$ (%)	$\Delta \mathbf{i}_t^{PMSR}$ (%)
Greenbook Forecasts	0.110	8.180
Topic-Sentiment Series	0.809	14.198
Forecasts + Topic-Sentiment Series	0.815	15.109
Greenbook Forecasts ($T - 1$)	-0.321	2.862
Topic-Sentiment Series ($T - 1$)	0.940	8.474
Forecasts + Topic-Sentiment Series ($T - 1$)	0.968	10.121

Notes: The results in the table are reported in terms of adjusted R^2 in percentage points. For the regressions involving text-based regressors, adjusted R^2 is computed based on the effective degrees of freedom, defined as $\text{tr}(X(X'X + \lambda I)^{-1}X')$. For regressions involving Greenbook forecasts only, adjusted R^2 is computed based on model degrees of freedom.

While regressing HF surprises on data released at the beginning of the announcement window is not the type of predictability discussed in Miranda-Agrippino and Ricco (2021), and Bauer and Swanson (2023a), it is important for recovering monetary policy shocks when using the PMSR to disentangle policy surprises from information surprises. From the third column of Table 4 it is shown that 15% of the variation in the monetary surprise instrument is driven by forecasts and topic-sentiment series that also explain the policy decision. It is, however, not possible to explain variation in market interest rates with similar information. This suggest that the co-movement on financial markets, that is the PMSR filter, can be partly explained explained with information from Riksbank documents. Once narrative information in included in the model, the additional value of the Greenbook forecasts is marginal.

Using lagged values, in line with the previous literature on predictability, I find similar effects. Neither Greenbook forecasts nor topic-sentiment series are predictive of monetary policy surprises, $\Delta \mathbf{i}_t$. This is expected for efficient markets. However, when those quantities are purged from central

bank information surprises with the PMSR filter, it is evident that especially the lagged narrative information captured with the topic-sentiment series are important predictors for current movements. Information from the last available Riksbank document systematically explain around 10% of the variation in the instrument used to identify monetary policy shocks.

To account for the predictability of contemporaneous Greenbook forecasts and topic-sentiment series, and purge the instrument for monetary policy surprises from these variables, I take the residuals $\boldsymbol{\eta} = \widetilde{\Delta \mathbf{i}}$ from regressing $\Delta \mathbf{i}^{PMSR}$ on $\mathbf{W}(v^*) = [\mathbf{X} \mathbf{Z}(v^*)]$ as text-based monetary policy surprises. In this setting, the optimization algorithm developed in Section 3.1 identifies $v^* = 80$ as the optimal threshold parameter for HDBSCAN in BERTopic, which gives 187 linear and squared topic-sentiment series in $\mathbf{Z}(v^*)$. This means that a total of 388 regressors was used in this specification of Equation 12. The highest loading topics all have clear economic interpretation and capture relevant themes for Swedish monetary policy, see Table 14 and 15 in Appendix B for details. For additional results, as well as the curve $R_{adj}^2(v) \forall v \in \mathcal{V}$, see Figure 6 and Table 11 in Appendix A. The results from the text-based regressions presented in this section are robust to the estimation sample, though slightly weaker in the earliest part of the sample, see Table 16 in Appendix C for estimation results in sub-samples, and Table 21-24 for the corresponding highest loading topics.

In $\widetilde{\Delta \mathbf{i}}$ I have an instrument for monetary policy surprises that capture unexpected variation in market rates around policy announcements. Compared to $\Delta \mathbf{i}^{PMSR}$, my text-based instrument is purged of the correlation with Riksbank forecasts and the narrative information surrounding those forecasts. Next, I identify text-based HF monetary policy shocks as the fitted values, $\widehat{\Delta s_t}$, from

$$\Delta s_t = \mu + \theta \times \widetilde{\Delta \mathbf{i}_t} + \psi_t. \quad (13)$$

The results of this estimation are reported in Table 5. Compared to the standard HF case, I find that text-based HF shocks are slightly more common.

The text-based HF shocks are plotted at quarterly frequency together with the standard HF shocks in Figure 4. Such shocks aggregated to monthly and annual frequency are found in Figure 8 in Appendix A.3.

Both $\hat{\varepsilon}_t^{HF}$ and $\hat{\varepsilon}_t^{HF-Text}$ identify roughly the same amount of variation in the policy rate as a shock and the correlation between them is 0.968. The general pattern is that both approaches iden-

Table 5: Estimation of High-Frequency Monetary Policy Shocks

Variable	Estimate
$\widetilde{\Delta i_t}$	2.551*** (0.458)
α	-0.0252 (0.019)
Observations	140
R^2_{adj}	0.179

Notes: The dependent variable is the change in the policy rate. Standard errors are reported in parentheses. The explanatory variable Δi_t^{PMSR} denotes the high-frequency monetary policy surprise around policy announcements, filtered with the PMSR filter.

tify shocks of similar direction but of different magnitude. Except for a few policy announcements, the text-based HF approach identifies a smaller shock.

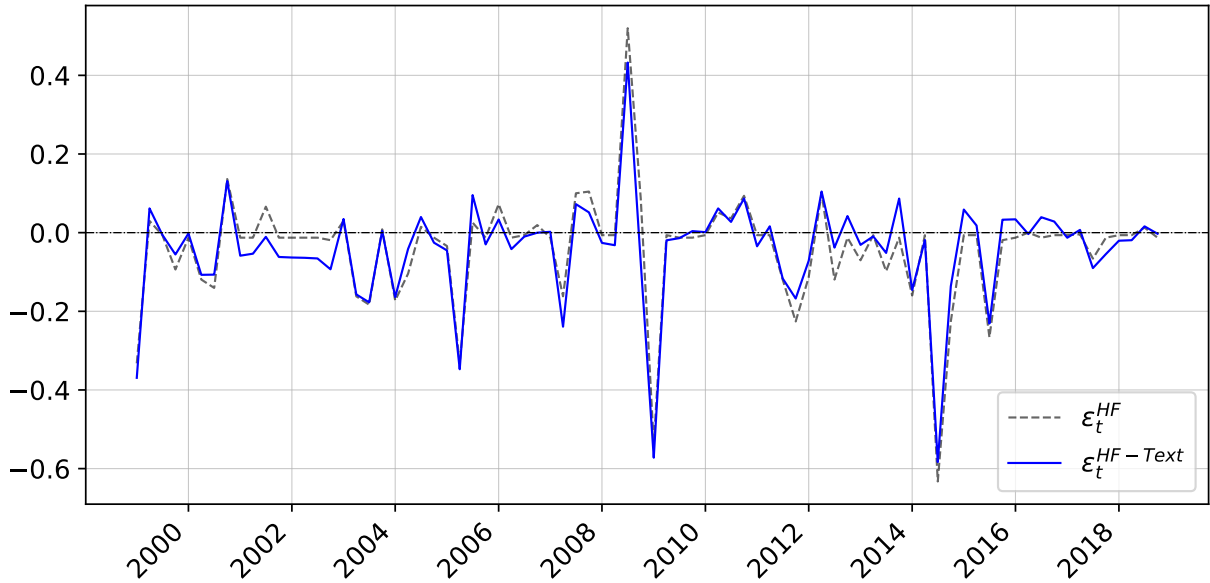


Figure 4: Standard and Text-based HF Shocks at Quarterly Frequency

3.3 Methodological Considerations

The proposed NLP pipeline, where BERTopic identifies topics and CentralBankRoBERTa classifies sentiment of Riksbank sentences, is applied uniformly across the Riksbank reports. One concern may be that since the Riksbank's communication framework has evolved over the sample, see Section 2.3, the corpus of Riksbank documents may also vary over time. In particular, if there are substantial differences in what, or how, content is discussed in the different types of documents, the

empirical model developed in Section 3.1 would need to account for that. If such heterogeneity is present in the corpus, the unsupervised topic model may base topic classification on regime-specific document characteristics rather than economically meaningful content⁹. To verify whether this issue is present in the Riksbank corpus, I report regime-by-regime descriptive statistics for the textual data, see Table 25 in Appendix D. Despite institutional changes in the communication framework, the language used in all documents appear standardized in terms of sentence- and document length and the share of sentences allocated a topic. Moreover, the 20 most prevalent topics identified with BERTopic are stable across communication regimes and capture key macroeconomic concepts such as inflation, labor markets, economic outlooks and policy discussions, see Table 27 in Appendix D. Together, these results indicate that there are no substantial heterogeneity in the textual data, justifying the uniform application of BERTopic and CentralBankRoBERTa across the corpus.

Regarding the models used in the NLP pipeline, it can be noted that while LLMs offer a conceptually more appealing modeling of the textual data over dictionary methods, the "black box" nature of such models might raise concerns. First, the training and fine-tuning of LLMs require large amounts of data. If a LLM is presented with data it saw in training, any resulting representation, classification or clustering would not be a result of the model architecture but rather a result of memorization. Since CentralBankRoBERTa is trained on central bank *speeches* in general, and fine-tuned on labeled *Federal Reserve speeches* in particular (Pfeifer and Marohl, 2023), it is unlikely that such a concern would be relevant for the sentiment analysis of the Riksbank documents. The embedding model in BERTopic, AllMini-LM-L6-V2, is fine-tuned to predict whether, given a pair of sentences, one sentence follow from the other using a dataset of more than 1 billion sentence-pairs (Wang et al., 2020). This training data mostly contain question-answer type of sentence pairs from internet forums such as Stack-Exchange, WikiAnswers and Reddit, none of which are likely to be contaminated by Riksbank documents.

Another concern is of temporal nature. Since a LLM sentiment classifier by construction generalizes the sentence-label pattern it saw in training, the information that enter the labeled training data is generalized to the sentiment classification of the Riksbank documents. If the training data was constructed such that current information about historical events influenced the labeling of

⁹For example, if the bulk of the inflation reports discuss concepts that become obsolete with changing macroeconomic conditions, or if the syntactic style evolves drastically with time, the embeddings of Riksbank sentences will reflect such characteristics. Since BERTopic identifies topics by clustering high-dimensional data in a space where direction represent semantic meaning, this type of regime-specific characteristics will introduce variation in embedding vectors unrelated to the economic content which may influence topic identification.

those historical events, the classifier is in fact using $t + h$ information when making assessments of the meaning of a sentence at time t . Take as an example the booming US housing market in 2007. Although it is well known that this was the catalyst of the global financial crisis, in 2007, a reporting about a booming housing market was likely good news. Hence, if current information about the 2007 reporting would influence the labeling of that reporting when constructing the training data, the model is making the above mentioned mistake. Importantly, if a sentence included in $\mathbb{Z}_t$ in, say, the text-based R&R regression is classified like this, estimating Equation (9) effectively projects Ω_{t+h} onto Δs_t which obscures the identification of monetary policy shocks.

The training data used for CentralBankRoBERTa’s sentiment classifier *describes* the meaning of a sentence for a particular economic agent (Pfeifer and Marohl, 2023). That is, the training label does not capture a *prediction* of the effect of the monetary policy communicated in an arbitrary training sentence. Since there was no information of time in the training data (see Pfeifer and Marohl, 2023), and since we pass no time-dimension through the model, the resulting sentiment classification do not condition on the temporal dimension. That is, the sentiment classifier is unaware of the time dimension. Visual inspection suggest that there are no systematic differences in the sentiment classification in the sample over time.

Note that however unlikely, if indeed $t + h$ information contaminates the sentiment classifications at t , it means that the classifier would ”correct” the t classification using superior $t + h$ information. The resulting shocks would in this case understate the true shocks.

Similar to CentralBankRoBERTa, BERTopic takes no time dimension into account when representing the Riksbank sentences as sentence embeddings that capture *semantic* and not *temporal* similarity. After the embedding step, BERTopic identifies topics by performing a clustering on the sentence embeddings. Therefore, such a temporal concern are not present in the topic modeling.

4 Macroeconomic Responses to Monetary Policy Shocks

To study the effect of the text-based monetary policy shocks on macroeconomic aggregates, I estimate impulse response functions (IRFs) with local projections (LPs), in line with Holm et al. (2021), Kilman (2022), and Andersen et al. (2023). While both LPs and VARs are commonly used in the literature to estimate IRFs, it can be shown that LPs has favorable finite-sample properties

such as being more robust to misspecification and the choice of lag length (Olea et al., 2025)¹⁰.

LPs estimate direct projections of the dependent variable at horizon $t + h$ on the shock that occur at time t along with an additional vector of controls. Following Jordà (2005), I estimate

$$y_{t+h} = \mu_h + \phi_h^{LP} \hat{\varepsilon}_t^i + \sum_{p=1}^P \psi_{p,h} X_{t-p} + \xi_{t,h} \quad (14)$$

for horizon $h = 0, 1, \dots, 12$ in quarterly data and $h = 0, 1, \dots, 5$ in yearly data. X_t denote control variables and I consider all four estimated shock series, $i \in \{RR, RR - Text, HF, HF - Text\}$. The choice of X_t is data-driven, obtained by estimating auxiliary VAR models of each outcome on the shock and lags of both variables, and by using AIC to select the lag length (Olea et al., 2025). This procedure selects two lags of the outcome and shock for the quarterly data, and one lag of the outcome and shock in yearly data¹¹. The results presented here are not sensitive to the selected lag length in either shock nor the outcome, see Section 5 for alternative specifications in the LP. In the interest of space, impulse responses estimated at annual frequency are reported in Appendix A.4.

The text-based monetary policy shocks are estimated quantities that depend on $\hat{\beta}_{v^*}^{ridge}$ and on the data-driven choice of v^* . That is, they are machine-learning generated regressors coming from a first-stage that is itself a model-selection procedure. When used in a second-stage estimation, here as impulse variables in local projections, this introduce the well-known generated regressor problem (see e.g. Chernozhukov et al. (2018) and Battaglia et al. (2024)). LPs treats the impulse variable as observed, not estimated, which means that the resulting standard errors will understate the true uncertainty around the impulse responses. That is, LP standard errors ignore the uncertainty in the text-based shocks that originate from the ridge estimation and the selection of v^* .

To account for this, I implement a bootstrap procedure that re-estimates the first-stage regression for each bootstrap replication, generating for each replication a new realization of the text-based shock, for both R&R and HF type shocks. The bootstrap text-based shock is used as impulse variable in a local projection of each outcome variable and horizon at both quarterly and annual frequency. To preserve the local dependence in the data, I use a moving block bootstrap

¹⁰A disadvantage with local projections are that the estimation condition of $H - 1$ leading, in addition to P lagged, observations which means that a smaller sample is effectively used to estimate response coefficients in LPs compared to VARs.

¹¹There are minor variation in the optimal lag lengths across the different variables at quarterly frequency (ranging from one to three lags in either outcome or shock). The choice of two lags each for outcome and shock is therefor a compromise, see Holm et al. (2021) and Kilman (2022) for similar reasoning. At yearly frequency, one lag in each outcome and shock is selected uniformly for all outcome variables. I test the robustness of the results to the local projection specification in Section 5.

when constructing bootstrap samples (Künsch, 1989; Lahiri, 2013). To reflect short-run persistence in macroeconomic conditions and policy decisions, I chose 12 consecutive observations as block length¹². Inference for the text-based monetary policy shocks are based on the resulting bootstrap distribution of impulse responses¹³. For more details on the bootstrap implementation, see Appendix E. In line with the previous literature I conduct inference based on Newey-West standard errors for the standard R&R and HF shocks.

I consider the four outcome variables mainly used in the empirical literature on the transmission of monetary policy, the policy rate, the unemployment rate, output and price level. I consider real and seasonally adjusted variables, see Table 6 in Appendix A for details. Throughout this section, I illustrate percentage responses in these four outcome variables to a one percentage point (100 basis points) contractionary monetary policy shock. The left column plot responses to R&R shocks and HF shocks are plotted to the right. I plot 95% confidence intervals for all IRFs.

4.1 Impulse Responses

Impulse responses estimated at quarterly frequency is presented in Figure 5, for responses in annual data see Appendix A.4. The responses differ across the two types of shocks, and there are important differences between the text-based shock and the standard shock within each type.

The policy rate increases on impact in response to a contractionary shock in all four shocks. This result, and the reported magnitude, is consistent with related papers in the Swedish setting, see Kilman (2022), Amberg et al. (2022), and Almerud et al. (2024)¹⁴. Responses to both text-based shocks at further horizons suggest a mean reverting behavior on behalf of the Riksbank as the policy rate significantly decreases in the first and fourth quarter, for $\hat{\varepsilon}_t^{HF-Text}$ and $\hat{\varepsilon}_t^{RR-Text}$, respectively. A similar pattern is found in both Swedish (Amberg et al., 2022; Almerud et al., 2024) and Norwegian (Holm et al., 2021) settings.

The responses in prices and unemployment present some puzzles. A contractionary shock in

¹²This represents 1.5 – 2 years of meetings for most of the years in the sample. I have verified that the choice of block length is not driving the conclusions that may be drawn from my LP estimations.

¹³This bootstrap implementation accounts for uncertainty associated with the estimation of ridge coefficients, the selection of regularization parameter, the aggregation of shocks to lower frequencies as well as sampling uncertainty in the underlying time series. Due to computational limitations, I am unable to propagate the uncertainty in selecting v^* into the confidence bands in the LP. For further discussion of this, and additional details on the bootstrap implementation, see Appendix E

¹⁴Aggregate responses in Amberg et al. (2022) are found in their online appendix. Both Amberg et al. (2022) and Almerud et al. (2024) use a VAR model to estimate responses to changes in the policy rate, the reported responses in the policy rate is therefore mechanical.

$\hat{\varepsilon}_t^{RR}$ and $\hat{\varepsilon}_t^{RR-Text}$ cause prices to increase on impact. While both shocks illustrate gradual decreases in subsequent quarters, only the response to $\hat{\varepsilon}_t^{RR-Text}$ suggest responses in line with theory as prices decrease significantly after five quarters. Less the increase on impact, the response on prices from $\hat{\varepsilon}_t^{RR-Text}$ is similar to the dynamics and magnitude reported in Almerud et al. (2024). The standard R&R shock decreases steadily, but is never significantly negative, consistent with Kilman (2022). This price puzzle is not present for either R&R-type shock in annual data, see Figure 9.

The response of unemployment to the R&R shocks is slightly negative in the initial quarter which is a small puzzle, see Kilman (2022) and Amberg et al. (2022) for similar results. The initial response does however vanish within the first quarter and presents the expected increase in unemployment after two quarters¹⁵. The peak response of +0.488% occur after seven quarters, a magnitude comparable to Amberg et al. (2022), however significant at the 95% level. The on-impact puzzle is not present in the annual data for either $\hat{\varepsilon}_t^{RR}$ or $\hat{\varepsilon}_t^{RR-Text}$, see Figure 9. Output behaves as expected as a contractionary policy shock decrease economic activity. The effect of the text-based shock is significant for a shorter period of time than the standard R&R shock, which suggest that output may be more resilient than previously thought. The effect reaches its trough after five quarters at -2.613% , a magnitude comparable to Kilman (2022) and Almerud et al. (2024)¹⁶.

The price puzzle observed for the R&R-type shocks are also prevalent for the HF shocks. CPI increases on impact to both impulses after which the price level decreases. The response to the standard HF shock is however significant in the sixth quarter which is contrary to prior expectations. While the response to $\hat{\varepsilon}_t^{HF-Text}$ is never outside the 95% confidence interval around zero after the third quarter, the point estimate suggest a persistency of the price level over the estimation horizon. Unemployment increases with a similar magnitude and delay in response to both HF shocks as to the R&R shocks, though not as smooth as the latter¹⁵. Output behaves as expected, but it can be noted that a shock from $\hat{\varepsilon}_t^{HF}$ and $\hat{\varepsilon}_t^{HF-Text}$ transmit to GDP faster than the R&R shocks. While economy takes as much time to recover from a contractionary shock in $\hat{\varepsilon}_t^{HF}$ as for the standard R&R shock, the response to $\hat{\varepsilon}_t^{HF-Text}$ is more persistent than its R&R based counterpart. The response to $\hat{\varepsilon}_t^{HF}$ achieves its trough at -6.01% after four quarter which is large compared to the related literature, although accompanied by much uncertainty reflected in the wide confidence bands. Peak response

¹⁵I have verified that the responses of both $\hat{\varepsilon}_t^{RR}$ and $\hat{\varepsilon}_t^{HF}$ level out to 0 within 16 quarters.

¹⁶The aggregate response to the shock used in Amberg et al. (2022) also reach its trough around this time. While my results display similar dynamics as Amberg et al. (2022), a comparison of magnitudes is not feasible as they use an industrial production index to proxy economic activity.

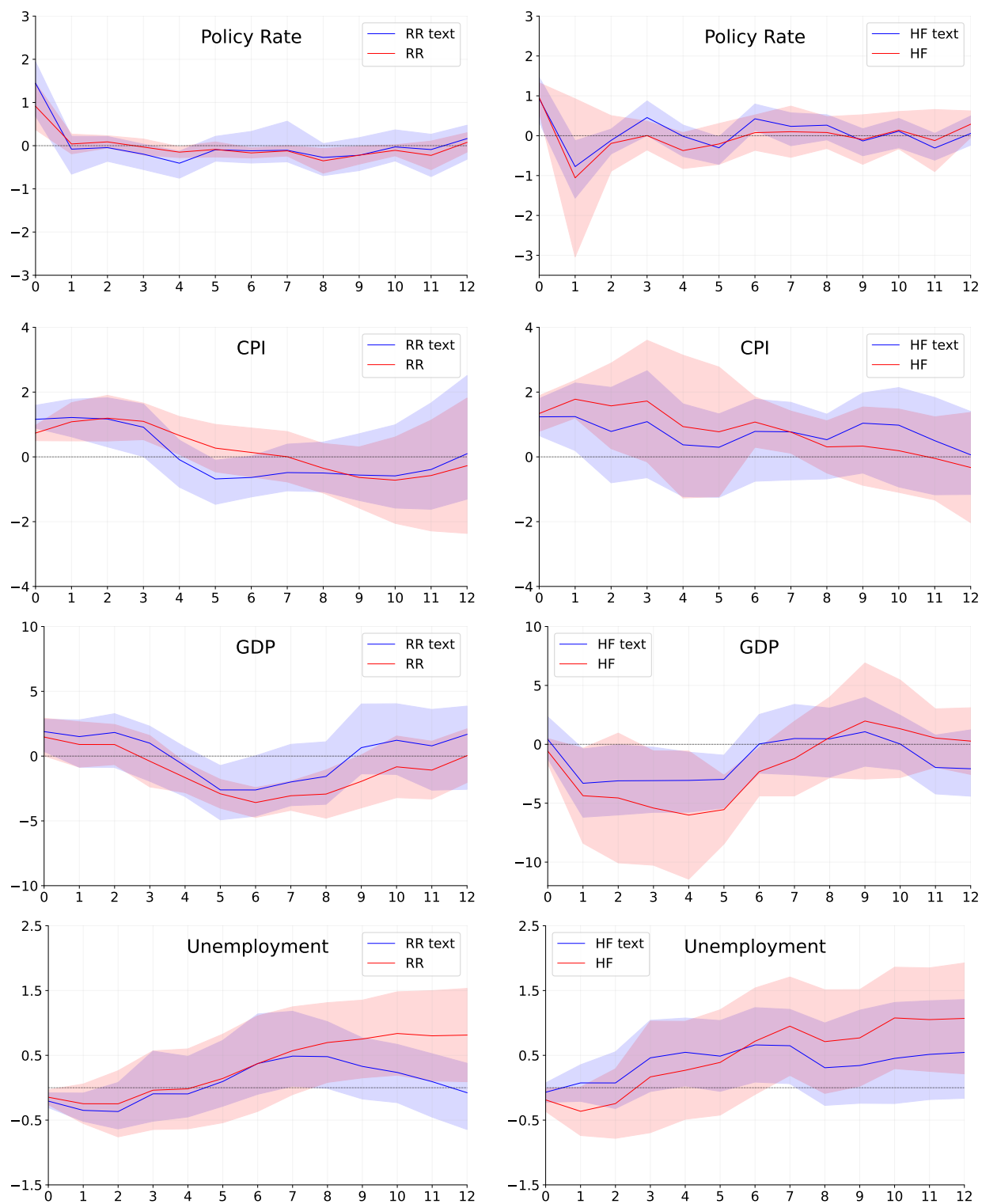


Figure 5: Impulse Response Functions Estimated at Quarterly Frequency

Note: Red shaded areas denote standard 95% confidence intervals, blue shaded areas denote 95% bootstrap confidence intervals. All specifications include two lags of the dependent variable and two lags of the shock.

to $\hat{\varepsilon}_t^{HF-Text}$ is -3.31% , which is more in line with the previous literature.

In sum, though the price level increases on impact, the results related to $\hat{\varepsilon}_t^{RR-Text}$ most resemble 'textbook' responses to a monetary tightening as economic activity falls following the shock which puts downwards pressure on prices which together transmit to an increase in unemployment.

5 Robustness Analysis

The method outlined in Section 3.1 for using topic-sentiment time series to identify monetary policy shocks was, for both the R&R and HF approach, a two-step approach. For the former, I included the variables as components in the central bank reaction function, taking the part of monetary policy that is uncorrelated with economic forecasts and narrative information as monetary policy shocks. In the HF framework, I found that the PMSR filter was correlated with information available in Riksbank documents, both the Greenbook forecasts and the narrative surrounding these forecasts. I therefore defined an improved instrument for monetary policy surprises that is orthogonal to such information, and use this new instrument to identify monetary policy shocks. An alternative approach is to project out the part of the standard monetary policy shock that correlate with the topic-sentiment series directly. Such an approach is similar to Bauer and Swanson (2023a) who remove the component of monetary policy surprises (no PMSR) that correlate with economic and financial data that pre-date the announcement. For $i \in \{RR, HF\}$, I run the following regressions:

$$\hat{\varepsilon}^i = \mathbf{W}(v)\beta + v, \quad (15)$$

again using the methodology for finding v^* outlined in Section 3.1, and take the residuals as monetary policy shocks. I refer to these shocks as $\hat{\varepsilon}_t^{RR-TextAlt}$ and $\hat{\varepsilon}_t^{HF-TextAlt}$. In the standard case, text-based regressors explain 20.2% of the variation in $\hat{\varepsilon}_t^{RR}$ and 15.8% of the variation in $\hat{\varepsilon}_t^{HF}$, for information about estimation, see Table 10 in Appendix A. The alternative text-based shocks capture much of the same variation as the standard R&R and HF shocks, as well as the text-based shocks identified in Section 3. $\hat{\varepsilon}_t^{RR-TextAlt}$ has a correlation of 0.87 and 0.968 with $\hat{\varepsilon}_t^{RR-Text}$ and $\hat{\varepsilon}_t^{RR}$, respectively. The correlation between $\hat{\varepsilon}_t^{HF-TextAlt}$ and the text-based and standard HF shock is 0.979 and 0.96, respectively. See Table 8 in Appendix A for details. Impulse responses for these

alternative shocks are reported in Figure 10 and 11 in Appendix F.1 for quarterly and annual data, respectively. Overall, the responses to the shocks identified with this alternative text-approach are consistent with the results presented in Section 4.

At quarterly frequency, the responses to $\hat{\varepsilon}_t^{RR-AltText}$ and $\hat{\varepsilon}_t^{HF-AltText}$ display similar dynamics and the response for the alternative shocks only appear outside the 95% confidence interval associated with $\hat{\varepsilon}_t^{RR-Text}$ and $\hat{\varepsilon}_t^{HF-Text}$ for the unemployment rate. For the R&R-type shocks, the point estimates of impulses from $\hat{\varepsilon}_t^{RR-AltText}$ at deviating horizons appear between those of the standard and text-based shock, so the overall conclusion does not change. Responses on early horizons to $\hat{\varepsilon}_t^{HF-AltText}$ suggest a small puzzle similar to that documented for the R&R-type shocks in Section 4.1. At annual frequency, the responses to $\hat{\varepsilon}_t^{RR-AltText}$ and $\hat{\varepsilon}_t^{HF-AltText}$ are never outside the 95% confidence interval of the responses associated with $\hat{\varepsilon}_t^{RR-Text}$ and $\hat{\varepsilon}_t^{HF-Text}$, respectively. That is, the results presented in Section 4.1 are not sensitive to how I introduce narrative information from Riksbank documents in either the R&R or HF frameworks.

It has been shown that the choice of controls and number of lags included in the local projection model are important for both the estimated responses (Miranda-Agrippino and Ricco, 2021) and the properties of the estimator (Olea et al., 2025). Figure 12 and 13 in Appendix F.2 presents impulse responses from $\hat{\varepsilon}_t^{RR-Text}$ and $\hat{\varepsilon}_t^{HF-Text}$ estimated in local projection models with different lag lengths in the outcome compared to the main specification. For quarterly data, I consider four and eight lags (the main specification uses two), and for the annual data I consider two and three lags (the main specification uses one). In Figure 14 and 15 in Appendix F.2 I include additional lags of the shock in the local projection of $\hat{\varepsilon}_t^{RR-Text}$ and $\hat{\varepsilon}_t^{HF-Text}$ on the monetary aggregates. I consider four and eight lags in the quarterly data and two and three lags in yearly data. In both cases, the responses are near identical to those presented in Figure 5 and 9 in Section 4 and in instances where the responses differ, I draw the same general conclusions about the effect of monetary policy on the economy.

As a third robustness test, I consider alternative specifications of the R&R Greenbook forecast approach. That is, alternative specifications of $\mathbf{X}$ in Equation (9) and (12). To differentiate between different specifications, I refer to the $\mathbf{X}$ used in the main text as $\mathbf{X}^{(v1)}$.

First, I allow for a possibly different reaction function in the period of negative policy rates by adding interactions of all variables in $\mathbf{X}^{(v1)}$ with a dummy defined in periods where the policy

rate is below 0.01. Here $\mathbf{X}^{(v2)} = [\mathbf{X}^{(v1)} d * \mathbf{X}^{(v1)}]$. Second, the Riksbank started publishing forecasts for the policy rate in 2007. I include the forecast for the next year's policy rate, $\tilde{s}$, as a regressor, $\mathbf{X}^{(v3)} = [\mathbf{X}^{(v1)} \tilde{s}]$, and use the sample from 2007 and onward. Third, I include the lag of the nominal SEK/USD exchange rate as a control. I use $\mathbf{X}^{(v4)} = [\mathbf{X}^{(v1)} \text{SEK/USD}]$ where *SEK/USD* is the dollar price in Swedish krona the day before the policy meeting.

Table 9 in Appendix A report estimation results from using the alternative Greenbook specifications. Figure 16 and 17 in Appendix F.3 present impulse responses using text-based shocks defined with $[\mathbf{X}^{(v1)}, \dots, \mathbf{X}^{(v4)}]$. In these figures it is clear that the results presented in the main text are not sensitive to the specification of the Greenbook forecasts.

6 Conclusion

In this paper I use data from 1996-2019 to identify R&R and HF monetary policy shocks for Sweden. Building on Aruoba and Drechsel (2026), I develop a method that rely on modern tools for textual analysis to construct regressors that capture the rich information in official documents available to the Riksbank ahead of each policy meeting. Modern deep learning methods are preferred as they allow for a modeling of linguistic nuances that older methods are unable to capture. Such text-based regressors are crucial for the identification of monetary policy shocks as they systematically predict shocks identified with standard methods. Text-based information should therefore be included both when constructing a proxy for the Riksbank information set, and when instrumenting high-frequent monetary policy surprises. If not, monetary policy shocks are not identified.

I define the part of the standard shock that is uncorrelated with my text-based regressors as text-based monetary policy shocks, and study the dynamic responses to all shocks with local projections in quarterly and annual data. Text-based R&R shocks display results that are mostly in line with 'textbook' responses as prices and output declines following a contractionary shock, while unemployment increases. While all shocks cause a price puzzle in the quarterly data, only the response to the text-based R&R shock is significantly negative over the projection horizon. A prize puzzle is still present for both types of HF shocks in annual data. All results are robust to an alternative implementation of the text-based regressors, the choice of lag length in the LP, and the control variables used in both the LP and the identification of the text-based monetary policy shocks.

In general, my results show that text-based information from central bank documents is crucial to identify monetary policy shocks. That is, the information set researchers could, and should, condition on when disentangling variation in policy rates that are systematic responses to economic conditions from non-systematic variation, is larger than previously thought.

In a broader perspective, the developed methodology for constructing and using text-based regressors may be extended to other empirical applications. For example, including speeches from the European Central Bank or the Federal Reserve would allow constructing text-based HF shocks for these large economies at speech-frequency, rather than meeting-frequency. Future research may also use a similar method to identify what central bank communication is 'central bank information' in Jarociński and Karadi's (2020) poor man's sign restriction.

References

- Acosta, Miguel, Andrea Ajello, Michael Bauer, Francesca Loria, and Silvia Miranda-Agrippino (2025). *Financial Market Effects of FOMC Communication: Evidence from a New Event-Study Database*. Working Paper 2025-30. Federal Reserve Bank of San Francisco.
- Almerud, Jakob, Dominika Krygier, Henrik Lundvall, and Mambuna Njie (2024). *Measuring Riksbank monetary policy: Shocks and macroeconomic transmission*. Tech. rep. Sveriges Riksbank Working Paper Series.
- Amberg, Niklas, Thomas Jansson, Mathias Klein, and Anna Rogantini Picco (2022). “Five facts about the distributional income effects of monetary policy shocks”. In: *American Economic Review: Insights* 4.3, pp. 289–304.
- Andersen, Asger Lau, Niels Johannesen, Mia Jørgensen, and José-Luis Peydró (2023). “Monetary policy and inequality”. In: *The Journal of Finance* 78.5, pp. 2945–2989.
- Andersson, Fredrik NG and Lars Jonung (2018). *Lessons for Iceland from the monetary policy of Sweden*. Tech. rep. Working Paper.
- Andersson, Fredrik NG and Josefin Kilman (2021). *A Study of the Romer and Romer Monetary Policy Shocks Using Revised Data*. Tech. rep. Working Paper.
- Antolín-Díaz, Juan and Juan F Rubio-Ramírez (2018). “Narrative sign restrictions for SVARs”. In: *American Economic Review* 108.10, pp. 2802–2829.
- Aruoba, S Borağan and Thomas Drechsel (2026). “Identifying monetary policy shocks: A natural language approach”. In: *American Economic Journal: Macroeconomics (Forthcoming)*.
- Ash, Elliott and Stephen Hansen (2023). “Text algorithms in economics”. In: *Annual Review of Economics* 15, pp. 659–688.
- Bahdanau, Dzmitry, Kyunghyun Cho, and Yoshua Bengio (2014). “Neural machine translation by jointly learning to align and translate”. In: *arXiv preprint arXiv:1409.0473*.
- 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.
- Battaglia, Laura, Timothy Christensen, Stephen Hansen, and Szymon Sacher (2024). “Inference for regression with variables generated by ai or machine learning”. In: .
- Bauer, Michael D and Eric T Swanson (2023a). “A reassessment of monetary policy surprises and high-frequency identification”. In: *NBER Macroeconomics Annual* 37.1, pp. 87–155.
- (2023b). “An alternative explanation for the “fed information effect””. In: *American Economic Review* 113.3, pp. 664–700.
- Bengio, Yoshua, Réjean Ducharme, Pascal Vincent, and Christian Jauvin (2003). “A neural probabilistic language model”. In: *Journal of Machine Learning Research* 3.Feb, pp. 1137–1155.
- Bernanke, Ben S, Jean Boivin, and Piotr Elias (2005). “Measuring the effects of monetary policy: a factor-augmented vector autoregressive (FAVAR) approach”. In: *The Quarterly Journal of Economics* 120.1, pp. 387–422.
- Bertsch, Christoph, Isaiah Hull, Robin L Lumsdaine, and Xin Zhang (2025). “Central bank mandates and monetary policy stances: Through the lens of federal reserve speeches”. In: *Journal of Econometrics*, p. 105948.
- Blanchard, Olivier and Roberto Perotti (2002). “An empirical characterization of the dynamic effects of changes in government spending and taxes on output”. In: *the Quarterly Journal of Economics* 117.4, pp. 1329–1368.

- Blei, David M, Andrew Y Ng, and Michael I Jordan (2003). “Latent dirichlet allocation”. In: *Journal of Machine Learning Research* 3.Jan, pp. 993–1022.
- Boukus, Ellyn and Joshua V. Rosenberg (2006). *The Information Content of FOMC Minutes*. Staff Report Paper No. 246. New York, NY: Federal Reserve Bank of New York.
- Buda, Gergely, Vasco M Carvalho, Giancarlo Corsetti, João B Duarte, Stephen Hansen, Afonso S Moura, Álvaro Ortiz, Tomasa Rodrigo, JV Rodríguez Mora, and G Alves da Silva (2025). “The short lags of monetary policy”. In: *CAMBRIDGE WORKING PAPERS IN ECONOMICS*.
- Caldara, Dario and Edward Herbst (2019). “Monetary policy, real activity, and credit spreads: Evidence from Bayesian proxy SVARs”. In: *American Economic Journal: Macroeconomics* 11.1, pp. 157–192.
- Campello, Ricardo JGB, Davoud Moulavi, and Jörg Sander (2013). “Density-based clustering based on hierarchical density estimates”. In: *Pacific-Asia Conference on Knowledge Discovery and Data Mining*. Springer, pp. 160–172.
- Chernozhukov, Victor, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, Whitney Newey, and James Robins (2018). *Double/debiased machine learning for treatment and structural parameters*.
- Cieslak, Anna, Stephen Hansen, Michael McMahon, and Song Xiao (2023). *Policymakers’ uncertainty*. Tech. rep. National Bureau of Economic Research.
- Cieslak, Anna and Annette Vissing-Jorgensen (2021). “The economics of the Fed put”. In: *The Review of Financial Studies* 34.9, pp. 4045–4089.
- Cloyne, James and Patrick Hürtgen (2016). “The macroeconomic effects of monetary policy: a new measure for the United Kingdom”. In: *American Economic Journal: Macroeconomics* 8.4, pp. 75–102.
- Cochrane, John H (2004). “Comments on Christina and David Romer “A new measure of monetary policy””. In: *July 2004 EFG meeting*.
- Coibion, Olivier (2012). “Are the effects of monetary policy shocks big or small?” In: *American Economic Journal: Macroeconomics* 4.2, pp. 1–32.
- Davison, Anthony Christopher and David Victor Hinkley (1997). *Bootstrap methods and their application*. 1. Cambridge university press.
- Delis, Manthos D, Annalisa Ferrando, Klaas Mulier, and Steven Ongena (2025). “The Poor, the Rich, and the Credit Channel of Monetary Policy”. In: *ECB Working Paper Series, No3058*.
- Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova (2019). “Bert: Pre-training of deep bidirectional transformers for language understanding”. In: *Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers)*, pp. 4171–4186.
- Dumais, Susan T (2004). “Latent semantic analysis.” In: *Annual Review of Information Science and Technology (ARIST)* 38, pp. 189–230.
- Frankel, Richard, Jared Jennings, and Joshua Lee (2022). “Disclosure sentiment: Machine learning vs. dictionary methods”. In: *Management Science* 68.7, pp. 5514–5532.
- Fusari, Francesco (2025). “Identifying monetary policy shocks through external constraints”. In: *Journal of Macroeconomics*, p. 103696.
- Gardner, Ben, Chiara Scotti, and Clara Vega (2022). “Words speak as loudly as actions: Central bank communication and the response of equity prices to macroeconomic announcements”. In: *Journal of Econometrics* 231.2, pp. 387–409.

- Gertler, Mark and Peter Karadi (2015). “Monetary policy surprises, credit costs, and economic activity”. In: *American Economic Journal: Macroeconomics* 7.1, pp. 44–76.
- Giannone, Domenico, Michele Lenza, and Giorgio E Primiceri (2021). “Economic predictions with big data: The illusion of sparsity”. In: *Econometrica* 89.5, pp. 2409–2437.
- Gorodnichenko, Yuriy, Tho Pham, and Oleksandr Talavera (2023). “The voice of monetary policy”. In: *American Economic Review* 113.2, pp. 548–584.
- Grootendorst, Maarten (2022). “BERTopic: Neural topic modeling with a class-based TF-IDF procedure”. In: *arXiv preprint arXiv:2203.05794*.
- Hansen, Stephen and Michael McMahon (2016). “Shocking language: Understanding the macroeconomic effects of central bank communication”. In: *Journal of International Economics* 99, S114–S133.
- Hansen, Stephen, Michael McMahon, and Andrea Prat (2017). “Transparency and Deliberation Within the FOMC: A Computational Linguistics Approach”. In: *The Quarterly Journal of Economics* 133.2, pp. 801–870.
- Hanson, Samuel G and Jeremy C Stein (2015). “Monetary policy and long-term real rates”. In: *Journal of Financial Economics* 115.3, pp. 429–448.
- Harrell, Frank E. Jr. (2025). *Regression Modeling Strategies*. Accessed: 2025-10-21. Free online publication, Department of Biostatistics, Vanderbilt University.
- Hastie, Trevor, Robert Tibshirani, Jerome Friedman, et al. (2009). *The Elements of Statistical Learning*.
- Hinton, Geoffrey, Oriol Vinyals, and Jeff Dean (2015). “Distilling the knowledge in a neural network”. In: *arXiv preprint arXiv:1503.02531*.
- Holm, Martin Blomhoff, Pascal Paul, and Andreas Tischbirek (2021). “The transmission of monetary policy under the microscope”. In: *Journal of Political Economy* 129.10, pp. 2861–2904.
- Huang, Allen H, Hui Wang, and Yi Yang (2023). “FinBERT: A large language model for extracting information from financial text”. In: *Contemporary Accounting Research* 40.2, pp. 806–841.
- Jacobson, Margaret M., Christian Matthes, and Todd B. Walker (2023). *Temporal Aggregation Bias and Monetary Policy Transmission*. Tech. rep. 2022-054r1. Washington: Board of Governors of the Federal Reserve System.
- 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.
- Jordà, Òscar (2005). “Estimation and inference of impulse responses by local projections”. In: *American Economic Review* 95.1, pp. 161–182.
- Kekre, Rohan and Moritz Lenel (2025). “The High-Frequency Effects of Dollar Swap Lines”. In: *American Economic Review: Insights* 7.1, pp. 107–123.
- Kilman, Josefin (2022). *Monetary policy shocks for Sweden*. Tech. rep. Working paper.
- Kriegel, Hans-Peter, Peer Kröger, and Arthur Zimek (2009). “Clustering high-dimensional data: A survey on subspace clustering, pattern-based clustering, and correlation clustering”. In: *Acm Transactions on Knowledge Discovery from Data (tkdd)* 3.1, pp. 1–58.
- Künsch, Hans R (1989). “The jackknife and the bootstrap for general stationary observations”. In: *The annals of Statistics*, pp. 1217–1241.
- 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.
- Lahiri, Soumendra Nath (2013). *Resampling methods for dependent data*. Springer Science & Business Media.

- Laséen, Stefan (2020). *Monetary policy surprises, central bank information shocks, and economic activity in a small open economy*. Tech. rep. Sveriges Riksbank Working Paper Series.
- Leeper, Eric M, Christopher A Sims, Tao Zha, Robert E Hall, and Ben S Bernanke (1996). “What does monetary policy do?” In: *Brookings Papers on Economic Activity* 1996.2, pp. 1–78.
- Leippold, Markus (2023). “Sentiment spin: Attacking financial sentiment with GPT-3”. In: *Finance Research Letters* 55, p. 103957.
- Liu, Yinhan, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov (2019). “Roberta: A robustly optimized bert pretraining approach”. In: *arXiv preprint arXiv:1907.11692*.
- 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.
- McInnes, Leland, John Healy, and James Melville (2018). “Umap: Uniform manifold approximation and projection for dimension reduction”. In: *arXiv preprint arXiv:1802.03426*.
- Mertens, Karel and Morten O Ravn (2014). “A reconciliation of SVAR and narrative estimates of tax multipliers”. In: *Journal of Monetary Economics* 68, S1–S19.
- Meyersson, Pernilla and Ann-Leena Mikiver (2018). “The Riksbank’s communication before, during and after the financial crisis”. In: *Sveriges Riksbank Economic Review* 4, pp. 35–47.
- Miranda-Agrippino, Silvia and Giovanni Ricco (2021). “The transmission of monetary policy shocks”. In: *American Economic Journal: Macroeconomics* 13.3, pp. 74–107.
- 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.
- Olea, José Luis Montiel, Mikkel Plagborg-Møller, Eric Qian, and Christian K Wolf (2025). “Local Projections or VARs? A Primer for Macroeconomists”. In: *Prepared for the NBER Macroeconomics Annual 2025*.
- Pfeifer, Moritz and Vincent P Marohl (2023). “CentralBankRoBERTa: A fine-tuned large language model for central bank communications”. In: *The Journal of Finance and Data Science* 9, p. 100114.
- Ramey, Valerie A (2016). “Macroeconomic shocks and their propagation”. In: *Handbook of Macroeconomics* 2, pp. 71–162.
- Riksbanken (2018). *Evaluation of the Riksbank’s Forecasts*. Tech. rep. Riksbank Studies.
- Romer, Christina D and David H Romer (1989). “Does monetary policy matter? A new test in the spirit of Friedman and Schwartz”. In: *NBER Macroeconomics Annual* 4, pp. 121–170.
- (2004). “A new measure of monetary shocks: Derivation and implications”. In: *American economic review* 94.4, pp. 1055–1084.
- (2023). “Presidential address: Does monetary policy matter? The narrative approach after 35 years”. In: *American Economic Review* 113.6, pp. 1395–1423.
- Sandström, Maria (2018). “The impact of monetary policy on household borrowing-a high-frequency IV identification”. In: *Riksbank Research Paper Series* 174.
- Shapiro, Adam Hale and Daniel J Wilson (2022). “Taking the fed at its word: A new approach to estimating central bank objectives using text analysis”. In: *The Review of Economic Studies* 89.5, pp. 2768–2805.
- Silva, Thiago Christiano, Kei Moriya, and Romain Michel Veyrune (2025). *From Text to Quantified Insights: A Large-Scale LLM Analysis of Central Bank Communication*. Tech. rep. International Monetary Fund.

- 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.
- Svensson, Lars E.O. (2010). “Transparency, Credibility, and Accountability in Monetary Policy”. In: *Sveriges Riksbank Economic Review* 1. Originally presented at the conference *Ten Years with an Independent Riksbank*, Stockholm, 11 September 2009, pp. 5–44. URL: <https://larseosvensson.se/files/papers/PolExpEval.pdf> (visited on 05/11/2026).
- Sveriges Riksbank (2017). *The CPIF as Target Variable for Monetary Policy*. Published 7 September 2017. Accessed: 2026-05-11. Available at: <https://www.riksbank.se/en-gb/about-the-riksbank/history/historical-timeline/2000-2018/the-cpif-as-target-variable-for-monetary-policy/>.
- (2023). *The Riksbank’s Communication Policy*. Communication Policy adopted by the Executive Board on 14 December 2023. Accessed: 2026-06-09. Available at: <https://www.riksbank.se/en-gb/about-the-riksbank/how-the-riksbank-is-governed/internal-regulatory-framework/policies/communication-policy/>.
- (2025). *How a Monetary Policy Decision Is Implemented*. Last updated 3 April 2025. Accessed: 2026-05-11. Available at: <https://www.riksbank.se/en-gb/monetary-policy/how-a-monetary-policy-decision-is-implemented/>.
- (2026a). *Complementary Monetary Policy Measures*. Last updated 20 January 2026. Accessed: 2026-05-11. Available at: <https://www.riksbank.se/en-gb/monetary-policy/what-is-monetary-policy/complementary-monetary-policy-measures/>.
- (2026b). *Monetary Policy Reports and Updates*. Accessed: 2026-02-05, Available at: <https://www.riksbank.se/en-gb/monetary-policy/monetary-policy-report/monetary-policy-reports-and-updates/>.
- (2026c). *The Inflation Target*. Last updated January 2026. Accessed: 2026-05-11. Available at: <https://www.riksbank.se/en-gb/monetary-policy/the-inflation-target/>.
- (2026d). *The Path to a Monetary Policy Decision*. Last updated January 2026. Accessed: 2026-05-11. Available at: <https://www.riksbank.se/en-gb/monetary-policy/the-path-to-a-monetary-policy-decision/>.
- Swanson, Eric T (2021). “Measuring the effects of federal reserve forward guidance and asset purchases on financial markets”. In: *Journal of Monetary Economics* 118, pp. 32–53.
- Uhlig, Harald (2005). “What are the effects of monetary policy on output? Results from an agnostic identification procedure”. In: *Journal of Monetary Economics* 52.2, pp. 381–419.
- Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin (2017). “Attention is all you need”. In: *Advances in Neural Information Processing Systems* 30.
- Wang, Wenhui, Hangbo Bao, Shaohan Huang, Li Dong, and Furu Wei (2020). “Minilmv2: Multi-head self-attention relation distillation for compressing pretrained transformers”. In: *arXiv preprint arXiv:2012.15828*.
- Wieland, Johannes F and Mu-Jeung Yang (2020). “Financial dampening”. In: *Journal of Money, Credit and Banking* 52.1, pp. 79–113.
- Xiao, Tong and Jingbo Zhu (2023). “Introduction to transformers: an nlp perspective”. In: *arXiv preprint arXiv:2311.17633*.
- Zhang, Yue and Zhiyang Teng (2021). *Natural Language Processing: a Machine Learning Perspective*. Cambridge University Press.

Appendix

A Monetary Policy Shocks - Additional Results

A.1 Data and Descriptive Statistics

Table 6: Description of Data

Panel A: Shock Data		
Variable	Source	Definition
GDP growth forecast	Kilman (2022)	Year-over-year growth in GDP at market prices
CPI inflation forecast	Kilman (2022)	Year-over-year growth in the all-items CPI
Policy rate	Kilman (2022)	Riksbank repo rate
Unemployment forecast	Kilman (2022)	Percent of the labor force
T-Bills (One-Month)	Amberg et al. (2022)	Daily change in one-month Treasury bill yields
OMXSPI	Amberg et al. (2022)	Daily change in the OMXSPI
Panel B: Macroeconomic Data		
Variable	Source	Definition
CPI	Kilman (2022)	All-items consumer price index
Real GDP	Kilman (2022)	Million SEK, deflated using the GDP deflator
Unemployment	Kilman (2022)	Percent of the labor force aged 16–64
Panel C: Text Data		
Variable	Source	Definition
Inflation Reports	Sveriges Riksbank	Monetary policy communication
Monetary Policy Reports	Sveriges Riksbank	Monetary policy communication
Monetary Policy Updates	Sveriges Riksbank	Monetary policy communication

Notes: The data from Kilman (2022) were provided by the author. The data from Amberg et al. (2022) were obtained from the replication package available at <https://www.aeaweb.org/articles?id=10.1257/aeri.20210262> (accessed: 2025-10-23). Riksbank documents from 1996–2016 were downloaded from <https://archive.riksbank.se/en/Web-archive/Published/Published-from-the-Riksbank/Monetary-policy/Monetary-Policy-Report/index.html?all=1.html>. The remaining reports were downloaded from <https://www.riksbank.se/en-gb/monetary-policy/monetary-policy-report/monetary-policy-reports-and-updates/> (accessed: 2025-01-22).

Table 7: Descriptive Statistics of Monetary Policy Shocks

Panel A: Romer & Romer Style Shocks					
Shock	Obs.	Mean	Std. Dev.	Min	Max
R&R	279	0.000	0.169	-1.273	0.548
R&R-T1	279	0.000	0.105	-0.772	0.339
R&R-T2	279	0.000	0.109	-0.763	0.370
R&R-T3	279	0.000	0.101	-0.645	0.347
R&R-T4	279	0.000	0.105	-0.724	0.356
R&R-AT	279	0.000	0.136	-1.013	0.456

Panel B: High-Frequency Style Shocks					
Shock	Obs.	Mean	Std. Dev.	Min	Max
HF	239	-0.014	0.081	-0.626	0.271
HF-T1	239	-0.014	0.083	-0.614	0.286
HF-T2	239	-0.014	0.084	-0.619	0.286
HF-T3	239	-0.014	0.083	-0.613	0.285
HF-T4	239	-0.014	0.083	-0.615	0.262
HF-AT	239	0.000	0.069	-0.497	0.285

Notes: R&R denotes standard Romer & Romer shocks and HF denotes High Frequency shocks, see Section 2.1 and 2.2 respectively. "T" denotes text-based shocks, identified in Section 3.1 and 3.2, the number following "T" indicates the specification of the Greenbook controls used to identify the shock. See Section 5 for details on alternative Greenbook specifications. "AT" denotes the alternative text-augmentation approach described in Section 5. The impulse responses reported in Section 4 are based on R&R, R&R-T1, HF, and HF-T1. The average value for the HF shocks differ on further decimals.

Table 8: Correlation Matrix of Monetary Policy Shocks

	R&R	R&R T1	R&R T2	R&R T3	R&R T4	R&R AT1	HF	HF T1	HF T2	HF T3	HF T4	HF AT1
R&R	1.000											
R&R_T1	0.780	1.000										
R&R_T2	0.815	0.968	1.000									
R&R_T3	0.786	0.973	0.974	1.000								
R&R_T4	0.804	0.976	0.984	0.981	1.000							
R&R_AT	0.968	0.87	0.899	0.877	0.891	1.000						
HF	0.417	0.416	0.414	0.415	0.409	0.418	1.000					
HF_T1	0.419	0.479	0.467	0.479	0.468	0.448	0.968	1.000				
HF_T2	0.424	0.479	0.468	0.479	0.469	0.451	0.97	0.997	1.000			
HF_T3	0.418	0.479	0.467	0.479	0.468	0.447	0.967	0.996	0.997	1.000		
HF_T4	0.419	0.478	0.467	0.478	0.467	0.448	0.968	0.999	0.997	0.994	1.000	
HF_AT	0.412	0.473	0.461	0.471	0.459	0.441	0.960	0.979	0.979	0.979	0.979	1.000

Notes: R&R denotes standard Romer & Romer shocks and HF denotes High Frequency shocks, see Section 2.1 and 2.2 respectively. "T" denotes text-based shocks, identified in Section 3.1 and 3.2, the number following "T" indicates the specification of the Greenbook controls used to identify the shock. See Section 5 for details on alternative Greenbook specifications. "AT" denotes the alternative text-augmentation approach described in Section 5. The impulse responses reported in Section 4 are based on R&R, R&R-T1, HF, and HF-T1.

A.2 Additional Estimation Results

Table 9: Estimation Results from Alternative Greenbook Specifications

Panel A: Romer & Romer Specifications						
Specification	Standard R&R		v^*	Text-Based R&R		
	#params	R^2_{adj}		# topics	#params	R^2_{adj}
X_1	14	0.329	59	234	482	0.718
X_2	27	0.386	75	440	213	0.704
X_3	15	0.555	98	318	152	0.756
X_4	15	0.337	90	354	170	0.727

Panel B: High-Frequency Specifications							
Specification	Standard HF		v^*	Text-HF Ridge		Text-HF OLS	
	#params	R^2_{adj}		#params	$R^2_{adj, Ridge}$	#params	$R^2_{adj, OLS}$
X_1	2	0.176	80	388	0.151	2	0.179
X_2	2	0.176	110	296	0.153	2	0.187
X_3	2	0.176	104	318	0.164	2	0.185
X_4	2	0.176	96	344	0.159	2	0.186

Notes: The text-based specifications use $W_i(v_i^*) = [X_i \ Z(v_i^*)]$ as regressors, where $Z(v_i^*)$ denotes the optimal topic-sentiment representation associated with Greenbook specification X_i , $i \in \{1, 2, 3, 4\}$. The R^2_{adj} in Panel A capture the systematic component of monetary policy while $R^2_{adj, OLS}$ in panel B directly capture the non-systematic component, the monetary policy shock. $R^2_{adj, Ridge}$ refers to the ridge regression of $W_i(v_i^*)$ on Δi^{PMSE} . #params describes the number of parameters used in each regression and v^* the optimal threshold parameter obtained via grid search over $\mathcal{V}$. The specification based on X_3 is estimated using data from 2007–2019 for the standard R&R and HF method. The corresponding text-based specification is estimated on the full sample because Riksbank documents are available throughout the period.

Table 10: Robustness Test: Projecting Text-Based Information from R&R and HF Shocks

Specification	Text-Based R&R				Text-Based HF			
	v^*	# topics	#params	R^2_{adj}	v^*	# topics	#params	R^2_{adj}
X_1	63	262	538	0.202	198	64	142	0.158
X_2	115	134	282	0.208	198	64	142	0.149
X_3	63	262	538	0.180	198	64	142	0.155
X_4	63	262	538	0.201	63	64	142	0.155

Notes: The results presented here are based on regressions where standard R&R and HF shocks are used as dependent variables, regressed on $W_i(v_i^*) = [X_i \ Z(v_i^*)]$, where $Z(v_i^*)$ denotes the optimal topic-sentiment representation associated with benchmark specification X_i . R^2_{adj} describes the variation in the standard shocks that is explained by variation in Greenbook and text-based regressors. The specification based on X_3 is estimated using data from 2007–2019 for the standard R&R and HF method. The corresponding text-based specification is estimated on the full sample because Riksbank documents are available throughout the period.

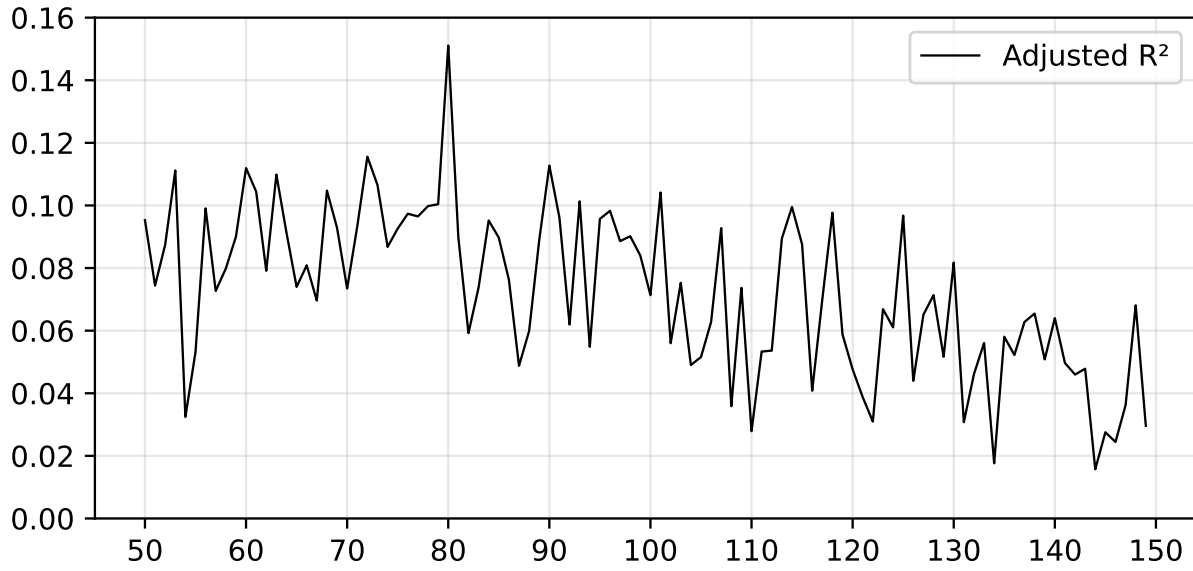


Figure 6: $R_{adj}^2(v)$ for $v \in \mathcal{V}$ for HF Identification

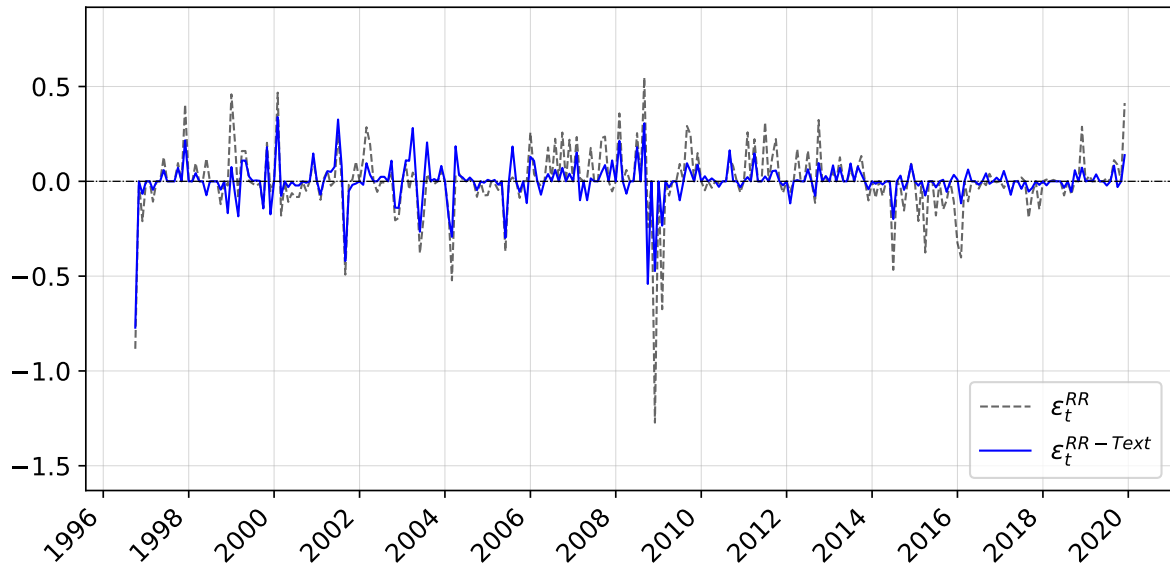
Table 11: Text-Based HF Shock Estimation

Panel A. Text Representation			
Embedding dimension (v^*)	80		
Number of "linear" text features	187		
Number of squared text features	187		
Number of R&R parameters	14		
Total regressors	388		
Panel B. Estimation Results			
	Text-based Ridge	Text-based OLS	Standard OLS
Observations	139	139	139
Number of regressors	388	2	2
Effective degrees of freedom	42.7	2	2
Optimal ridge penalty (λ^*)	124.22	—	—
Adjusted R^2	0.151	0.179	0.176
Training MSE	0.00180	0.0518	0.0519
LOOCV MSE	0.00238	0.0539	0.0542

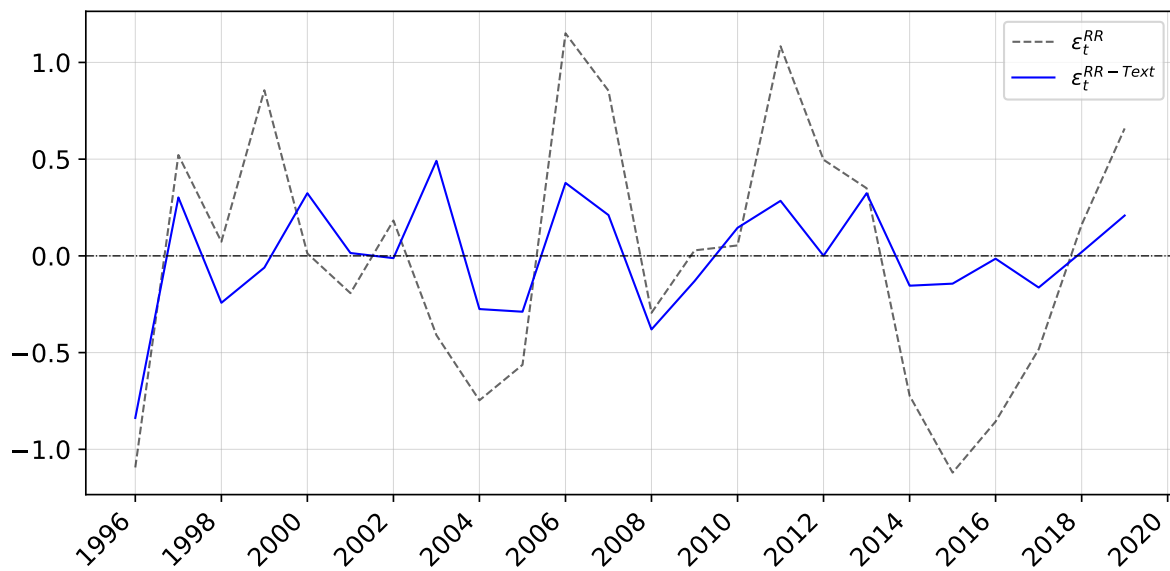
Notes: Standard OLS refers to the standard HF specification in Section 2.2 where policy rate changes are regressed on monetary policy surprises filtered with PMSR. Text-based Ridge refers to the regression in 3.2 where this measure of policy surprises are regressed on the optimal set of topic-sentiment series and standard Greenbook forecasts, $W(v^*) = [X \ Z(v^*)]$. The ridge penalty parameter λ^* is selected by minimizing the leave-one-out cross-validation mean squared error (LOOCV MSE). Adjusted R^2 are for the ridge regression computed based on the effective degrees of freedom, computed as $\text{tr}(X(X'X + \lambda I)^{-1}X')$. Text-based OLS refer to a regression where the residual from the text-based ridge regression, the improved instrument for monetary policy surprises, is regressed on policy rate changes. In this second step, adjusted R^2 represents the non-systematic component of monetary policy and is computed based on the model degrees of freedom.

A.3 Shocks at Monthly and Yearly Frequency

Figure 7: Standard and Text-Augmented R&R Shocks

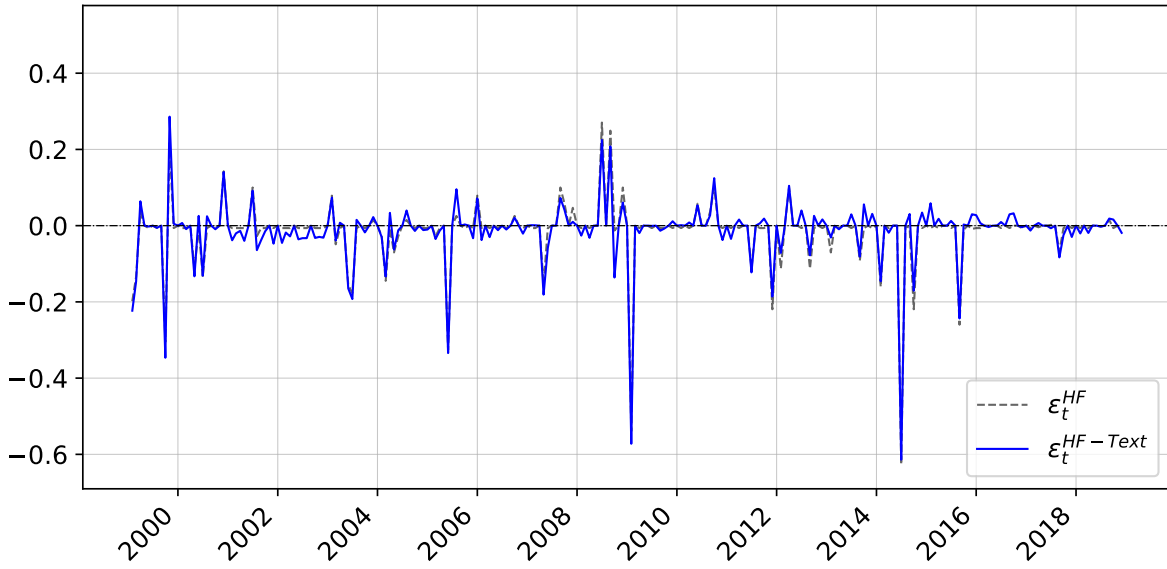


(a) Monthly Frequency

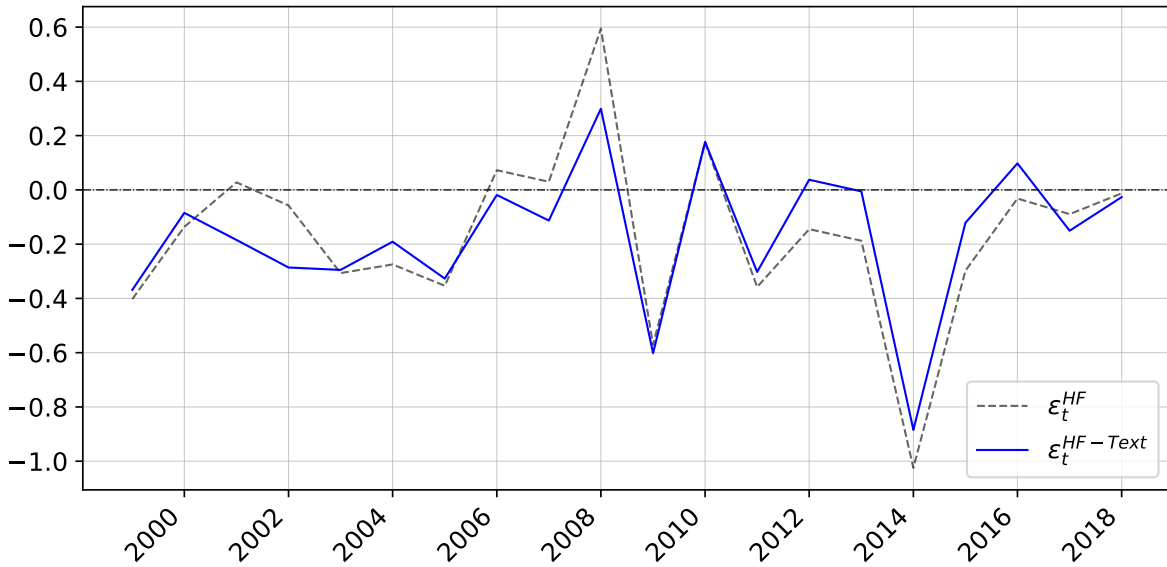


(b) Yearly Frequency

Figure 8: Standard and Text-Augmented HF Shocks



(a) Monthly Frequency



(b) Yearly Frequency

A.4 Annual Responses

Impulse responses to shocks aggregated to annual frequency are presented in Figure 9. Although it remains standard practice to aggregate the monetary policy shocks estimated at meeting frequency by summing up all shocks within a year, this practice may introduce a time aggregation bias in responses (Jacobson et al., 2023). That is, if the monetary variables are expected to respond

to the shock within the aggregation window, aggregating the data then effectively remove such dynamics by imposing a uniform timing assumption. While shocks in early quarters has had more time to affect outcome variables than shocks in late quarters, aggregation gives all shocks equal weight within the period (year). Indeed, a year with a 50 basis points contractionary and two 25 basis points expansionary policy shocks would be summarized as a year with no policy shocks¹⁷. While not addressing this issue directly, the bootstrap confidence intervals reflect some uncertainty related to this practice, see Appendix E for more details.

That said, much of the results found at the quarterly frequency is also present in the annual data, see Section 4.1. The dynamics of the responses to R&R-type shocks are all in line with theoretical predictions as productivity falls following a contractionary shock which increases unemployment and puts downwards pressure on prices. In comparison to responses from $\hat{\varepsilon}_t^{RR}$, responses to $\hat{\varepsilon}_t^{RR-Text}$ are large in magnitude and are more uncertain. While $\hat{\varepsilon}_t^{RR}$ and $\hat{\varepsilon}_t^{RR-Text}$ capture much of the same variation ($corr(\hat{\varepsilon}_t^{RR}, \hat{\varepsilon}_t^{RR-Text}) = 0.78$), the shock series themselves are different, see Figure 3. Recall that $\hat{\varepsilon}_t^{RR}$ represents 67.1% of the variation in Δs_t while $\hat{\varepsilon}_t^{RR-Text}$ represent only 28.2%. These differences manifest, with varying magnitude depending on bootstrap sample, in each bootstrap replication. This induce large uncertainty around the estimated impulse responses.

Apart from the price puzzle also documented in the quarterly data, the responses to HF shocks align with expectations. For the HF-type shocks, recall that while text-based regressors were important to identify an instrument for monetary policy surprises orthogonal to available information, the shocks identified with this improved instrument are almost identical to standard HF shocks, see Figure 4. The impact of the uncertainty in $\hat{\varepsilon}_t^{HF-Text}$ that stem from $\hat{\beta}_{v*}^{ridge}$ is therefore not large, which is reflected in the substantial overlap between standard and bootstrap confidence intervals.

I observe a mean reverting behavior of the policy rate to a contractionary shock in $\hat{\varepsilon}_t^{RR}$ and $\hat{\varepsilon}_t^{RR-Text}$, with a response larger in magnitude for the text-based shock. The price puzzle identified in the quarterly data, see Figure 5, is not present in the annual data for either version of the R&R shock. Instead, a contractionary shock significantly reduces the price level three years after impact, in line with expectations. The magnitude is again larger in the response to the text-based shock.

Output decreases two years after a shock in $\hat{\varepsilon}_t^{RR-Text}$, however not significantly. For the remain-

¹⁷While this is also a critique towards quarterly aggregation, all policy relevant variables are not available at frequencies higher than quarterly, why analysis on quarterly data serves a reasonable compromise.

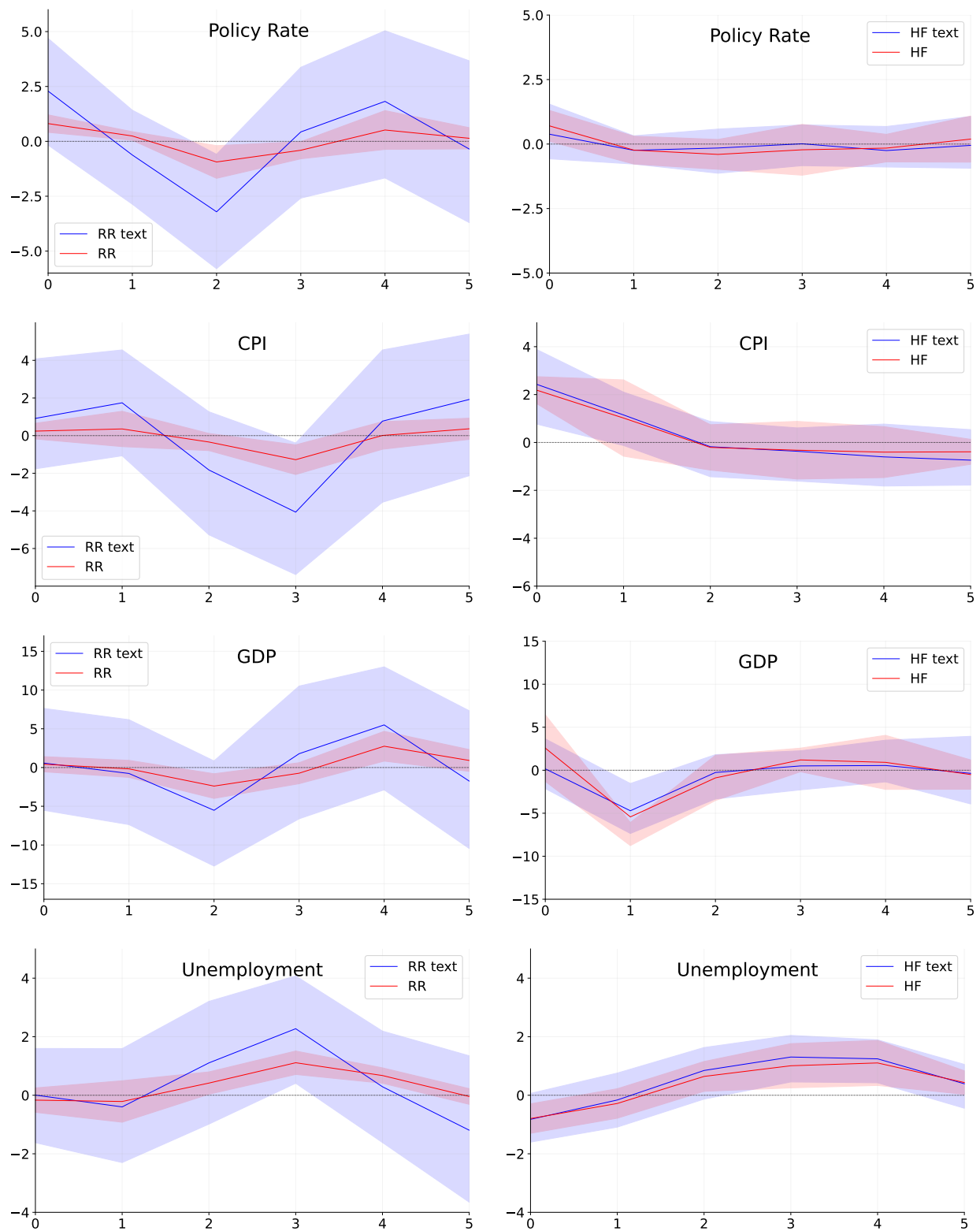


Figure 9: Impulse Response Functions Estimated at Annual Frequency

Note: Red shaded areas denote standard 95% confidence intervals, blue shaded areas denote 95% bootstrap confidence intervals. All specifications include two lags of the dependent variable.

-ing periods, the effect is zero. This is in contrast to the response of output to $\hat{\varepsilon}_t^{RR}$ which indicates a significant increase in the fourth period, which is a puzzle. Similar to the responses in price level, the puzzling on-impact fall of unemployment reported for the quarterly frequency is not present in the annual data. Rather, unemployment increase three years after both the standard and text-based R&R shock. Compared to $\hat{\varepsilon}_t^{RR-Text}$, responses to $\hat{\varepsilon}_t^{RR}$ is smaller but more persistent.

Different from the responses to R&R shocks, there is no mean reverting behavior in the policy rate response to either $\hat{\varepsilon}_t^{HF}$ or $\hat{\varepsilon}_t^{HF-Text}$. The price puzzle observed in the quarterly data is also present in the annual data, in responses to both HF-type shocks. Neither do the impulse responses indicate any tendency of a reduction in the price level over the estimation horizon. The response in output is of similar magnitude and duration for $\hat{\varepsilon}_t^{HF}$ and $\hat{\varepsilon}_t^{HF-Text}$ as for $\hat{\varepsilon}_t^{RR-Text}$. The HF shocks, however, transmit faster, one year after impact, and document significant effects. The response to unemployment from the standard HF shock present a puzzle as the initial response to a contractionary shock is a significantly lower unemployment. This reverts back to the expected increase in unemployment after two years and stays significant until after four years. The impact of unemployment to $\hat{\varepsilon}_t^{HF-Text}$ is free of this puzzle.

As for the quarterly data, the results in the annual data related to $\hat{\varepsilon}_t^{RR-Text}$ most resembles 'textbook' responses to a policy tightening. An unexpected policy increase decreases economic activity two years after the shock, although not significant at the 95% level. This puts downwards pressure on prices and drives up the unemployment rate, both which moves three years after the shock.

B Details on Highest-loading Topics

B.1 Text-based R&R

Table 12: Highest-loading Topics for Text-based R&R Regression

Rank	Topic Number	Topic Interpretation	Top Keywords	Representative Sentences
1	80	Unemployment reporting	unemployment, percent, end, just, amounted, fell, fall, period, employment, rose	"Unemployment was 6.2 percent." "Unemployment was 6.5 percent."
2	105	Emerging economies and Asian growth	emerging, economies, asia, growth, asian, america, slow-down, china, russia, grow	"Growth is still high in emerging economies, especially in Asia and Eastern Europe." "The emerging economies of Asia are continuing to grow rapidly."
3	3	Repo rate discussions	repo, rate, path, raised, cut, scenario, unchanged, forward, rates, implied	"At the end of the forecast period, the repo rate is expected to be over 2 per cent." "The repo rate is therefore being cut to 0.25 per cent and the repo-rate path is being revised down."
4	117	Revised GDP forecast	revised, gdp, forecast, upwards, growth, compared, downwards, quarter, update, forecasts	"The forecast for GDP growth has been revised up for all years in the forecast period." "At the same time, the forecast for GDP growth has been revised down."
5	20	Fiscal budget balance	fiscal, budget, balance, surplus, ceiling, public, government, structural, expenditure, tightening	"Fiscal policy is currently covered by three important restrictions; the expenditure ceiling, the surplus target and the local government balance requirement." "The direction of fiscal policy is usually measured in terms of the structural budget balance (the cyclically adjusted balance)."
6	48 (SQ)	Information sources	sources, reuters, ecowin, imf, eurostat, source, riksbank, bureau, national, figure	"Sources: Reuters EcoWin and the Riksbank Figure 3:4." "Sources: Reuters EcoWin and the Riksbank Figure 3:4."
7	14	Swedish economic growth	swedish, economy, sweden, growth, gdp, quarter, strong, activity, economic, grow	"Growth in the Swedish economy has been relatively high since 2014." "The Swedish economy is strong."
8	1	Inflation (CPI and CPIF)	inflation, cpif, cpi, energy, cent, expected, expectations, underlying, low, lower	"Excluding energy prices, CPIF inflation was lower and amounted to 1.7 per cent, which was also in line with the Riksbank's forecast." "CPIF inflation was 1.8 per cent and CPIF excluding energy was 1.5 per cent."
9	45 (SQ)	Energy and Electricity Prices	tcw, sek, exchange, scale, eur, real, left, oct, index, right	SEK/TCW exchange rate. , SEK/TCW exchange rate. , SEK/TCW exchange rate.
10	19	Resource utilisation	resource, utilisation, normal, resources, unutilised, period, economy, forecast, higher, capacity	This is expected to lead to a lower resource utilisation." "Resource utilisation will thus increase."

Notes: Rank reflects the magnitude of the coefficient for the corresponding topic-sentiment time series and topic number refers to BERTopic's internal numbering of topics. "(SQ)" means that the squared version of corresponding topic-sentiment time series entered the ridge regression. Top keywords are computed with c-TF-IDF and representative sentences are sentences closest to the cluster centroid of the relevant topic. The topic interpretation is the authors description and interpretation of the topic.

Table 13: Highest-loading Topics for Text-based R& R Regression, continued

Rank	Topic Number	Topic Interpretation	Top Keywords	Representative Sentences
11	39	Companies, prices and costs	companies, price, prices, costs, increases, raise, demand, consumers, pass, competition	"Then companies' costs normally rise and they also find it easier to pass on their cost increases to consumer prices." , "The companies' costs then increase too."
12	50 (SQ)	Industrial capacity	capacity, utilisation, industry, manufacturing, nier, institute, plant, bottleneck, primary, firms	"Current capacity utilisation in the manufacturing industry in percent." , "Moreover, the proportion of firms reporting machinery and plant capacity as a bottleneck has decreased recently , which presumably implies less need to expand production facilities."
13	126 (SQ)	Households and company confidence	optimistic, households, pessimistic, optimism, views, confidence, survey, finances, tendency, future	"However, according to the Consumer Tendency Survey, households have become more optimistic about the future." , "Moreover, households and companies are still relatively optimistic."
14	34 (SQ)	Labour market shortages	shortage, shortages, reporting, tendency, labour, recruitment, companies, business, construction, survey	"According to the Economic Tendency Survey, a large proportion of companies report that they have a shortage of labour (see Figure 3:11)." , "Shortage figures in the business sector indicate labour shortages in the construction industry in particular, but also within the private service sector. "
15	8 (SQ)	Housing market	housing, construction, house, investment, prices, apartments, owned, tenant, market, houses	"This can lead to a fall in demand for housing and thus a fall in investment in housing." , "Historically, housing prices have covaried with housing investment and housing investment, as a share of GDP, has also increased significantly in recent years (see Figure 4:18 and Figure 4:20)."
16	52	Data sources	eurostat, research, institute, sources, national, bureau, source, economic, analysis, commission	"Sources: Bureau of Economic Analysis, Eurostat, national sources and the Riksbank Figure 3:6", "Eurostat, national sources"
17	116 (SQ)	Gross capital formation	gross, formation, capital, fixed, xed, housing, business, authorities, total, fixed	"Deviation from baseline: Household consumption -1.0, Public consumption -1.5, Fixed gross capital formation -0.5, Exports -1.2' , 'The underlying path for the growth rate in gross capital formation has not changed, but has been displaced slightly into the future."
18	63 (SQ)	ECB monetary policy	ecb, bank, policy, central, european, expansionary, reserve, cut, unchanged, federal	"In the euro area, the European Central Bank (ECB) left its monetary policy unchanged at the meeting in July." , "The ECB's policy rate is now 1.0 per cent."
19	28	Swedish export market	exports, swedish, export, demand, growth, goods, markets, weak, imports, abroad	"Average growth on the swedish export market will be 8 per cent per year during the forecast period, and swedish exports are expected to increase in line with this figure." , "The Riksbank is making a downward revision to its assessment of Swedish export market growth, particularly during the current year. "
20	44	Energy prices and inflation	energy, electricity, prices, inflation, rising, expected, price, rise, partly	"The fact that electricity prices are now expected to remain relatively high during a long period of time means that the risk of contagion to other prices and costs has increased somewhat." , "Energy prices have been unexpectedly high, however, while other inflation has been somewhat lower than anticipated."

Notes: Rank reflects the magnitude of the coefficient for the corresponding topic-sentiment time series and topic number refers to BERTopic's internal numbering of topics. "(SQ)" means that the squared version of corresponding topic-sentiment time series entered the ridge regression. Top keywords are computed with c-TF-IDF and representative sentences are sentences closest to the cluster centroid of the relevant topic. The topic interpretation is the authors description and interpretation of the topic.

B.2 Text-based HF

Table 14: Highest-loading Topics for Text-based HF Regression

Rank	Topic Number	Topic Interpretation	Top Keywords	Representative Sentences
1	86	Resource utilisation	utilisation, resource, normal, uses, overall, indicator, indicators, riksbank, assessment, statistical	"The Riksbank's overall assessment is that resource utilisation in the economy is higher than normal." , "The Riksbank's overall assessment is that resource utilisation is still lower than normal."
2	42	Consumer and company confidence	confidence, consumer, indicators, sentiment, households, companies, sector, household, manufacturing, historical	"Confidence among both households and companies is at a very low level.", "Confidence among households and companies is very high."
3	107	Taxation policies	tax, cuts, local, taxes, income, government, transfers, revenue, expenditure, revenues	"The proposed measures include tax increases and spending cuts as well as the postponement of tax cuts that were planned earlier." , "Moreover, the increase in income this year will be particularly large as a result of tax cuts."
4	43 (SQ)	Electricity prices	electricity, energy, prices, inflation, rising, expected, price, rise, fall, falling	"This is because both electricity and oil prices are now lower." , "Higher oil prices but lower electricity prices."
5	118	GDP forecast revisions	revised, gdp, forecast, upwards, growth, compared, downwards, quarter, update, forecasts	"The forecast for GDP growth has been revised up for all years in the forecast period." , "At the same time, the forecast for GDP growth has been revised down."
6	55 (SQ)	Consumption goods inflation	prices, swedish, goods, price, import, sweden, food, imported, consumer, producer	"Producer prices for consumer goods are still rising faster than normal, both with regard to imported goods and goods produced in Sweden.", "The main explanation for the low domestic inflation stems from low consumer price increases on goods."
7	94 (SQ)	GDP growth forecasts	gdp, growth, expected, grow, relatively, years, high, year, historical, faster	"GDP growth is expected to be somewhat higher this year and next year." , "GDP growth is therefore expected to be somewhat lower in these years."
8	116	Geopolitical uncertainty	uncertainty, political, considerable, setbacks, developments, attacks, abroad, terrorist, economic, uncertain	"Uncertainty about political developments abroad is giving rise to major uncertainty regarding economic development and increasing the risk of setbacks.", "There is great uncertainty over both political and economic developments abroad."
9	15 (SQ)	Bond yields	yields, bond, government, bonds, yield, fallen, german, germany, rates, differential	"At the same time, yields on higher risk bonds, such as corporate bonds, have fallen in line with government bond yields.", "Swedish government bond yields have risen slightly in line with US yields."
10	118 (SQ)	GDP forecast revisions	revised, gdp, forecast, upwards, growth, compared, downwards, quarter, update, forecasts	"The forecast for GDP growth has been revised up for all years in the forecast period." , "At the same time, the forecast for GDP growth has been revised down."

Notes: Rank reflects the magnitude of the coefficient for the corresponding topic-sentiment time series and topic number refers to BERTopic's internal numbering of topics. "(SQ)" means that the squared version of corresponding topic-sentiment time series entered the ridge regression. Top keywords are computed with c-TF-IDF and representative sentences are sentences closest to the cluster centroid of the relevant topic. The topic interpretation is the authors description and interpretation of the topic.

Table 15: Highest-loading Topics for Text-based HF Regression, continued

Rank	Topic Number	Topic Interpretation	Top Keywords	Representative Sentences
11	68	European debt crisis	greece, debt, problems, euro, countries, sovereign, area, crisis, support, public	"Moreover, the problems with public debt in some European countries have continued.", "But in the euro area, the upturn in long -term rates causes problems for countries with sovereign debt problems."
12	132 (SQ)	Household and company confidence	optimistic, households, pessimistic, optimism, confidence, future, survey, tendency, finances, views	"However, according to the Consumer Tendency Survey, households have become more optimistic about the future." , "Moreover, households and companies are still relatively optimistic."
13	55	Consumption goods inflation	prices, swedish, goods, price, import, sweden, food, imported, consumer, producer	"Producer prices for consumer goods are still rising faster than normal, both with regard to imported goods and goods produced in Sweden.", "The main explanation for the low domestic inflation stems from low consumer price increases on goods."
14	43	Electricity prices	electricity, energy, prices, inflation, rising, expected, price, rise, fall, falling	"This is because both electricity and oil prices are now lower." , "Higher oil prices but lower electricity prices."
15	15	Bond yields	yields, bond, government, bonds, yield, fallen, german, germany, rates, differential	"At the same time, yields on higher risk bonds, such as corporate bonds, have fallen in line with government bond yields.", "Swedish government bond yields have risen slightly in line with US yields."
16	85	Expected GDP growth	gdp, cent, grow, year, expected, just, growth, ir, 2015, 2016	"This year, GDP is expected to grow by 3.5 per cent and thereafter by 2.8 and 2.4 per cent per year (see Table 5).", "All in all, GDP growth is expected to increase from just over 2 per cent this year to around 3.5 per cent next year."
17	99	Implied forward interest rates	forward, implied, funds, rates, fed, refi, curve, futures, premiums, securities	"A more detailed account of implied forward interest rates and how they can be used in the analysis has been presented in earlier inflation reports, e.g. 1996:4(December), p. 17.", "In that the monetary stance was adjusted for this information, the short implied forward interest rates fell continuously during the year."
18	116	Geopolitical uncertainty	uncertainty, political, considerable, setbacks, developments, attacks, abroad, terrorist, economic, uncertain	"Uncertainty about political developments abroad is giving rise to major uncertainty regarding economic development and increasing the risk of setbacks.", "There is great uncertainty over both political and economic developments abroad."
19	87 (SQ)	Money holdings	m3, money, m0, supply, m2, holdings, coins, banknotes, deposits, aggregate	"Money supply measured as M0 and M3.", "The liquidity aggregate comprises the non-bank public's holdings of M3, certificates, treasury bills, national savings accounts, National Debt Office Accounts, premium bonds and private bonds."
20	26	Riksbank forecasts	rikssbank, forecasts, forecast, assessment, model, growth, wage, forecasters, inflation, models	"Inflation was lower than the Riksbank expected in the forecasts made in 2004, although well in line with the forecasts made in 2005.", "This is lower than the Riksbank's forecast."

Notes: Rank reflects the magnitude of the coefficient for the corresponding topic-sentiment time series and topic number refers to BERTopic's internal numbering of topics. "(SQ)" means that the squared version of corresponding topic-sentiment time series entered the ridge regression. Top keywords are computed with c-TF-IDF and representative sentences are sentences closest to the cluster centroid of the relevant topic. The topic interpretation is the authors description and interpretation of the topic.

C Sub-sample Analysis

C.1 Estimation Results

Table 16: Model Fit Across Monetary Policy Regimes

Panel A: Romer & Romer Type Shocks					
Regime	Sample	Standard R&R		Text-based R&R	
		R^2_{adj}	#params	R^2_{adj}	#params
Pre 2008	1996-10 – 2007-12	0.352	14	0.607	100
Post 2008	2008-01 – 2019-12	0.324	14	0.663	248
Pre ZLB	1996-10 – 2014-10	0.375	14	0.665	160
Post ZLB	2014-11 – 2019-12	0.324	14	0.872	228
Full sample	1996-10 – 2019-12	0.329	14	0.718	482

Panel B: High-Frequency Type Shocks							
Regime	Sample	Standard HF		Text-HF Ridge		Text-HF OLS	
		$R^2_{adj,OLS}$	#params	$R^2_{adj,Ridge}$	#params	$R^2_{adj,OLS}$	#params
Pre 2008	1996-10 – 2007-12	0.333	2	0.049	290	0.332	2
Post 2008	2008-01 – 2019-12	0.112	2	0.305	286	0.122	2
Pre ZLB	1996-10 – 2014-10	0.181	2	0.108	206	0.194	2
Post ZLB	2014-11 – 2019-12	0.094	2	0.764	102	0.042	2
Full sample	1996-10 – 2019-12	0.175	2	0.151	388	0.179	2

Notes: Standard R&R denotes the Romer and Romer specification described in Section 2.1 and standard HF denotes High Frequency identification of Section 2.2. Estimation in sub-samples follow the procedure described in Section 2.1, 2.2, 3.1 and 3.2. The R^2_{adj} in Panel A capture the systematic component of monetary policy while $R^2_{adj,OLS}$ in panel B directly capture the non-systematic component, the monetary policy shock. $R^2_{adj,Ridge}$ refers to the ridge regression of $\mathbf{W}(v^*)$ on $\Delta \mathbf{i}^{PMR}$. #params describes the number of parameters used in each regression.

C.2 Highest-loading Topics in Sub-samples

Table 17: Highest-loading Topics for Text-based R&R Regression pre 2008

Rank	Topic Number	Topic Interpretation	Top Keywords	Representative Sentences
1	37	Stock indices	'stock', 'index', 'OMX', 'market', 'implied', 'risk', 'rose', 'percent', 'volatility', 'germany'	"The Stockholmsbörsen OMX index rose by around 17 percent, compared with around 9 per cent in the United States and 5 per cent in Germany." , "The recovery in the OMX index in recent years has coincided with a fall in the market, Ås risk assessment (implied volatility)."
2	0 (SQ)	Inflation assessments	inflation, assessment, effects, interest, rates, leads, adjustment, transitory, target, measures	"The assessment leads to the conclusion that, after adjustments for transitory effects from indirect taxes, subsidies and interest rates, the rate of inflation twelve to twenty-four months ahead will be somewhat below the Riksbank's target.", "On account of the time lag before monetary policy measures affect inflation, the construction of monetary policy is based on assessments of inflation in the coming two years."
3	26 (SQ)	Public inflation expectations	expectations', 'inflation', 'percent', 'ahead', 'households', 'year', 'survey', 'years', 'report', slightly'	"Surveys published since the June Report show a slight tendency for inflation expectations to become somewhat lower.", "Inflation expectations have moved up slightly since the October Report but continue to be in line with the inflation target in both the short and the long term."
4	39 (SQ)	Labour costs	unit', 'costs', 'labour', 'increase', 'productivity', 'year', 'forecast', 'lower', 'rise', 'wage	"Unit labour costs throughout the economy are expected to increase by 4.8 per cent this year.", "Growth in unit labour costs has been sluggish since 2002 (see Figure 37)."
5	13	Investments	investment', 'stock', 'corporate', 'markets', 'share', 'sector', 'market', 'financial', 'capital', 'earnings	"Thus, weaker demand for other corporate sector investment could have a considerable effect on total corporate sector investment.", "A high return on capital, high share prices and strong demand are expected to stimulate investment demand in the forecast period. "
6	29	Electricity prices	electricity', 'prices', 'energy', 'inflation', 'expected', 'higher', 'price', 'fall', 'contagion	"In the next few months, consumer electricity prices are expected to fall further." "Higher oil prices but lower electricity prices."
7	10	Wage negotiations	wage', 'increases', 'agreements', 'wages', 'sector', 'agreement', 'drift', 'round', 'local', 'negotiations	"The remaining portion, i.e. the difference between total wage increases and centrally agreed wage increases, consists of locally agreed wage increases – wage drift in the more traditional sense, i.e. wage increases outside the scope of the agreements – and structural effects caused by different composition in the populations measured.", "That may affect the wage negotiations during 2004 as well as the rate of wage increases in the following years."
8	4	Trade and price developments	prices', 'goods', 'price', 'exports', 'imported', 'imports', 'import', 'export', 'international', 'services	"More new players have entered the market and contributed to further price cuts.", "To some extent, however, this price effect is countered by a weaker development of international prices for manufactured exports in national currencies as well as of international agricultural and food prices. "
9	36 (SQ)	Labour shortages	shortages', 'shortage', 'labour', 'construction', 'computer', 'skilled', 'services', 'technicians', 'business', 'manufacturing	"Labour shortages may also arise in technological and scientific occupations.", "It is only in the construction sector that the labour shortage is starting to reach high levels. "
10	21	Output gap and its estimation	gap', 'output', 'hp', 'filter', 'method', 'estimates', 'prescott', 'uc', 'hodrick', 'pf	"The increased production opportunities therefore lead to an increase in both output and the output gap.", "The trends are calculated using a Hodric-Prescott filter"

Notes: Rank reflects the magnitude of the coefficient for the corresponding topic-sentiment time series and topic number refers to BERTopic's internal numbering of topics. "(SQ)" means that the squared version of corresponding topic-sentiment time series entered the ridge regression. Top keywords are computed with c-TF-IDF and representative sentences are sentences closest to the cluster centroid of the relevant topic. The topic interpretation is the authors description and interpretation of the topic.

Table 18: Highest-loading Topics for Text-based R&R Regression post 2008

Rank	Topic Number	Topic Interpretation	Top Keywords	Representative Sentences
1	25 (SQ)	Riksbank forecast model	forecasts', 'riksbank', 'forecast', 'model', 'principle', 'assessment', 'risks', '10', 'does', 'outcome	"In the Riksbank's forecasts, the risks of both stronger and weaker development are, in principle, balanced.", "The Riksbank's forecast is approximately in line with the model forecasts (see Figure 4:12)."
2		US housing market	'us', 'housing', 'banking', 'financial', 'stress', 'market', 'problems', 'pessimism', 'weakening', 'risk'	"A weakening of the US housing market would lead to renewed problems for the banking sector and to increased pessimism", "If the housing market in the United States continues to develop weakly, this could lead to problems for the US banking sector and increased pessimism among the households, which in turn would lead to the risk of unemployment becoming entrenched at a high level."
3	36	Brexit	eu', 'withdrawal', 'kingdom', 'uk', 'referendum', 'united', 'british', 'leave', 'agreement', 'uncertainty	"The United Kingdom's withdrawal from the EU", "Uncertainty over the United Kingdom's withdrawal from the EU has increased further since July."
4	15	Sovereign debt crisis	'sovereign', 'debt', 'bonds', 'bond', 'interest', 'high', 'rates', 'rate', 'greece', 'problems'	"These high bond rates are adding to the problems relating to sovereign debt in that they entail increased interest expenditure.", "This may be an indication that the market is pricing a high probability of some form of write-off of the Greek sovereign debt."
5	23	Exchange rates	exchange', 'real', 'rate', 'weaker', 'stronger', 'equilibrium', 'long', 'impact', 'relative', 'effects	"At present, the real exchange rate is weaker than this level, which means that the exchange rate should strengthen in real terms in a few years' time (see Figure 4:5).", "The forecast for the exchange rate in the longer term is based on assessment of the development of the real exchange rate."
6	27 (SQ)	Repo rate cuts	repo', 'riksbank', 'rate', 'path', 'cut', 'decided', 'executive', 'board', '25', 'scenario	"The Executive Board of the Riksbank has therefore decided to cut the repo rate to 1 per cent and to adjust the repo-rate path downwards.", "The Executive Board of the Riksbank has therefore decided to cut the repo rate to 0.25 per cent."
7	41	ECB lending, sovereign debt crisis	'ecb', 'loans', 'creditors', 'government', 'bonds', 'bank', 'greece', 'payments', 'impact', 'ceased'	"As government bonds have often been provided as collateral for the ECB loans, creditors with non-covered bank bonds will now have lower priority in the event of a suspension of payments.", "As the risk of suspensions of payments has increased, the positive impact of the ECB loans has faded during the spring and summer."
8	26	Euro area GDP growth	euro', 'area', 'growth', 'quarter', 'gdp', 'second', 'france', 'weak', 'germany', 'italy	"GDP growth in the euro area 1.6 per cent, third quarter.", "GDP growth in the euro area 1.7 per cent in first quarter."
9	8 (SQ)	Housing construction	housing', 'construction', 'investment', 'prices', 'house', 'owned', 'tenant', 'apartments', 'homes', 'family	"The lower housing prices are expected to lead to significantly lower housing construction in the coming years, compared with the recent years high levels.", "The situation on the housing market has improved, but the construction of housing is still at a low level."
10	3	Repo rate discussions	repo', 'rate', 'path', 'raised', 'cut', 'scenario', 'lower', 'main', 'level', 'mortgage	"The repo-rate path is also much lower than was forecast in April.", "As inflation is lower than in the main scenario, the repo rate will also be lower (see Figure 2:8)."

Notes: Rank reflects the magnitude of the coefficient for the corresponding topic-sentiment time series and topic number refers to BERTopic's internal numbering of topics. "(SQ)" means that the squared version of corresponding topic-sentiment time series entered the ridge regression. Top keywords are computed with c-TF-IDF and representative sentences are sentences closest to the cluster centroid of the relevant topic. The topic interpretation is the authors description and interpretation of the topic.

Table 19: Highest-loading Topics for Text-based R&R Regression pre ZLB

Rank	Topic Number	Topic Interpretation	Top Keywords	Representative Sentences
1	15 (SQ)	Lending and mortgages	banks, lending, capital, mortgage, credit, loans, borrowing, companies, households, requirements	"It is mainly households who have increased their borrowing from both banks and mortgage institutions.", "This would increase the banks' funding costs for mortgages and would probably mean that mortgage rates would be raised more than other lending rates."
2	69 (SQ)	Domestic price pressures	pressures, cost, pressure, domestic, inflationary, capacity, expected, rise, moderate, weaker	"Domestic cost pressures are thus expected to increase over the coming years.", "Domestic price pressure has also been somewhat weaker than foreseen earlier."
3	26 (SQ)	Resource utilisation	resource, utilisation, normal, resources, period, unutilised, forecast, economy, lower, slightly	"Resource utilisation will thus increase too.", "Resource utilisation is still high."
4	5	Monetary policy decisions	repo, rate, path, raised, scenario, implied, assumption, forward, cut, rates	"This would call for a higher repo-rate path.", "The repo rate is therefore being cut to 0.25 per cent and the repo-rate path is being revised down."
5	39	Energy markets	electricity, energy, prices, inflation, price, expected, market, rising, rise, power	"Energy prices will also increase.", "Higher oil prices but lower electricity prices."
6	10	National and international fiscal policy	budget, fiscal, public, balance, debt, surplus, policy, government, central, countries	"In the scenario 'public finances crisis in the euro area' it is assumed that the uncertainty over the budget consolidation in a number of European countries will increase.", "Instead, the discretionary fiscal policy in the central government budget for that year has become somewhat more expansionary over time, according to the indicator that measures the effect of rule changes in the central government budget."
7	35 (SQ)	Working hours	hours, worked, number, working, employment, average, time, employed, overtime, quarter	"Both employment and the number of hours worked increased.", "During the coming period, average working hours are expected to rise as the number of hours worked increases."
8	48 (SQ)	Information sources	sources, reuters, ecowin, imf, eurostat, source, riksbank, bureau, national, figure	"Sources: Reuters EcoWin and the Riksbank Figure 3:4.", "Sources: Reuters EcoWin and the Riksbank Figure 3:4."
9	7	Inflation forecasts	inflation, forecasts, forecast, percent, models, expected, uncertainty, period, scenario, years	"US inflation is expected to be low during the forecast period.", "Inflation expected to be 2 percent in two years."
10	32	International outlooks	euro, area, growth, gdp, germany, weak, united, states, quarter, countries	"Euro area: GDP growth and unemployment rate.", "GDP growth in the euro area has been lower than expected this year."

Notes: Rank reflects the magnitude of the coefficient for the corresponding topic-sentiment time series and topic number refers to BERTopic's internal numbering of topics. "(SQ)" means that the squared version of corresponding topic-sentiment time series entered the ridge regression. Top keywords are computed with c-TF-IDF and representative sentences are sentences closest to the cluster centroid of the relevant topic. The topic interpretation is the authors description and interpretation of the topic.

Table 20: Highest-loading Topics for Text-based R&R Regression post ZLB

Rank	Topic Number	Topic Interpretation	Top Keywords	Representative Sentences
1	78 (SQ)	Private sector investments	investment, investments, sector, business, capacity, increase, contribute, need, demand, utilisation	"As the outlook for industrial activity has weakened, growth in other business sector investment has also slowed and will contribute to a slight reduction in total business sector investment as a share of GDP in the coming years." , However, the increasingly high level of resource utilisation means that the corporate sector's need for new investment will increase and investment growth will rise in the coming years. "
2	10	Expansionary monetary policy (negative rates)	repo, rate, raised, unchanged, cut, path, begin, indicates, half, 25	"The repo rate path reflects the fact that there is still a greater probability that the rate will be cut than that it will be raised in the near future, and that slow increases in the repo rate will not begin until the start of 2018 (see figure 1:10)." , "The repo rate is held at -0.50 per cent, but the repo□rate path reflects the fact that there will still be a greater probability that the rate will be cut than that it will be raised in the near future, and that slow increases in the repo rate will not begin until the start of 2018."
3	103	Insurance savings	insurance, savings, collective, understand, reports, publishing, external, parties, schemes, follow	"The insurance companies' relatively large holdings of shares, together with the low interest rates, also make them vulnerable to falling share prices." , "Collective insurance savings consist of savings that households do not control themselves, e.g. premium pensions and collective agreements insurances."
4	7 (SQ)	Confidence indicators	confidence, indicators, indicator, sector, manufacturing, households, tendency, index, industry, managers	"Confidence among households is also very high." , "Total Business Sector Consumer Confidence Indicator (CCI)Economic Tendency Indicator Source: National Institute of Economic Research Figure 3:7."
5	63	Riksbank inflation expectations	cpif, reach, percent, close, expected, inflation, cpi, just, 2016, 2017	"CPIF inflation is expected to be close to 2 percent during 2016." , "CPI inflation is expected to reach 2 per cent at around the same time as the CPIF."
6	55 (SQ)	Riksbank inflation assessments	inflation, riksbank, close, target, assessment, conditions, adapt, information, study, ensure	"There are thus a number of factors that could lead to both lower and higher inflation than is now being assumed, and the Riksbank is prepared to adapt monetary policy to ensure inflation is close to the target of 2 per cent." , "Overall, the Riksbank's assessment is that the conditions for close-to-target inflation remain good."
7	69 (SQ)	The Swedish banking sector	banks, profitability, funding, credit, stable, lending, good, rates, swedish, low	"The Swedish banks' profits have been high and stable in recent years and although there are differences between the banks, profitability is on the whole good." , "The profitability of Swedish banks has been high and stable in recent years despite low and negative interest rates."
8	50	Information sources	eurostat, national, sources, institute, bureau, analysis, research, source, economic, figure	"Sources: Bureau of Economic Analysis, Eurostat, national sources, Office for National Statistics and the Riksbank" , Sources: Bureau of Economic Analysis, Eurostat, national sources, Office for National Statistics and the Riksbank"
9	30	Expansionary monetary policy (unconventional measures)	monetary, policy, riksbank, expansionary, target, make, inflation, prepared, safeguard, anchor	The expansionary monetary policy underlines the Riksbank's aim to safeguard the role of the inflation target as nominal anchor for price□setting and wage formation." , "The Riksbank is also preparing further measures that can be used to make monetary policy even more expansionary."
10	17	US-China trade conflict	trade, tariffs, china, conflict, states, united, tariff, barriers, chinese, wto	"The trade conflict between the United States and China is continuing to create significant uncertainty over world trade and global growth." , " Last year, the United States introduced tariffs on half of all imports from China."

Notes: Rank reflects the magnitude of the coefficient for the corresponding topic-sentiment time series and topic number refers to BERTopic's internal numbering of topics. "(SQ)" means that the squared version of corresponding topic-sentiment time series entered the ridge regression. Top keywords are computed with c-TF-IDF and representative sentences are sentences closest to the cluster centroid of the relevant topic. The topic interpretation is the authors description and interpretation of the topic.

Table 21: Highest-loading Topics for Text-based HF Identification pre 2008

Rank	Topic Number	Topic Interpretation	Top Keywords	Representative Sentences
1	10 (SQ)	Bond rates	rates, term, bond, long, real, rate, bonds, nominal, low, short	"Interest rates for the short-est maturities indicate market assessments of monetary policy in the near future, while long bond rates mainly mirror long-term confidence in economic policy as a whole.", "Historically, however, bond rates are still relatively low (Fig. 21)."
2	22	Wage bargaining	agreements, wage, agreement, negotiations, employees, round, bargaining, local, sector, expire	"The difference from previous wage bargaining round is that the white-collar workers' agreements are often somewhat lower due to their historically having had higher wage increases outside the scope of the agreement.", "New wage agreements will be concluded in the spring for employees in the local government sector.",
3	136	Retail sales	retail, sales, turnover, trade, car, everyday, retailing, whole-sale, infrequent, stores	"Retail trade in electronics and furniture rose comparatively rapidly , as did trade in cars and leisure articles.", "Retail turnover and car sales are both weaker at present."
4	24	Riksbank forecasts	riksbank, forecasts, forecast, inflation, forecasters, growth, assessment, models, gdp	"The underlying domestic price tendency was accordingly somewhat lower than the Riksbank envisaged in the October Report. ", "Riksbank forecast for 2000-02."
5	0	Riksbank policy discussions	inflation, target, policy, monetary, based, assessments, measure, term, long, expectations	"In the somewhat longer term, the expectations are more in line with the inflation target.", "On account of the time lag before monetary policy measures affect inflation, the construction of monetary policy is based on assessments of inflation in the coming two years."
6	94	Data sources	statistics, cpi, sweden, 10123456, sources, 12, month, change, percentage, 1995	"Percentage 12-month change Sources: Statistics Sweden and the Riksbank.", "Annual percentage change CPI UNDEX Sources: Statistics Sweden and the Riksbank."
7	124	TCW weighting	tcw, weighting, comprise, includes, weights, 85, arrived, approximately, using, summary	"These together comprise approximately 85 per cent of the total TCW weighting.", "The TCW weighting includes eleven countries, which comprise approximately 85 per cent of the total TCW of twenty countries."
8	83	Expected GDP growth	gdp, growth, expected, year, somewhat, upwards, 2005, compared, revised, states	"GDP growth is also expected to be relatively high this year and next year.", "GDP growth is forecast to be somewhat higher in 2005."
9	73	Pricing of forward rates	forward, implied, rates, day, securities, calculated, curve, funds, fed, implicit	"For forward rates between 4 and 14 years this difference has fallen marginally to about 1.5 per cent since the June Report (Fig. 50).", "Market pricing of forward contracts indicates an expectation of tighter policy stance in the United States of around 0.25 percentage points in spring at the earliest."
10	107	Labour market developments	vacancies, job, number, notices, redundancy, new, unfilled, vacancy, jobs, vacant	"The development of different indicators of the labour market situation, such as the number of new job vacancies and redundancy notices, points to an upturn in the demand for labour (see Figure 30).", "The number of new job vacancies continued to rise at the same time as there was a slight fall in the number of redundancy notices."

Notes: Rank reflects the magnitude of the coefficient for the corresponding topic-sentiment time series and topic number refers to BERTopic's internal numbering of topics. "(SQ)" means that the squared version of corresponding topic-sentiment time series entered the ridge regression. Top keywords are computed with c-TF-IDF and representative sentences are sentences closest to the cluster centroid of the relevant topic. The topic interpretation is the authors description and interpretation of the topic.

Table 22: Highest-loading Topics for Text-based HF Identification post 2008

Rank	Topic Number	Topic Interpretation	Top Keywords	Representative Sentences
1	57	Resource utilisation	utilisation, resource, uses, normal, overall, assessment, indicator, indicators, statistical, riksbank	"The Riksbank's overall assessment is that resource utilisation is at present lower than normal.", "Resource utilisation lower than normal but will rise."
2	134	Domestic prices	prices, produced, sweden, goods, producer, food, producers, price, consumer, paid	"Producer prices for consumer goods are still rising faster than normal, both with regard to imported goods and goods produced in Sweden.", "The upturn is partly explained by prices in the producer channel increasing more rapidly than normal, both prices of imported goods and of goods produced in Sweden (see Figure 3:5)."
3	89 (SQ)	ILO Definitions	ilo, focusing, according, body, organization, life, definition, percent, approximate, note	"According to the ILO definition, full-time students who are seeking work are included in the unemployment concept.", "According to the ILO definition, unemployment averaged 6.2 per cent during the period 1980-2006 and 7.4 per cent during the period 1998-2006."
4	16 (SQ)	Private sector investments	investment, investments, stocks, business, sector, increase, total, need, quarter, expected	"Business sector investment is also expected to fall next year.", "According to Statistics Sweden's investment survey, the business sector's investment plans indicate an increase of investments in 2010, compared with the previous year."
5	3 (SQ)	Repo rate deliberations	repo, rate, path, raised, cut, scenario, lower, main, level, mortgage	"As inflation is lower than in the main scenario, the repo rate will also be lower (see Figure 2:8).", "The repo-rate path is therefore lower than in the main scenario."
6	117 (SQ)	Riksbank press conference information	conference, press, day, held, meeting, following, decision, following, monetary, days	"A press conference is held on the day following the monetary policy meeting.", "A press conference is held on the day following the monetary policy meeting."
7	127 (SQ)	Importance of KIX	transactions, aggregate, kix, important, international, countries, sweden, transaction, weighted, aggregation	"KIX is an aggregate of the countries that are important to Sweden's international transactions.", "KIX is an aggregate of the countries that are important to Sweden's international transactions."
8	115 (SQ)	International outlooks - Euro area and the Baltics	baltic, swedish, consequences, economy, developments, sweden, worsen, solutions, crisis, area	"If the situation in the euro area were to worsen again or if the recovery were to be faster than forecast, this would have consequences for developments in Sweden.", "A deterioration in the Baltic countries would have negative consequences for the Swedish economy."
9	40 (SQ)	Unit labour costs	unit, costs, labour, productivity, increase, means, year, hour, average, expected	"For the full year 2012, unit labour costs are expected to increase by 2.6 per cent.", "The rate of increase in unit labour costs is expected to rise during the forecast period."
10	43 (SQ)	Stock market developments	stock, volatility, equity, markets, exchanges, risen, indices, fallen, levels, share	"The Stockholm Stock Exchange has fallen substantially, in line with other global stock markets, and volatility has been high.", "Stock prices on the Swedish stock market have fallen since the monetary policy decision from July and volatility has risen."

Notes: Rank reflects the magnitude of the coefficient for the corresponding topic-sentiment time series and topic number refers to BERTopic's internal numbering of topics. "(SQ)" means that the squared version of corresponding topic-sentiment time series entered the ridge regression. Top keywords are computed with c-TF-IDF and representative sentences are sentences closest to the cluster centroid of the relevant topic. The topic interpretation is the authors description and interpretation of the topic.

Table 23: Highest-loading Topics for Text-based HF Identification pre ZLB

Rank	Topic Number	Topic Interpretation	Top Keywords	Representative Sentences
1	31	Monetary policy decisions	repo, riksbank, rate, decided, path, executive, board, cut, unchanged, 25	"The Executive Board of the Riksbank has therefore decided to cut the repo rate to 0.25 per cent.", "The Executive Board of the Riksbank has therefore decided to hold the repo rate unchanged at 0.25 per cent and the repo rate is expected to remain at this level until autumn 2010."
2	72 (SQ)	Figures reporting model forecasts	limits, horizontal, tolerance, respectively, lines, broken, target, annual, change, cpi	"The horizontal broken lines at 2, 1 and 3 per cent, respectively, are the Riksbank's inflation target and the tolerance limits for the annual change in the CPI.", "The horizontal lines at 2, 1 and 3 per cent, respectively, are the Riksbank's target and tolerance limits for the annual percentage rate of increase in the CPI."
3	23 (SQ)	Resource utilisation	resource, utilisation, normal, resources, unutilised, period, forecast, economy, lower, picture	"Resource utilisation will thus increase too.", "Resource utilisation will thus increase."
4	18	Productivity growth	productivity, growth, strong, 1990s, high, weak, trend, cyclical, recent, period	"When productivity growth is higher than 1.5 percent as an annual percentage change, the period with weak productivity is considered to be over.", "Productivity growth is still high."
5	10	Oil prices	oil, price, prices, barrel, crude, usd, higher, high, petrol, opec	"The oil price has remained at around USD 50 per barrel since March.", "The expected fall in oil and commodity prices will also counteract a rapid upturn in inflation."
6	15 (SQ)	Mortgage market (focused on the 2008 crisis)	mortgage, lending, banks, institutions, term, credit, borrowing, households, capital, companies	"The contraction has also had an impact on short-term mortgage rates, as the interbank rates affect the mortgage institutions' short-term financing costs.", "However, in Sweden there are as yet no clear signs, in the official statistics, of a slowdown in bank lending to households and companies."
7	37 (SQ)	Riksbank forecasts	riksbank, forecasts, forecast, inflation, forecasters, growth, assessment, gdp, 2005, expected	"Inflation was lower than the Riksbank expected in the forecasts made in 2004, although well in line with the forecasts made in 2005.", "The Riksbank forecasts GDP to increase by almost 2 per cent this year and by just less than 3 per cent in the coming years."
8	59 (SQ)	International and domestic consumer confidence	confidence, consumer, indicators, households, household, indicator, historical, institute, area, dence	"Consumer confidence rose slightly in February and March, but the confidence indicators for both households and companies are still at low levels.", "Confidence indicators for households have also risen and the improvement among German households was particularly striking."
9	61 (SQ)	Information sources	sources, reuters, ecowin, imf, eurostat, source, riksbank, bureau, national, figure	"Sources: Reuters EcoWin and the Riksbank Figure 3:4.", "Sources: Reuters EcoWin and the Riksbank Figure 3:4."
10	38 (SQ)	Working hours	hours, worked, number, working, employment, average, time, employed, overtime, quarter	"Both employment and the number of hours worked increased.", "During the coming period, average working hours are expected to rise as the number of hours worked increases."

Notes: Rank reflects the magnitude of the coefficient for the corresponding topic-sentiment time series and topic number refers to BERTopic's internal numbering of topics. "(SQ)" means that the squared version of corresponding topic-sentiment time series entered the ridge regression. Top keywords are computed with c-TF-IDF and representative sentences are sentences closest to the cluster centroid of the relevant topic. The topic interpretation is the authors description and interpretation of the topic.

Table 24: Highest-loading Topics for Text-based HF Identification post ZLB

Rank	Topic Number	Topic Interpretation	Top Keywords	Representative Sentences
1	17	Labour market demographics	born, population, immigration, force, participation, age, labour, people, refugee, employment	"Among Swedish-born persons, the labour force participation rate is now very high." , "It takes time for new arrivals to establish themselves on the labour market, however, and foreign-born persons have on average a lower labour force participation rate and employment rate than the working-age population as a whole."
2	3 (SQ)	Wage costs	wage, costs, increases, unit, agreements, labour, wages, productivity, growth, increase	"Demand in the Swedish economy has been high for a long period, wage growth is expected to rise somewhat and companies' costs, measured in terms of unit labour costs, will increase by about 2 per cent per year." , "Wage growth according to short-term wage statistics is expected to be 2.6 per cent this year."
3	42	Labour shortages	shortages, shortage, reporting, companies, labour, recruitment, tendency, survey, sector, business	"The proportion of companies reporting labour shortages increased further in the second quarter according to the Economic Tendency Survey." , "According to the National Institute of Economic Research's Economic Tendency Survey, the proportion of companies reporting labour shortages has risen to a high level and the average recruitment time in the business sector is long (see Figure 4:11)."
4	7	Housing investment and construction	housing, construction, investment, prices, starts, supply, homes, demand, fall, price	"The part of investment that has increased the most in recent years has been housing construction." , "The recent decline in housing prices is expected to lead to some slowdown in housing construction and thereby in housing investment."
5	24 (SQ)	Implications of US-China trade conflict	trade, tariffs, china, states, conflict, united, tariff, countries, goods, barriers	B"ut if the trade conflict between the United States and China were to spread and lead to more countries establishing trade barriers, the effects will be greater." , "Increased uncertainty due to the trade conflict between the United States and China is also subduing imports (see the article "World trade in constant change")."
6	30	Expansionary monetary policy	monetary, policy, riksbank, expansionary, target, make, inflation, prepared, safeguard, anchor	The expansionary monetary policy underlines the Riksbank's aim to safeguard the role of the inflation target as nominal anchor for price-setting and wage formation." , "The Riksbank is also preparing further measures that can be used to make monetary policy even more expansionary."
7	11 (SQ)	Household and company lending conditions	households, lending, companies, loans, borrowing, rates, bank, funding, rate, income	"The banks' continuing good funding conditions mean that lending rates to households and companies remain on historically low levels." , "Lending rates for households and companies are still low."
8	12	Government bond purchases	bonds, government, purchases, sek, billion, nominal, holdings, riksbank, bond, purchase	"Government bond purchases will continue until the end of 2016." , "At the monetary policy meeting on 28 April, the Executive Board chose to hold the repo rate unchanged, but to extend the government bond purchases by SEK 40-50 billion."
9	19	Stock market developments	stock, markets, equity, volatility, prices, market, exchanges, risen, financial, share	"Stock prices on the Swedish stock market have fallen since the monetary policy decision from July and volatility has risen." , "Equity prices have risen in Sweden and abroad, while volatility has fallen on the stock market."
10	14 (SQ)	Oil prices	oil, price, fall, prices, effects, production, lower, supply, opec, sanctions	"A price fall due to reduced demand for oil will not have the same positive effects on the global economy as a price fall due to an increased supply of oil." , "The effects of the fall in oil prices on inflation will thereby be increased."

Notes: Rank reflects the magnitude of the coefficient for the corresponding topic-sentiment time series and topic number refers to BERTopic's internal numbering of topics. "(SQ)" means that the squared version of corresponding topic-sentiment time series entered the ridge regression. Top keywords are computed with c-TF-IDF and representative sentences are sentences closest to the cluster centroid of the relevant topic. The topic interpretation is the authors description and interpretation of the topic.

D The Riksbank Corpus

Table 25 present descriptive statistics of the Riksbank corpus by communication regime and document type. Based on the fourth column of Table 25, it can be noted that the sentences used in the documents across all regimes are on average of similar length. It is also clear that MPRs generally include more material than inflation reports, and that MPUs are short in comparison. While there are differences in document length based on type, it can be noted that the quantity of total communication via these reports in each regime is rather stable (ranging from 961.9 to 1288.5)¹⁸

Table 25: Summary statistics across communication regimes.

Document Type	Years	# Documents	Average Sentence Length (words)	Average Document Length (sentences)
Inflation Report	1996–2006	42	20.020	1191.360
MPR	2007–2008	6	20.190	1628.330
MPU	2007–2008	2	19.310	269.000
MPR	2009–2014	18	21.210	1636.940
MPU	2009–2014	17	20.300	415.470
MPR	2015–2019	30	22.320	961.900

Notes: MPR denotes Monetary Policy Report and MPU denotes Monetary Policy Update. Up until 2009, Riksbank documents were published quarterly and the vast majority were released in March, June, September and December. Between 2009 and 2014, MPRs were published in February, July and October and MPUs were published in April, September and December. In the remainder of the sample, MPRs replaced MPUs in April, September and December.

While length of documents provide some information about the corpus, more important, from an text-analytical point of view, is information about the document content across the communication regimes. If there are substantial differences in what is discussed in the different types of documents, for example if the bulk of the inflation reports discuss concepts that become obsolete with the passing of time and changing macroeconomic conditions, the empirical model developed in the main paper would need to account for that. To verify whether this is the case in the Riksbank corpus, I estimate BERTopic on all textual data and report for each document type and communication regime the share of sentences allocated a topic (*Topic Share*), the number of identified topics and the average share of each document that consists of the top 20 topics (*Top 20 Share*). The results are reported in Table 26. This estimation is based on the full-sample optimal value for the threshold parameter in the text-based R&R setting ($v^* = 59$). This practice of estimating the model on the

¹⁸In the period 2007-2008, the average document length for all published reports were 1288.5, between 2009 and 2014 it was 1043.43.

Table 26: Summary statistics across communication regimes.

Document Type	Years	Topic Share	# Topics	Top 20 Share
Inflation Report	1996–2006	0.618	98	0.680
MPR	2007–2008	0.507	93	0.583
MPU	2007–2008	0.515	65	0.675
MPR	2009–2014	0.483	97	0.609
MPU	2009–2014	0.614	94	0.639
MPR	2015–2019	0.515	99	0.653

Notes: MPR denotes Monetary Policy Report and MPU denotes Monetary Policy Update. Topic Fraction refers to the share of sentences classified as non-outlier in each regime by BERTopic when the topic model is estimated on the full sample. The number of topics reflect how many unique topics are present in each regime. The top 20 share reflects the fraction of all sentences in each regime that belong to the 20 most frequent topics.

full sample and displaying results for sub-periods deliberately disregards potential heterogeneity in document content with respect to time.

The share of sentences allocated a topic is rather stable over the communication regimes and document types, as is the average share of the text that appear in the 20 most frequent topics. This suggest similarity of document content and structure over time. Since a majority of the text in the documents across each communication regime appear in the top 20 topics, investigating these provides additional reassurance of document similarity over time. Table 27 report top three key-words for each top 20 topic ranked by occurrence for each document type over time. Representative sentences for each topic is left out in the interest of space, but can be provided upon request.

While the exercise of identifying topics with BERTopic is fully data-driven, it is clear that (i) the topic model identifies topics with economic interpretation and meaning and (ii) there is substantial overlap in the top 20 topics across the document types and communication regimes. The striking similarity in document content over communication regimes and document types reported in Table 27 provides no motivation for adjusting the NLP pipeline developed in Section 3.1.

Table 27: Top 20 Topics by Document Type and Communication Regime

1996-2006	2007-2008 (MPR)	2007-2008 (MPU)
(0) inflation, prices, target	(2) inflation, expectations, forecasts	(1) inflation, cpif, cpi
(1) inflation, cpif, cpi	(7) oil, price, barrel	(27) euro, area, gdp
(10) lending, loans, banks	(5) wage, agreements, increases	(2) inflation, expectations, forecasts
(7) oil, price, barrel	(4) krona, appreciation, weaker	(7) oil, price, barrel
(9) g32, g101, g116	(6) consumption, household, saving	(4) krona, appreciation, weaker
(6) consumption, household, saving	(37) hours, worked, number	(6) consumption, household, saving
(5) wage, agreements, increases	(15) yields, bond, government	(14) swedish, economy, growth
(2) inflation, expectations, forecasts	(8) housing, construction, investment	(37) hours, worked, number
(24) inflation, 99, port	(18) investments, stocks, sector	(23) jan, dec, mar
(3), repo, rate, path	(9) g32, g101, g116	(20) fiscal, budget, balance
(4) krona, appreciation, weaker	(1) inflation, cpif, cpi	(8) housing, construction, investment
(27) euro, area, gdp	(45) sek, exchange, scale	(9) g32, g101, g116
(26) labour, market, employment	(90) repo, sweden, rate	(33) represent, lines, forecast
(12) expansionary, monetary, policy	(31) debt, household, ratio	(5) wage, agreements, increases
(11) exports, imports, trade	(11) exports, imports, trade	(3), repo, rate, path
(18) investments, stocks, sector	(40) statistics, sweden, sources	(18) investments, stocks, sector
(22) that, raised, long	(20) fiscal, budget, balance	(26) labour, market, employment
(15) yields, bond, government	(13) rates, real, term	(16) productivity, growth, 1990s
(20) fiscal, budget, balance	(3), repo, rate, path	(72) japan, growth, asia
(25) born, labour, population	(12) expansionary, monetary, policy	(38) employment, force, fall
2009-2014 (MPR)	2009-2014 (MPU)	2015-2019
(1) inflation, cpif, cpi	(1) inflation, cpif, cpi	(1) inflation, cpif, cpi
(4) krona, appreciation, weaker	(21) riksbanken, scb, källor	(3), repo, rate, path
(2) inflation, expectations, forecasts	(2) inflation, expectations, forecasts	(2) inflation, expectations, forecasts
(3), repo, rate, path	(4) krona, appreciation, weaker	(8) housing, construction, investment
(5) wage, agreements, increases	(6) consumption, household, saving	(17) inflation, sweden, chapter
(14) swedish, economy, growth	(3), repo, rate, path	(4) krona, appreciation, weaker
(6) consumption, household, saving	(11) exports, imports, trade	(5) wage, agreements, increases
(8) housing, construction, investment	(8) housing, construction, investment	(13) rates, real, term
(7) oil, price, barrel	(5) wage, agreements, increases	(19) resource, utilisation, normal
(12) expansionary, monetary, policy	(12) expansionary, monetary, policy	(30) riksbank, forecasts, model
(11) exports, imports, trade	(16) productivity, growth, 1990s	(14) swedish, economy, growth
(13) rates, real, term	(14) swedish, economy, growth	(16) productivity, growth, 1990s
(21) riksbanken, scb, källor	(29) exchange, real, rate	(6) consumption, household, saving
(16) productivity, growth, 1990s	(27) euro, area, gdp	(9) g32, g101, g116
(33) represent, lines, forecast	(59) food, commodity, agricultural	(58) note, cpix, undinhx
(15) yields, bond, government	(34) shortage, reporting, tendency	(66) gdp, exports, inventory
(28) exports, swedish, demand	(13) rates, real, term	(7) oil, price, barrel
(37) hours, worked, number	(17) inflation, sweden, chapter	(34) shortage, reporting, tendency
(35) gdp, quarter, percent	(28) exports, swedish, demand	(23) jan, dec, mar
(23) jan, dec, mar	(7) oil, price, barrel	(39) companies, prices, costs

Notes: The number in parenthesis refer to BERTopic's internal numbering of topics, the ordering of topics reflect topic prevalence in each regime and the top three keywords are selected by class-based TF-IDF. See Grootendorst (2022) for details on c-TF-IDF.

E Bootstrap Confidence Intervals

The monetary policy shocks used in the local projection analysis are estimated rather than observed. The text-based R&R shocks are residuals from a ridge regression and the text-based HF shocks are fitted values from a regression where the single regressor itself is a residual from a ridge regression. In addition, the text-based shocks used in the paper are the result of a model selection algorithm. That is, the optimization algorithm developed in Section 3.1 selects the optimal parameterization of $\mathbf{W}$, $\mathbf{W}(v^*)$, by estimating Equation (9) for each $v \in \mathcal{V}$. The text-based shocks can therefore be written as $\hat{\varepsilon}_t^{i-text} = \hat{\varepsilon}_t^{i-text}(\hat{\beta}_{v^*, \hat{\lambda}}^{ridge}, v^*)$ for $i \in \{RR, HF\}$, where $\hat{\beta}_{v^*, \hat{\lambda}}^{ridge}$ denotes the ridge estimator at v^* and $\hat{\lambda}$ is the ridge penalty selected using leave-one-out cross-validation.

The local projection standard errors treat the shocks as observed rather than estimated, which ignores the uncertainty that come from the first-stage estimation procedure. Standard confidence intervals will therefor understate the true uncertainty around the impulse response estimates.

To account for this generated-regressor problem (see e.g. Chernozhukov et al. (2018) and Battaglia et al. (2024) on the inference on machine-learning generated regressors), I base inference on a bootstrap procedure that creates a bootstrap distribution of impulse response estimates for each outcome variable and horizon. The proposed bootstrap re-estimates, for each bootstrap sample, the shock-construction model (Equation (9) for the R&R case and Equation (12) and (13) for the HF case) and local projection model for all outcome variables and horizons, and stores the resulting estimates of the impulse response estimates. I then construct bootstrap confidence intervals based on the quantiles of the distribution of bootstrap impulse response estimates.

To propagate the uncertainty of finding v^* into such bootstrap confidence bands, v_b^* should be identified for each bootstrap sample $b \in B$. That is, Equation (9) and (12) should be estimated for all $v \in \mathcal{V}$, for all $b \in B$. Currently, fitting BERTopic once takes me around 4 minutes. With a grid of 100 values for v and with at least 500 bootstrap replications, this computation would take at least $4 \times 100 \times 500 = 200.000$ minutes ≈ 139 days, which is not feasible for me. The solution I implement is therefore unable to properly account for the uncertainty that arise from my selection of v^* , but rather takes the v^* identified from the true sample as given and propagates the uncertainty from $\hat{\beta}_{\hat{\lambda}}$ into the LP confidence bands. Details on my bootstrap implementation is found below.

E.1 Bootstrap Algorithm

Let $\mathcal{D} = \{y_t, W_t\}_{t=1}^T$ denote the dataset used in the first-stage shock estimation¹⁹. To preserve serial dependence in the underlying time series, I use a moving block bootstrap to construct bootstrap samples. Specifically, I construct overlapping blocks of length L , $B_1 = (\mathcal{D}_1, \dots, \mathcal{D}_L)$, $B_2 = (\mathcal{D}_2, \dots, \mathcal{D}_{L+1}), \dots$, where $\mathcal{D}_t$ denotes the complete observation vector at date t . Bootstrap samples are generated by drawing blocks with replacement and concatenating them until a pseudo-sample of length T is obtained. To reflect short run persistence in macroeconomic conditions and policy decisions, I choose $L = 12$. Due to changes in meeting frequency, this corresponds to around 1.5 – 2 years of meeting observations for most of the sample.

For each bootstrap replication $b = 1, \dots, B$, the following steps are performed:

1. Generate a moving block bootstrap sample: $\mathcal{D}^{(b)} = \{y_t^{(b)}, W_t^{(b)}\}_{t=1}^T$
2. Re-estimate the ridge regression on $\mathcal{D}^{(b)}$:
 - Select the optimal regularization parameter, $\hat{\lambda}^{(b)} = \arg \min_{\lambda} CV^{(b)}(\lambda)$,
 - Estimate ridge coefficients $\hat{\beta}^{(b)} = \hat{\beta}_{\hat{\lambda}^{(b)}}$.
3. Construct bootstrap shocks: $\hat{\varepsilon}_t^{(b)} = y_t - X_t' \hat{\beta}^{(b)}$.
4. Aggregate the shock series to quarterly and annual frequencies by summation.
5. For frequency f , outcome variable y and horizon h , re-estimate all local projections with the bootstrap shock, $\hat{\varepsilon}_t^{(b)}$, as impulse variable. Store the impulse response coefficients, $\hat{\theta}_{f,y,h}^{(b)}$.

With B repetitions, I get the bootstrap distribution $\left\{ \hat{\theta}_{f,y,h}^{(1)}, \hat{\theta}_{f,y,h}^{(2)}, \dots, \hat{\theta}_{f,y,h}^{(B)} \right\}$ for each combination f, y, h .

E.2 Confidence Intervals

Let $\hat{\theta}_{f,y,h}$ be the full-sample local projection impulse response estimate for frequency f , outcome variable y and horizon h . To conduct inference on $\hat{\theta}_{f,y,h}$, I use the distribution of bootstrap

¹⁹The regressors used in the first-stage shock estimation consists of the benchmark R&R specification and the optimal set of topic-sentiment series. That is $W_t = W_t(v^*)$. To ease notation, I refer to this quantity simply as W_t in this section.

estimates. However, as the bootstrap distribution is generated in a two-step procedure that involves model section on bootstrap samples, it may be the case that $\hat{\theta}_{f,y,h} \neq \frac{1}{B} \sum_{j=1}^B \hat{\theta}_{f,y,h}^{(j)}$. To account for this, following Davison and Hinkley (1997), I employ a centered percentile bootstrap, which corrects for such finite-sample shifts by centering the bootstrap distribution before constructing confidence intervals. First, I define the centered bootstrap distribution, $\left\{ \tilde{\theta}_{f,y,h}^{(1)}, \tilde{\theta}_{f,y,h}^{(2)}, \dots, \tilde{\theta}_{f,y,h}^{(B)} \right\}$ where $\tilde{\theta}_{f,y,h}^{(b)} = \hat{\theta}_{f,y,h}^{(b)} - \frac{1}{B} \sum_{j=1}^B \hat{\theta}_{f,y,h}^{(j)}$ capture the deviation in estimating $\hat{\theta}_{f,y,h}$ with bootstrap sample b relative to the average estimate of $\hat{\theta}_{f,y,h}$ across all bootstrap replications. This centered bootstrap distribution now contains deviations from the bootstrap mean for the local projection estimate for frequency f , outcome variable y and horizon h . The $100(1 - \alpha)\%$ confidence interval for $\hat{\theta}_{f,y,h}$ is then given by

$$\hat{\theta}_{f,y,h} + q_{\alpha/2}(\tilde{\theta}_h^{(b)}), \hat{\theta}_{f,y,h} + q_{1-\alpha/2}(\tilde{\theta}_h^{(b)}),$$

where $q_p(\cdot)$ denotes the empirical p -quantile of the bootstrap distribution.

F Robustness Tests

F.1 Alternative Text-Analytic Approach

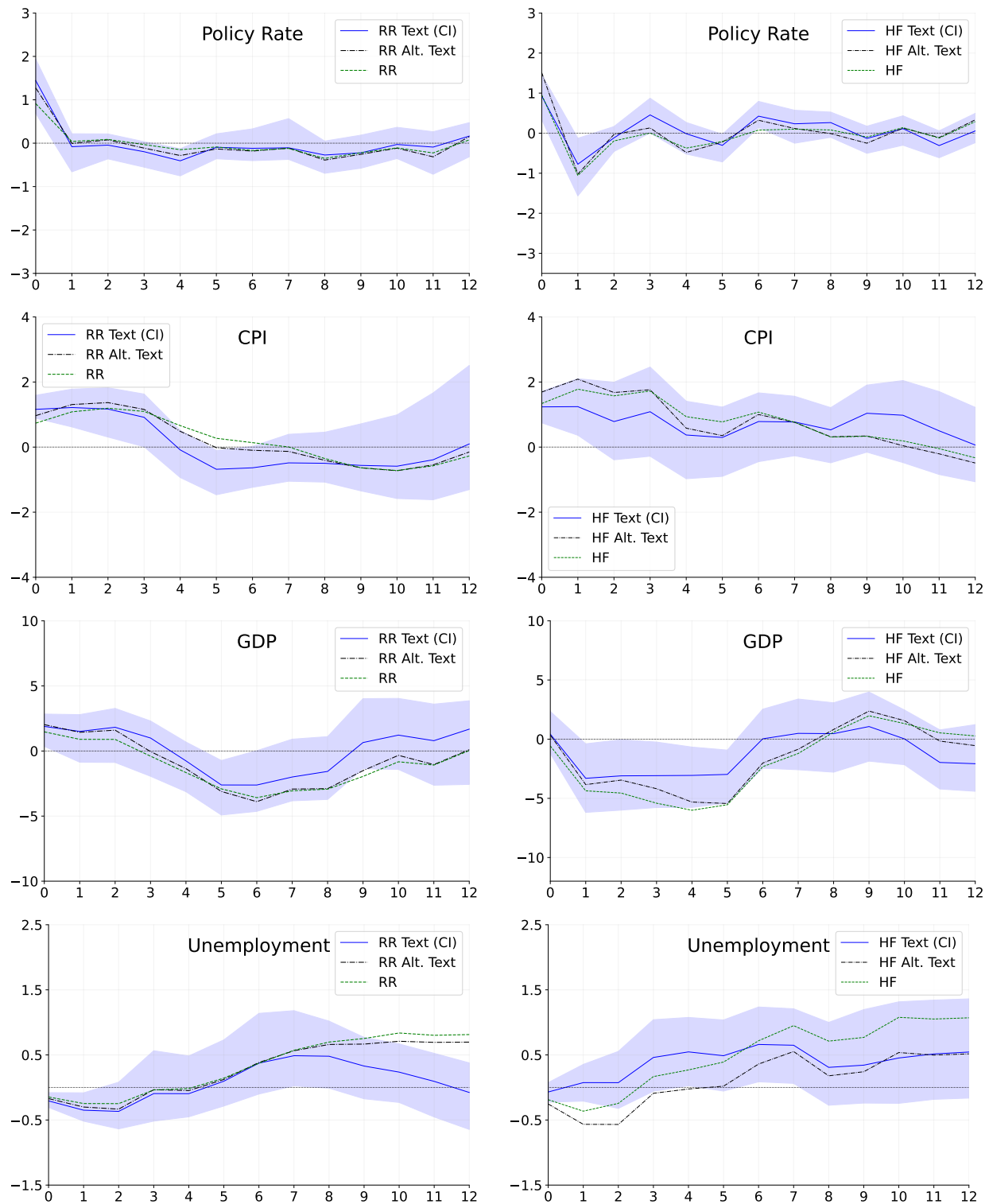


Figure 10: LPs of Shocks Identified with Alternative NLP Approaches, Quarterly Data

Notes: Two lags of the outcome and shock are included as controls. 95% bootstrap confidence intervals are plotted in blue.

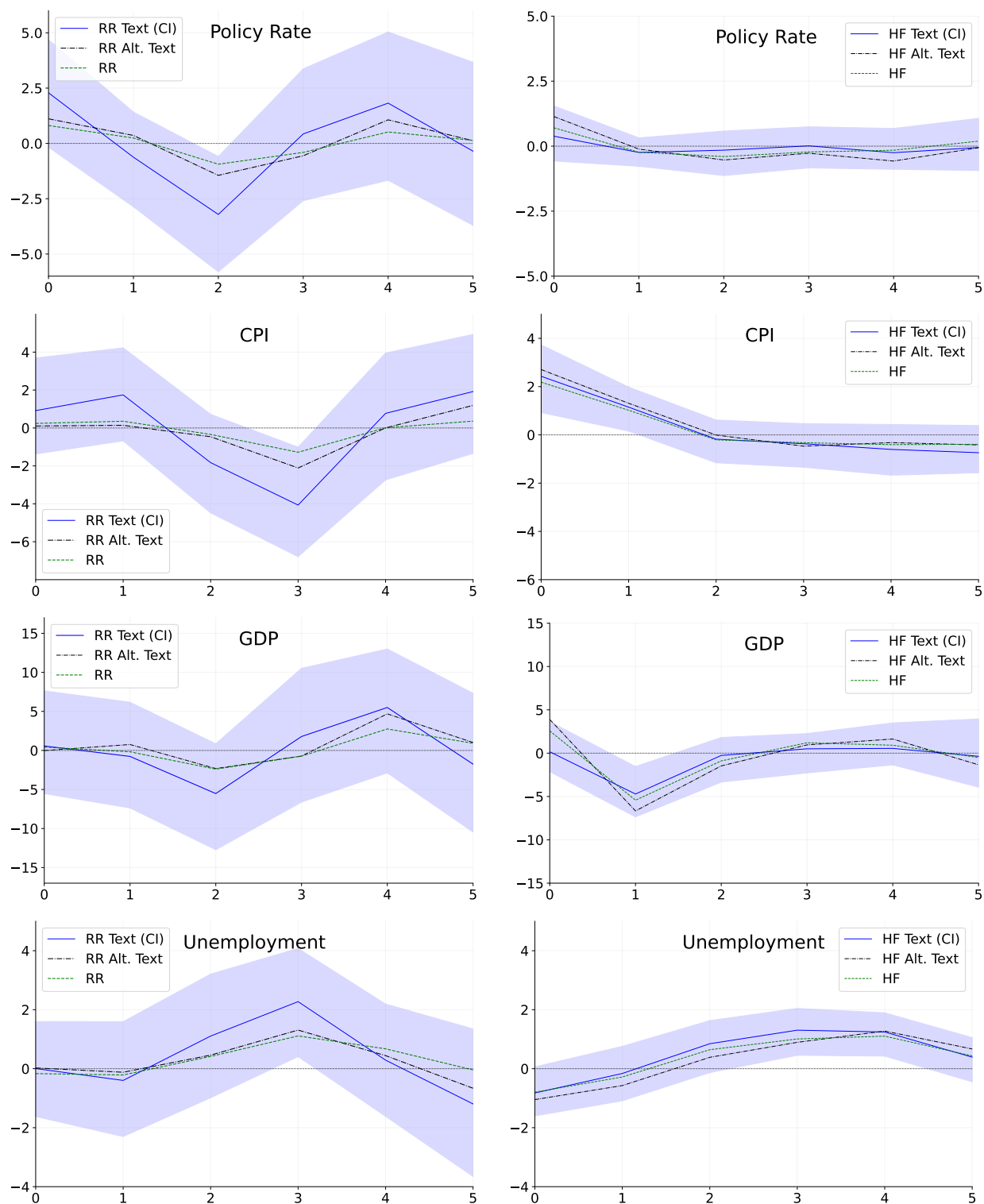


Figure 11: LPs of Shocks Identified with Alternative NLP Approaches, Annual Data

Notes: One lag of the outcome and shock are included as controls. 95% bootstrap confidence intervals are plotted in blue.

F.2 Alternative LP Specifications

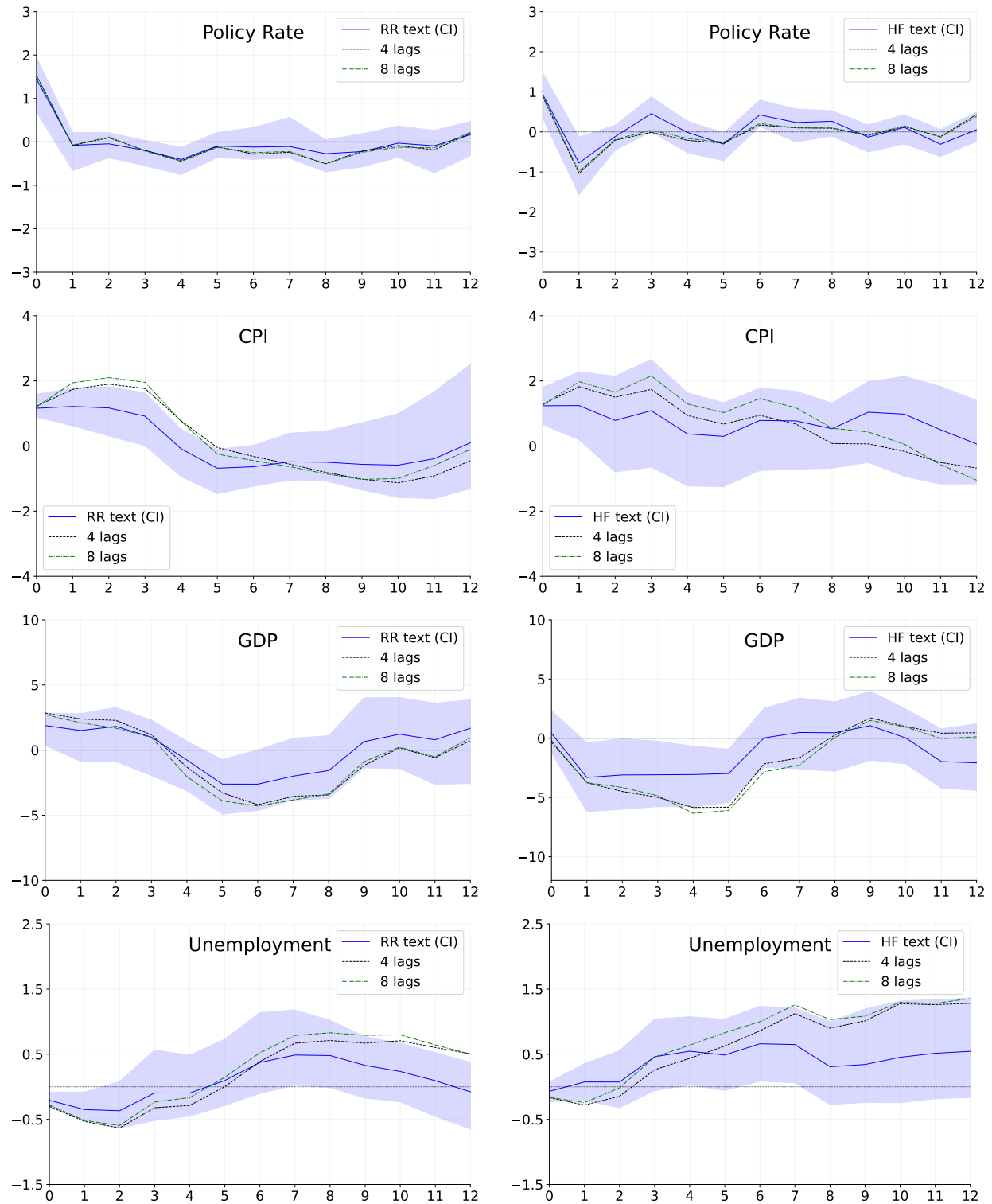


Figure 12: LPs Estimated with Further Lags of the Outcome, Quarterly Data

Notes: Baseline includes two lags of the outcome. Two lags of the shock is always included. 95% bootstrap confidence intervals are plotted in blue.

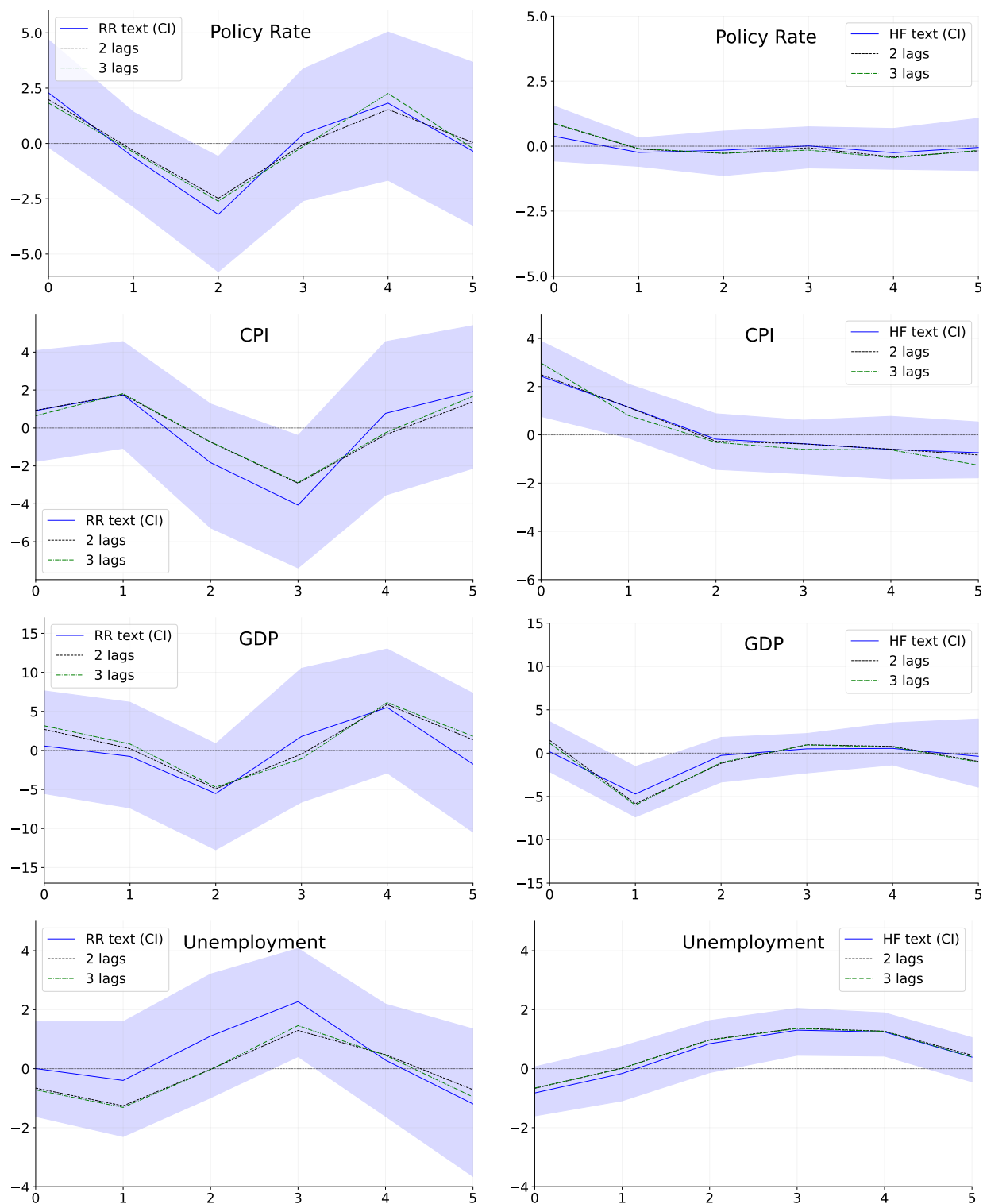


Figure 13: LPs Estimated with Further Lags of the Outcome, Annual Data

Notes: Baseline includes one lag of the outcome. One lag of the shock is always included. 95% bootstrap confidence intervals are plotted in blue.

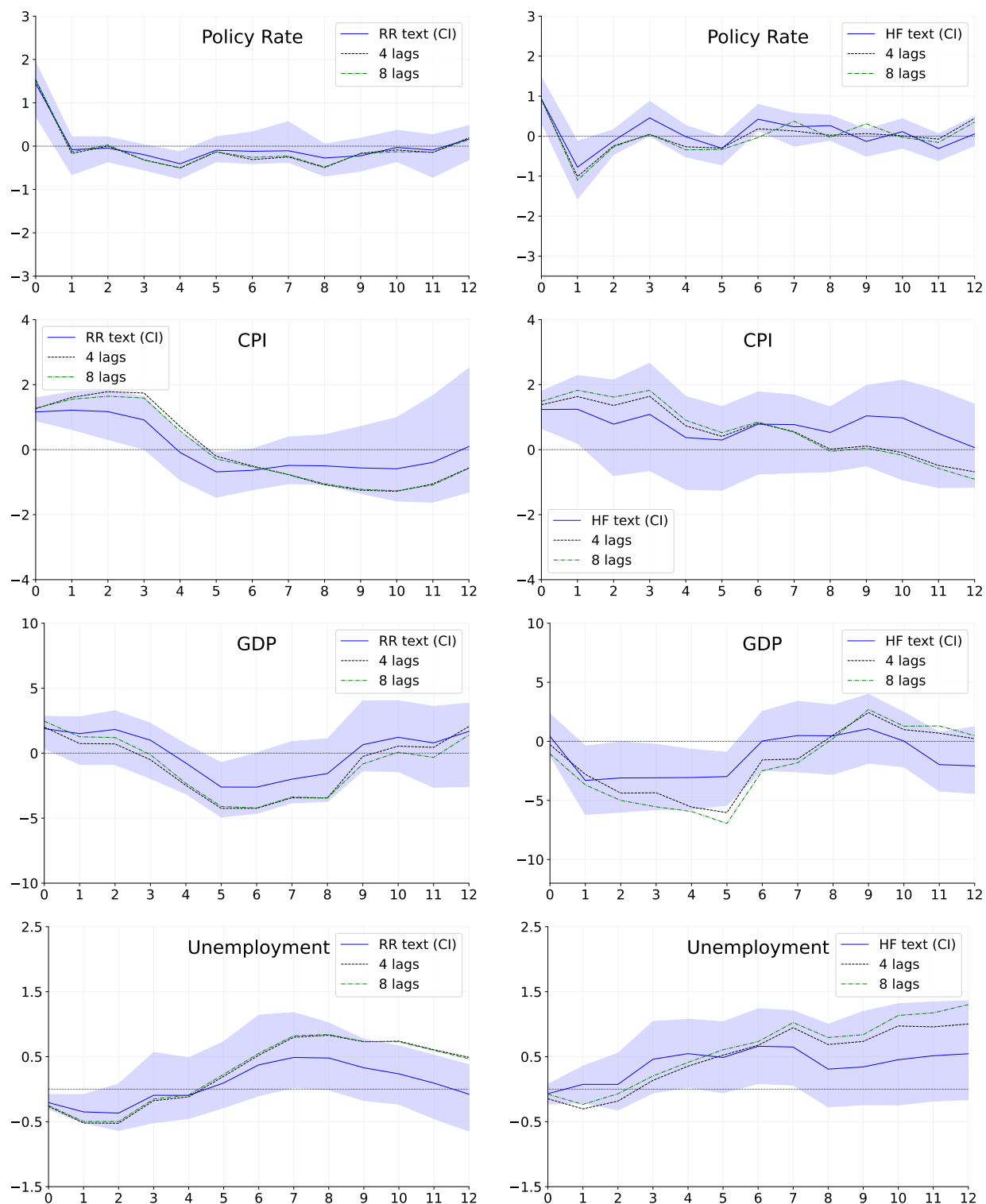


Figure 14: LPs Estimated with Further Lags of the Shock, Quarterly Data

Notes: Baseline includes two lags of the shock. Two lags of the outcome is always included. 95% bootstrap confidence intervals are plotted in blue.

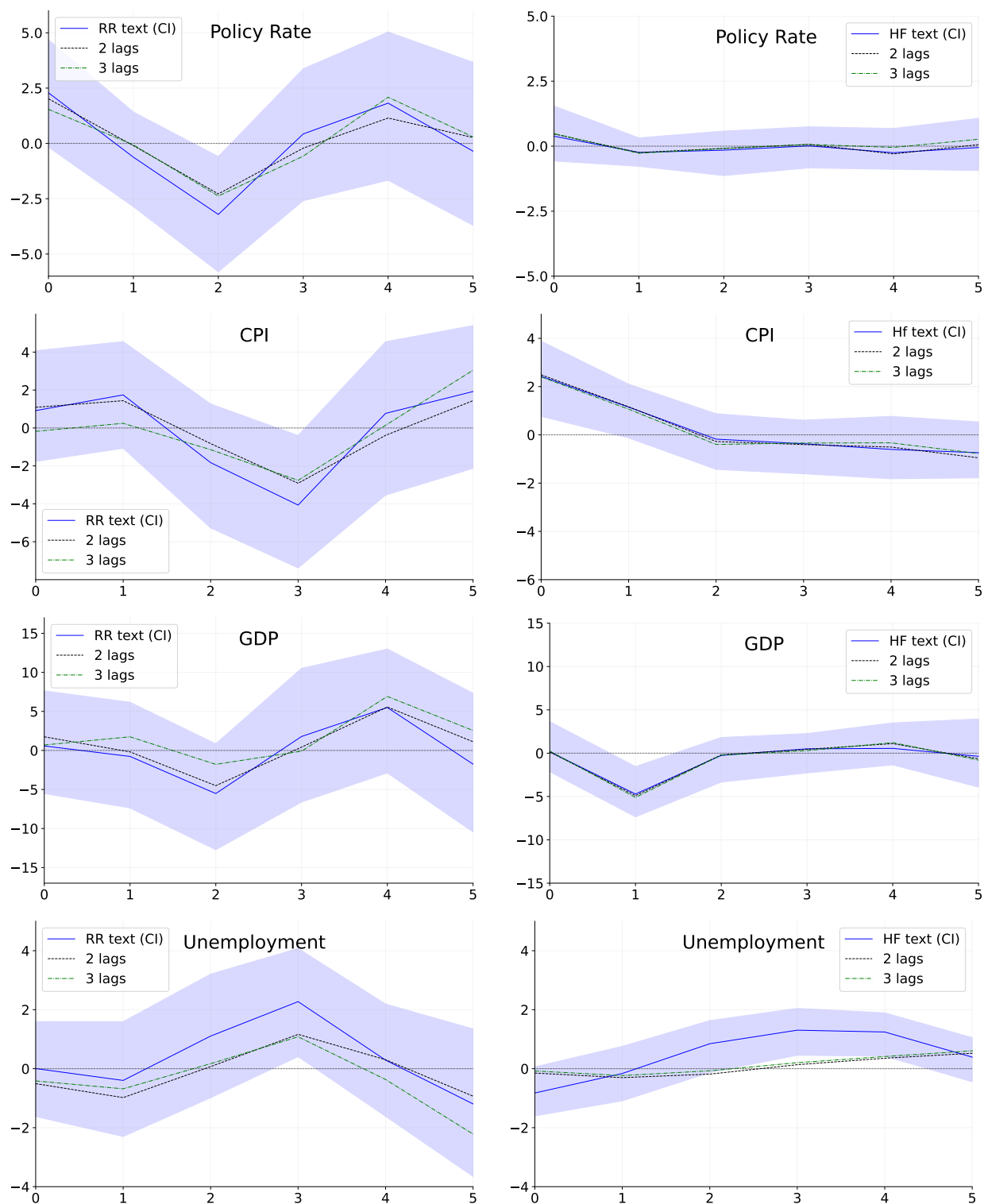


Figure 15: LPs Estimated with Further Lags of the Shock, Annual Data

Notes: Baseline includes one lag of the shock. One lag of the outcome is always included. 95% bootstrap confidence intervals are plotted in blue.

F.3 Alternative Greenbook Specification

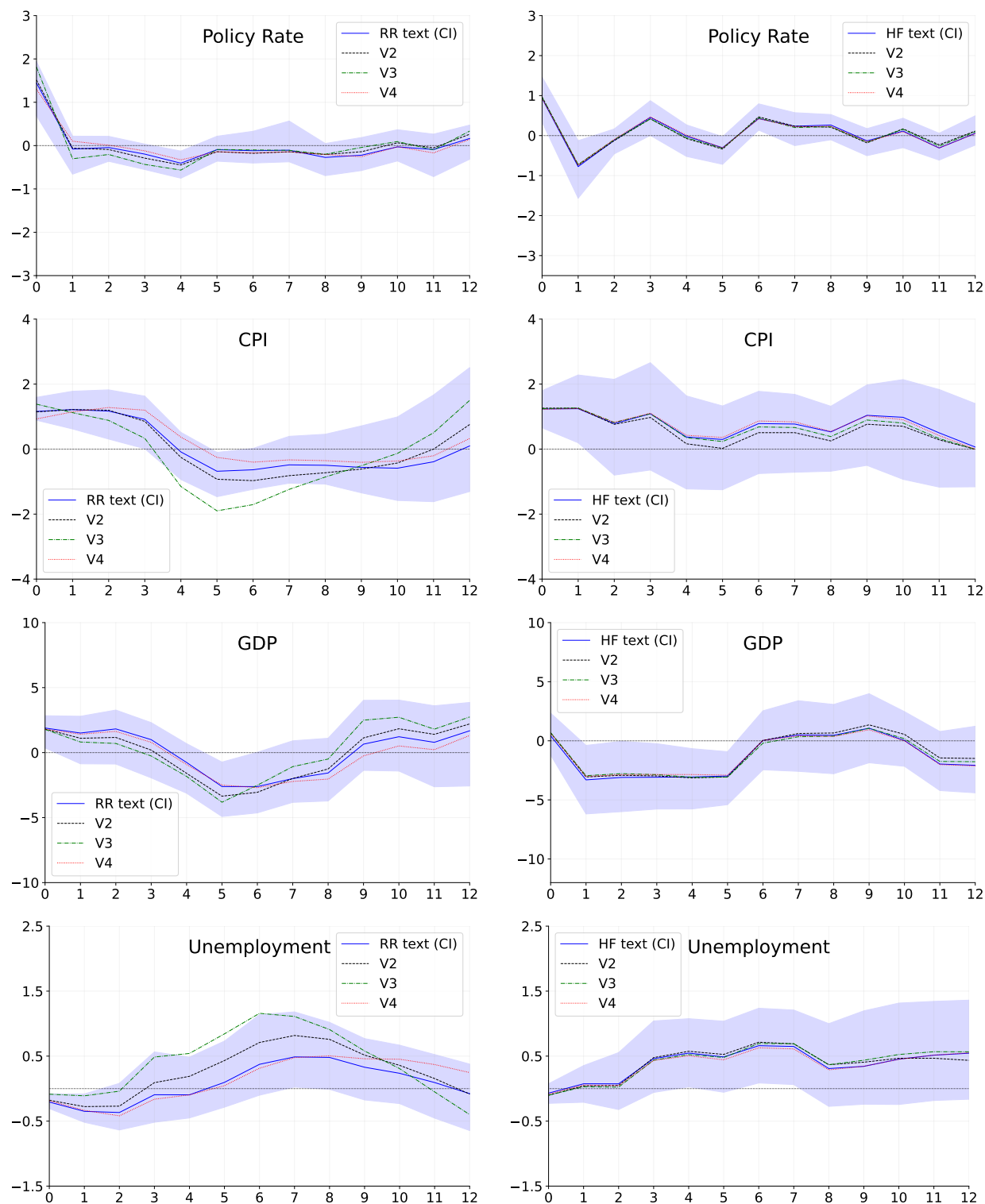


Figure 16: LPs for Shocks Estimated with Alternative Greenbook Specifications, Quarterly Data

Notes: Two lags of the outcome and shock has been used as control in each projection. 95% bootstrap confidence intervals are plotted in blue.

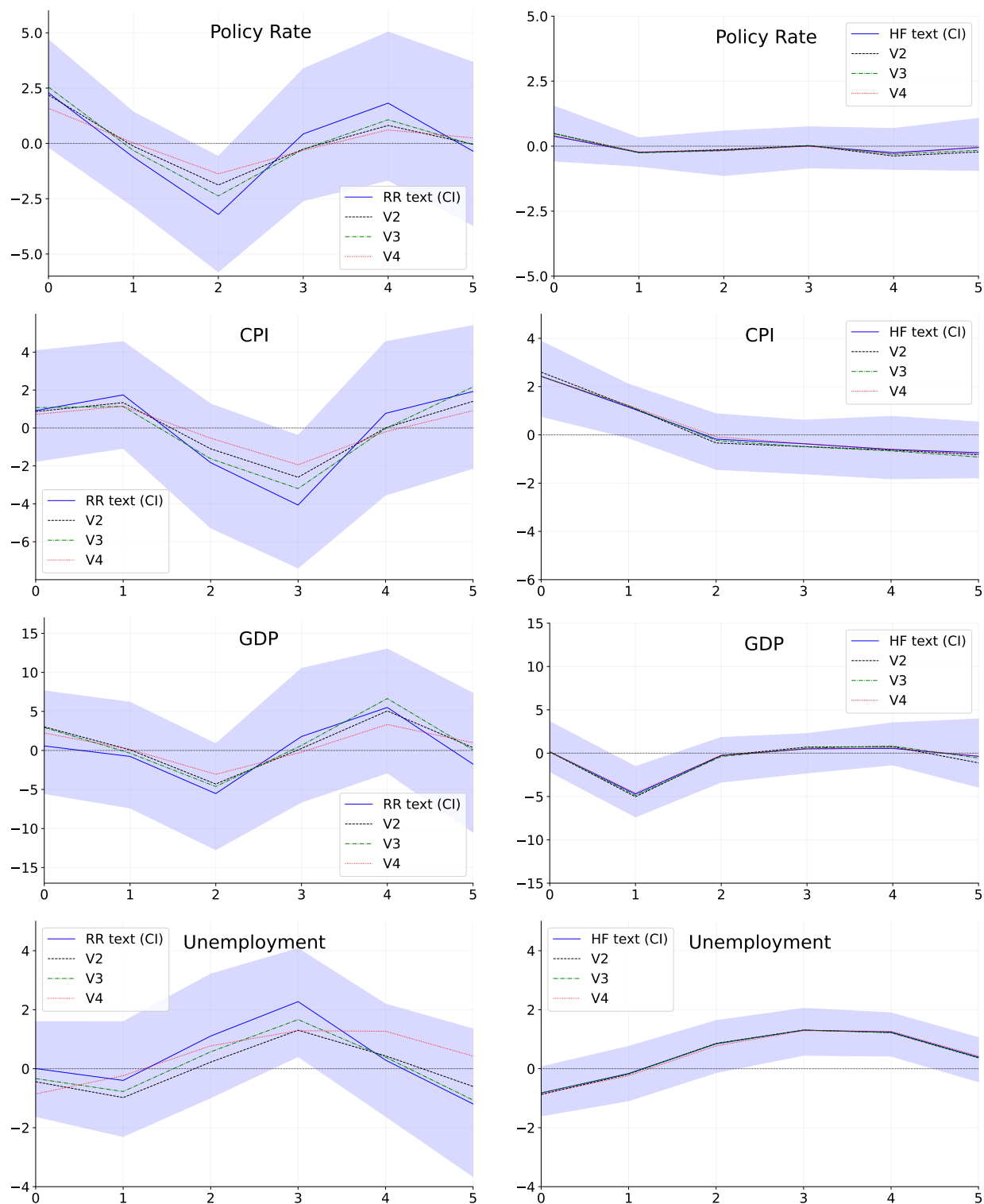


Figure 17: LPs for Shocks Estimated with Alternative Greenbook Specifications, Annual Data

Notes: One lag of the outcome and shock has been used as control in each projection. 95% bootstrap confidence intervals are plotted in blue.

I LLM Primer (Intended as Online Appendix)

Disclaimer and Limitations

In this paper, I use two LLMs to construct topic-sentiment series. CentralBankRoBERTa is used to classify the sentiment of sentences while BERTopic is used to identify common themes, topics, in the entire corpus. Both models build on Devlin et al.’s (2019) BERT model, CentralBankRoBERTa in that it is a fine-tuned version of Liu et al.’s (2019) RoBERTa which is based on BERT, BERTopic in that it by default uses a BERT-based embedding model (Grootendorst, 2022). The aim of this appendix is to provide a formal, but still intuitive, account on how this type of models work. In doing so, I will provide an introduction to embedding models in general, and to self-attention networks (SANs) in particular as these are workhorses in the Transformers architecture, upon which the current state-of-the-art language models (BERT and GPT) builds. Then I will briefly cover how CentralBankRoBERTa is trained, and how sentence-level embeddings (used in default BERTopic and in the paper) differ from word embeddings.

In order to deliver a formal account of SANs while explaining the intuition behind the different steps, I will abstract away from a few things. First, I do not describe how parameters are learned during training, for the purpose of this appendix I assume solution concepts exist. They do exist, see Vaswani et al. (2017) and Xiao and Zhu (2023) for a discussion of Transformers and Devlin et al. (2019) on BERT. Second, I will not explain the full Transformers architecture which would be needed to fully understand the construction of embeddings in models such as BERT and BERTopic. It could however be useful to think of the additional components as similar to the nonlinear ($\tanh(\cdot)$) activation in the neural n-gram model, see below. For a richer introduction to transformers, see Vaswani et al. (2017). Third, I disregard the tokenization step of modern neural language models. While it is conceptually useful to think of the inputs, w_i , as words represented by one-hot encoded vectors and subsequent embeddings, modern implementations rely on sub-word tokenization schemes which differ from this simplification.

II Introduction

Statistical methods operate on numbers. The starting point in the analysis of textual data is therefore to convert words into numbers, i.e. to construct a numerical representation of text. One simple way to represent the words $W_{1:n} = [w_1, \dots, w_n]$ is to assign each word a one-hot encoding, i.e. $w_1 = [1, 0, \dots, 0]$, $w_2 = [0, 1, 0, \dots, 0]$, $\dots$, $w_n = [0, 0, \dots, 1]$ where $w_i \in \mathbb{R}^n \forall i$. Such representation is central in how dictionary methods^a work. While simple, such an encoding carry no semantic, or contextual, information about the words. One consequence of this is an inability to differentiate the meaning of the word *not* in the sentences $s_1 = \{\text{this is not good}\}$ and $s_2 = \{\text{this is not bad}\}$ as w_{not} has the same vector in both s_1 and s_2 . An embedding, in contrast, is a way of mapping each word into a vector space where similar words have similar vectors. Since they allow for a rich representation of words, embedding models are the workhorses of modern textual analysis tools.

^aDictionary methods are a way of assessing the *sentiment* of text which was popularized in Loughran and McDonald (2011). In general, such methods rely heavily on a simple counting of (one-hot encoded) words who belong to pre-defined lexicons.

Formally, let $\mathbb{E} \in \mathbb{R}^{n \times d}$ represent the embedding matrix of the data, where the embedding dimension is d and the number of observations is n , then the embedding of w_i is computed by multiplying the one-hot vector of the word with the embedding matrix. That is $emb(w_i) = w_i \mathbb{E} = [e_{i,1}, e_{i,2}, \dots, e_{i,d}] \in \mathbb{R}^d$, a dense^b representation of w_i in which each element encodes a certain feature of w_i . Importantly, $\mathbb{E}$ is a parameter in a neural network and thus a quantity that is learned during training (Xiao and Zhu, 2023). Once trained, $emb(w_i)$ is fixed, why $\mathbb{E}$ is often referred to as an embedding look-up table (Zhang and Teng, 2021). This means that although $emb(w_i)$ is a dense vector representation of w_i and able to capture semantic information, it is static in the sense that each word in the vocabulary will have a fixed vector representation across sentences (Zhang and Teng, 2021). Consequently, $emb(not)$ will be identical in s_1 and in s_2 which is not ideal.

As such, I will only consider neural language models that allow for a *context-dependent* representation of words which here is defined as the hidden states of words in sequence encoder layers in a neural language model. Concretely, I will introduce Self-Attention Networks (SANs), a sequence encoder network that pool bidirectional information along the sequence (of word embeddings) to allow for a rich representation of words. SANs are at the core of the transformers architecture introduced in Vaswani et al. (2017) upon which the well known Bidirectional Encoder Representation from Transformers (BERT) model builds (Devlin et al., 2019).

Before introducing SANs, let us briefly introduce the components in a neural network. A neural network can be seen as a function approximator in that it takes an input, processes it through one or more intermediate transformations to produce an output such as a probability or numerical prediction. Given its prediction task, the network is able to learn a, possibly nonlinear and highly complex, relation between inputs and outputs. Here, the inputs are vectors of real numbers (the word embeddings) and the intermediate transformations are called hidden layers, see Figure 18.

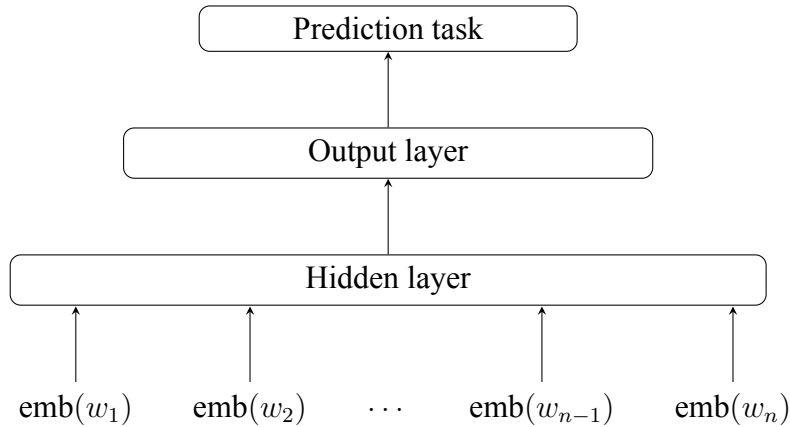


Figure 18: A schematic of a neural network for prediction.

Notation-wise, the inputs are the embeddings, $\mathbf{X} = [emb(w_1), emb(w_2), \dots, emb(w_n)]^T \in \mathbb{R}^{n \times d}$ in which the i th row of $\mathbf{X}$ contains the embedding of the i th word, $emb(w_i)$, which is a row vector of dimension d . To ease notation, the input layer is typically referred to as an initial hidden layer, $\mathbf{H}_0 = \mathbf{X}$, a convention I follow. The hidden layer is denoted $\mathbf{H} = [h_1, \dots, h_n]^T$, in which the inputs are subject to linear transformations followed by nonlinear activation and fed up

^bEmbedding dimensions are often between 100-1000, typically much smaller than the dimension of the data (Zhang and Teng, 2021). In the sense that $d \ll n$, $\mathbb{E}$ is a dense representation of the words.

in the network. In Bengio et al. (2003)’s neural n-gram language model, this transformation is $\mathbf{H} = \tanh(W^h \mathbf{X} + b^h)$ where W^h and b^h are parameters of the hidden layer which, together with $\mathbf{X}$, are learned during training. That is, the embeddings $\mathbf{X} = [\text{emb}(w_1), \text{emb}(w_2), \dots, \text{emb}(w_n)]^T$ are explicitly treated as another parameter in the model (Zhang and Teng, 2021). The corresponding SAN operations will be covered in detail in Section III.a. Individual h_i $i \in 1, \dots, n$ in the above formula are referred to as the hidden representation, or hidden state, of word i . I will only consider networks where $h_i \in \mathbb{R}^d$, i.e. networks in which the dimensions of the hidden representations are the same as the inputs. In a model with one hidden layer, we take $\mathbf{H}$ as the embeddings of $W_{1:n}$.

We are, however, not restricted to using one single hidden layer, deep neural networks typically consists of many hidden layers stacked on top of each other. One example of this is the BERT base model which consists of 12 SAN (along with Feedforward, Residual, Layer normalization and positional encoding network^c) layers. In this case, we refer to the k^{th} hidden layer as $\mathbf{H}_k = [h_1^k, \dots, h_n^k]^T$ where an arbitrary h_i^k is referred to as the hidden representation, or hidden state, of word i at layer k . With a model with K hidden layers, we take either the final hidden layer $\mathbf{H}_K$ or $\mathbf{H} = \sum_{k=1}^K s_k \mathbf{H}_k$, where s_k are weights that sum up to one, as the context-dependent embeddings of $W_{i:n}$. Such embeddings are context dependent as they are calculated dynamically for a given sentence using parameters that capture the surrounding words, the context (Zhang and Teng, 2021).

III Contextualized Embeddings

III.a Self-Attention Networks

As indicated in Section 1, taking $\text{emb}(w_1), \dots, \text{emb}(w_n)$ as our embeddings means that we represent each vocabulary word with a fixed vector. We are thus unable to account for contextual nuances which, as seen in the example sentences s_1 and s_2 , is very important for the understanding of words. Words do, after all, obtain their meaning from the context in which they appear.

One popular way of accounting for the context in a word sequence is to use a set of mathematical operations called *self-attention*. Self-attention enter a neural language model as a hidden layer, and it has the appeal that it allow for a modeling of dependencies along the sequence without regard to the distance in the sequence (Bahdanau et al., 2014). Before self-attention, the modeling of dependence was done through recurrent and/or convolutional networks (RNNs and CNNs), compared to which self-attention offer technical advantages in the modeling of said dependence structure as well as several computational advantages (Vaswani et al., 2017). An illustration of a neural network with two SAN layers can be found in Figure 19.

Here, the SANs are hidden layers in a model that predicts, w_i given its context using a masked language modeling objective which is one of BERTs learning tasks (Devlin et al., 2019). What is important here is that each hidden representation of words account for the surrounding context instantaneously, and not sequential as would have been the case with a recurrent layer (Vaswani et al., 2017). This is what all $\text{emb}(w_j)$ pointing to all h_j^1 which in turn points to all h_j^2 is meant to illustrate, i.e $h_i^k = f(h_1^{k-1}, \dots, h_n^{k-1})$, for some (nonlinear) function f , and not $h_i^k = f(h_{i-1}^{k-1}, h_i^{k-1})$ which would have been the case in a recurrent layer. In such a recurrent layer, context enter via

^cThese type of network will not be covered here, but can be thought of as replacing the $\tanh(\cdot)$ operation of the neural n-gram model. Formally, these network introduces a nonlinear transformation to the output of the self-attention layers (weighted averages of query-key-value information) to capture nonlinear patterns in the data while making sure the information content in the inputs are not lost (Devlin et al., 2019; Xiao and Zhu, 2023).

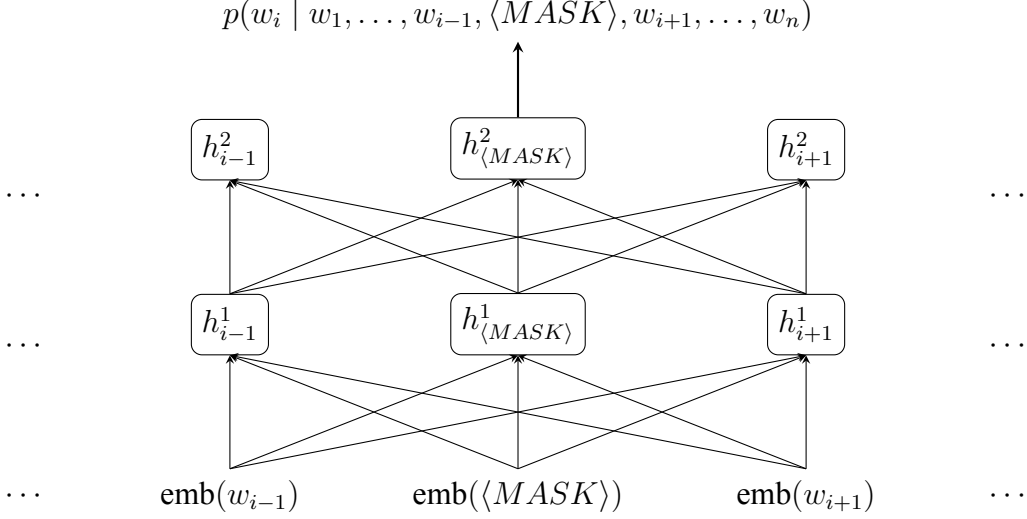


Figure 19: Contextualized embeddings using SAN, inspired by Zhang and Teng (2021)

h_{i-1}^k only which can be computed both one-, and bidirectional (Zhang and Teng, 2021).

Using notation from Section 1, a K layer version of the model in Figure 19 can be written as:

$$\begin{aligned}
 \mathbf{H}_0 &= [\text{emb}(w_1), \dots, \text{emb}(w_{i-1}), \text{emb}(\langle \text{MASK} \rangle), \text{emb}(w_{i+1}), \dots, \text{emb}(w_n)]^\top, \\
 \mathbf{H}^k &= \text{SAN}(\mathbf{H}^{k-1}), \quad \forall k \in \{1, \dots, K\}, \\
 \mathbf{p} &= \text{softmax}(W\mathbf{h}_i^k).
 \end{aligned} \tag{16}$$

Let us now turn to what happens in each SAN layer, i.e. parameterize the function $f()$. Similar to the transformer architecture, introduced in Vaswani et al. (2017), we use Query-Key-Value (QKV) self-attention which define three types of information to be used when learning a representation of words. The three parameters of QKV attention are denoted Q , K and V which are, as always, learned during training, and we use their weighted sum^d for representing the embeddings.

Intuitively, self-attention can be interpreted through an analogy of a database retrieval process (Zhang and Teng, 2021). Each word in a sentence can be thought of as a record in a database, where the record is defined by three fields, a query, a key, and a value. When computing self-attention, each word formulates a search request through its query, which is then compared against the keys associated with all other words in the sequence. The degree of similarity between a query and the available keys determines the relevance of each word in relation to the search, which is what we call attention score. These scores are normalized to form attention weights, which indicate how important each context word is. A word’s final representation is then a weighted average of the values it has retrieved from all words across the sequence (Bahdanau et al., 2014).

This means we can assign an interpretation to the matrices Q , K and V . In this regard, queries captures what information each (hidden representation of a) word ”asks” for, the keys capture how each word ”advertises” the information it has while the values contain the actual information content the word can pass over to other words (Zhang and Teng, 2021).

^dI focus on scaled dot product attention, used in Vaswani et al. (2017)’s Transformers and in Devlin et al. (2019)’s BERT. Other weighting schemes exist.

Formally, with hidden states from layer $k - 1$, QKV attention is defined as:

$$\mathbf{H}_k = \text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^\top}{\sqrt{d_k}}\right)V \quad (17)$$

where $Q = \mathbf{H}_{k-1}W^Q$, $K = \mathbf{H}_{k-1}W^K$ and $V = \mathbf{H}_{k-1}W^V$, $Q, K, V \in \mathbb{R}^{n \times d}$. I will now explain each of the components in turn.

1. Projection into queries, keys and values

In equation 17, $\mathbf{H}_{k-1}$ is projected onto query, key and value spaces

$$\begin{aligned} Q &= \mathbf{H}_{k-1}W^Q \\ K &= \mathbf{H}_{k-1}W^K \\ V &= \mathbf{H}_{k-1}W^V \end{aligned} \quad (18)$$

with $W^Q, W^K, W^V \in \mathbb{R}^{d \times d}$. For each word i , this means that $q_i = h_i^{k-1}W^Q$, $k_i = h_i^{k-1}W^K$ and $v_i = h_i^{k-1}W^V$ where $q_i, k_i, v_i \in \mathbb{R}^d$. Formally, Q, K, V is a way of splitting the representation of each tokens into three different roles whose content are learned during training.

To compute the first hidden layer, $\mathbf{H}_1$, we project $\mathbf{H}_0 = \mathbf{X}$ onto Q, K and V .

2. Attention scores

Once we have Q, K, V , we compute similarity scores between the queries and keys using scaled dot product:

$$S = \frac{QK^\top}{\sqrt{d}} \in \mathbb{R}^{n \times n} \quad (19)$$

where $S_{ij} = \frac{q_i k_j^\top}{\sqrt{d}}$ is a scalar capturing the relevance of word j in the representation of word i (Xiao and Zhu, 2023). Similar to before, d is the embedding dimension and n the dimension of the data.

3. Softmax normalization

When attention scores have been computed, they are normalized into attention weights with a row-wise softmax operation. That is,

$$A = \text{softmax}(S) = \text{softmax}\left(\frac{QK^\top}{\sqrt{d_k}}\right) \quad (20)$$

where A_{ij} captures how much attention word i places on word j . If A_{ij} is large, word j is important in the representation of word i .

4. Weighted sum of queries, keys and values

Finally, the hidden states of the current layer is computed by combining the information of all context words, weighted by their attention weights:

$$\begin{aligned}\mathbf{H}_k &= AV \in \mathbb{R}^{n \times d} \\ h_i^k &= \sum_{j=1}^n A_{ij} v_j \in \mathbb{R}^d\end{aligned}\tag{21}$$

5. An Illustrative Example

Assume we want to calculate embeddings for the words in the sentence "the cat sat on the mat". Then q_{sat} could encode "who sat?", k_{sat} could encode "verb, past tense" and v_{sat} could encode some spacial information. If we were to compute the attention for "sat", we would want its query to look at all keys in the context (i.e. the other words in the sentence) and expect to find the key for "cat" highly relevant. It may also give some weight to "mat". Here, k_{cat} and k_{mat} could encode "subject-like noun" and "plausible object", respectively. Since the word "the" would not as useful for representing "cat" given this query, the score $S_{cat,the} = q_{cat} k_{the}^\top$ would likely be small. Indeed, for the query "who sat?", the word "on" would also likely be unimportant. Then the representation of "sat" will be updated as a weighted average of the values of "cat", "mat", "the" and "on".

III.b Multi-head Attention

The past section points to a crucial limitation to the self-attention network as defined in Section III.a, namely that the importance of words j in the representation of word i depends heavily on the query. If q_{sat} in the above example was "where did something sit?" instead of "who sat?", the attention weight for "mat" would likely be much higher than the weight for "cat". This limitation follows from us using only one set of projections, W^Q, W^K, W^V when computing queries, keys and values, thus forcing the model to learn one kind of relationship at a time (Vaswani et al., 2017). Multi-head attention introduced in Vaswani et al. (2017) solves this issue by repeating the self-attention mechanism in parallel on subspaces of the embedding space (Xiao and Zhu, 2023). The result of each such operation is called an *attention head*, all of which are concatenated and projected back to the embedding space (Vaswani et al., 2017). While it is not possible to dictate what type of queries will be asked, having more attention heads rather allow for more queries that "ask questions" which allows the model to capture different types of information, as long as it helps minimizing loss during training. Formally, we compute each attention head h using equation 17 as

$$head_h = Attention(Q_h, K_h, V_h) \in \mathbb{R}^{n \times d_h}\tag{22}$$

where Q_h, K_h, V_h are again projections, but now computed with $W_h^Q, W_h^K, W_h^V \in \mathbb{R}^{d \times d_h}$ where the attention head dimension d_h is smaller than the embedding dimension. It is common to use the same d_h for all attention heads, and $d_h = d/m = 64$ where m is the number of attention heads has been empirically identified as a sweet spot of performance and stability (Vaswani et al., 2017; Devlin et al., 2019). In the case of BERT base, the model has an embedding dimension of $d = 768$ which mean it uses $m = 12$ attention heads each with $d_h = 64$.

Once all attention heads have been computed, we can concatenate them and project them back to the embedding dimension, d , with $W^O \in \mathbb{R}^{(m \cdot d_h) \times d}$ and use that as our hidden states:

$$\mathbf{H}_k = \text{MultiHeadAttention}(Q, K, V) = [\text{head}_1 \oplus \dots \oplus \text{head}_m]W^O \in \mathbb{R}^{n \times d} \quad (23)$$

Since W^O can be seen as weights, this projection back to the embedding dimension allows the model to learn which attention heads capture more relevant information for the final representation of the words.

III.c CentralBankRoBERTa

With this knowledge of how embeddings are created, it is possible to better understand how CentralBankRoBERTa makes sentiment classifications. First, it is important to recall that thinking about a sentence being represented as $s_i = (w_1, \dots, w_n)$ where $w_i \forall i$ is a word is conceptually correct, but nevertheless a simplification. In practice, BERT models begin with passing all words through a tokenizer which allows constructing a tokenized representation of each sentence. The tokenizer is a method of splitting words into sub-words which can increase the linguistic nuances captured with embeddings, see Zhang and Teng (2021) for more details. Let $x^{(i)} = ([CLS], \tilde{w}_1, \dots, \tilde{w}_m, [SEP])$ be the tokenized representation of s_i where $[CLS]$ is a special token that is learned to represent the entire sentence, $\tilde{w}_i$ is a tokenized representation of w_i (one word may be represented by several tokens so $n \neq m$) and $[SEP]$ indicates end of sentence^e. For a given sentence, the embeddings of the tokenized inputs are computed using self-attention as described in Section III.a and III.b.

CentralBankRoBERTa, is a binary classifier which uses the $[CLS]$ token to make predictions. With the embedding of this token for the i^{th} sentence, $\mathbf{h}_{[CLS]}^{(i)} \in \mathbb{R}^d$, CentralBankRoBERTa performs the following linear transformation for each sentence i :

$$\mathbf{z}^{(i)} = \mathbf{W}\mathbf{h}_{[CLS]}^{(i)} + \mathbf{b},$$

where $\mathbf{W} \in \mathbb{R}^{2 \times d}$ and $\mathbf{b} \in \mathbb{R}^{2 \times 1}$ are parameters in the model and the vector $\mathbf{z}^{(i)} = (z_1^{(i)}, z_2^{(i)})$ contain scalar scores. As before, d represents the embedding dimension which is 768 in RoBERTa-base, the model used in CentralBankRoBERTa (Liu et al., 2019; Pfeifer and Marohl, 2023). These scores are then converted into probabilities using the softmax function:

$$p_\theta(y = k | x^{(i)}) = \frac{\exp(z_k^{(i)})}{\sum_{j=1}^2 \exp(z_j^{(i)})}, \quad k \in \{1, 2\},$$

where $p_\theta(y = k | x^{(i)})$ is the probability that sentence s_i belongs to class k given its tokenized representation $x^{(i)}$.

With a training dataset $\mathcal{D} = \{(x^{(i)}, y^{(i)})\}_{i=1}^N$ where $y^{(i)} = (y_1^{(i)}, y_2^{(i)}) \in \{0, 1\}^2$ is a one-hot encoded label indicating whether sentence s_i is positive or negative, the model parameters are

^eThe two pre-training objectives in BERT is masked language prediction, which is described in Figure 19, and next sentence prediction (NSP). The NSP task is performed using the $[CLS]$ token only, which, based on the inner mechanics of QKV-attention, makes the $[CLS]$ token capture information that represents the entire sentence s_i .

trained by minimizing cross-entropy loss between predicted probabilities and true labels^f:

$$\mathcal{L}(\theta) = - \sum_{i=1}^N \sum_{k=1}^2 y_k^{(i)} \log(p_{\theta}(y = k|x^{(i)})),$$

$$\theta^* = \arg \min_{\theta} \mathcal{L}(\theta).$$

where θ contains all parameters of the neural network. This means that embedding, transformers and classification parameters are all updated based on the classification task.

Let $p_1 = p_{\theta^*}(y = +1|s_i)$ be the probability that sentence s_i is positive and $p_2 = p_{\theta^*}(y = -1|s_i)$ the probability that a sentence is negative under the optimal set of parameters, θ^* . The predicted sentiment of the sentence s_i is then given by the following rule:

$$\text{sentiment}(s_i) = \begin{cases} +1 & \text{if } p_1 > p_2, \\ -1 & \text{if } p_1 < p_2, \end{cases}$$

so that $\text{sentiment}(s_i) \in \{-1, 1\}$.

III.d All-MiniLM-L6-V2

In *Heterogenous Labor Market Impacts of Monetary Policy Shocks*, we use the default embedding model in BERTopic called All-MiniLM-L6-V2, a model that can be understood as a compact version of the BERT model that has been further trained to produce sentence-level embeddings. While BERT is trained on word (token)-level prediction tasks, All-MiniLM-L6-V2 is fine-tuned with a contrastive learning objective designed to capture semantic similarity between *sentences*. Formally, such a learning objective means that the model is presented with pairs of sentences and learns to distinguish positive (true) from negative (false) pairs. This results in the model aligning semantically similar sentences closer in the embedding space (Wang et al., 2020). When applied to the Riksbank documents, this property of our embedding model means that sentences with related meaning or with similar economic content are represented by vectors that reflects this semantic proximity. To use a clustering algorithm in such an embedding space is thus expected to result in groups of sentences in which similar topics are discussed. This is the main idea behind BERTopic.

To construct sentence-level word embeddings, All-MiniLM-L6-V2 first internally produces the word embeddings by feeding the words in each sentence through stacked multi-head SAN layers as described in Section III.b. For the sentence $s = w_1, \dots, w_n$ this gives contextualized word embeddings $h_1, \dots, h_n \in \mathbb{R}^d$ which are turned into sentence embeddings, $\mathbf{s}$, by mean-pooling the individual word embeddings^g:

$$\mathbf{s} = \frac{1}{n} \sum_{i=1}^n h_i \quad (24)$$

Formally, All-MiniLM-L6-v2 is built upon MiniLM-L6-H384-uncased, which is itself a distilled variant of BERT base (Wang et al., 2020). Distillation is a form of model compression

^fFor more details about the training procedure, see RoBERTa documentation: https://github.com/huggingface/transformers/blob/main/src/transformers/models/roberta/modeling_roberta.py and CentralBankRoBERTa documentation: <https://github.com/Moritz-Pfeifer/CentralBankRoBERTa>. Latest access: 16/10 2025.

^gSee the All-MiniLM-L6-V2's training script, `train_script.py`, on their HuggingFace Page for more details. Accessed 3/10-2025.

in which a smaller “student” network is trained to reproduce the behavior of a larger “teacher” network, thereby retaining most of the representational capacity at a fraction of the computational cost (Hinton et al., 2015; Wang et al., 2020). The MiniLM distillation reduces the number of hidden layers from BERT’s twelve to six and compresses the hidden dimension from 768 to 384, while preserving the Transformer architecture with self-attention (Wang et al., 2020). That is, the embedding dimension with All-MiniLM-L6-v2 is $d = 384$ compared to BERT’s $d = 768$. As of October 2025, MiniLM-L6-v2 ranks among the leading models for sentence embeddings in terms of performance, efficiency, and popularity, see the DataLoop website: https://dataloop.ai/library/model/sentence-transformers_all-minilm-l6-v2/.