

# DISCUSSION PAPER SERIES

DP21390

## **IDENTIFYING MONETARY POLICY SHOCKS IN NEWSPAPERS USING GPT**

Felix Betz, Peter Bofinger, Jonas Dix and Leonie  
Streit

**MONETARY ECONOMICS AND  
FLUCTUATIONS**

**CEPR**

# IDENTIFYING MONETARY POLICY SHOCKS IN NEWSPAPERS USING GPT

*Felix Betz, Peter Bofinger, Jonas Dix and Leonie Streit*

Discussion Paper DP21390

Published 15 April 2026

Submitted 14 April 2026

Centre for Economic Policy Research  
187 boulevard Saint-Germain, 75007 Paris, France  
2 Coldbath Square, London EC1R 5HL  
Tel: +44 (0)20 7183 8801  
[www.cepr.org](http://www.cepr.org)

This Discussion Paper is issued under the auspices of the Centre's research programmes:

- Monetary Economics and Fluctuations

Any opinions expressed here are those of the author(s) and not those of the Centre for Economic Policy Research. Research disseminated by CEPR may include views on policy, but the Centre itself takes no institutional policy positions.

The Centre for Economic Policy Research was established in 1983 as an educational charity, to promote independent analysis and public discussion of open economies and the relations among them. It is pluralist and non-partisan, bringing economic research to bear on the analysis of medium- and long-run policy questions.

These Discussion Papers often represent preliminary or incomplete work, circulated to encourage discussion and comment. Citation and use of such a paper should take account of its provisional character.

Copyright: Felix Betz, Peter Bofinger, Jonas Dix and Leonie Streit

# IDENTIFYING MONETARY POLICY SHOCKS IN NEWSPAPERS USING GPT

## Abstract

One of the central challenges in identifying the causal effects of monetary policy is the inherent endogeneity of its conduct. This paper introduces a novel identification strategy that leverages LLMs to detect monetary policy shocks from newspaper coverage following European Central Bank (ECB) policy decisions. Based on a dataset of 7,620 articles from eleven major European newspapers, we classify each policy decision as unexpectedly restrictive, unexpectedly expansionary, or as expected. The resulting narrative-based surprise series captures immediate post-announcement perceptions and shows a close alignment with established High Frequency Identification (HFI) measures with notable exceptions during times of financial turmoil. We subsequently analyze the potential influence of the information effect on our series and find that the majority of identified surprises are unlikely to be driven by information effects.

JEL Classification: E52, E58, C88

Keywords: Monetary policy shocks, Natural language processing, Large Language Models

Felix Betz - felix.betz@uni-wuerzburg.de  
*University Würzburg*

Peter Bofinger - bofinger@t-online.de  
*University Würzburg and CEPR*

Jonas Dix - Jonas.Dix@uni-wuerzburg.de  
*University Würzburg and University Bamberg*

Leonie Streit - Leonie.Streit@uni-wuerzburg.de  
*University Würzburg*

## Acknowledgements

The authors gratefully acknowledge financial support from the Banque Centrale du Luxembourg (BCL). The views in this paper are those of the authors and do not necessarily reflect those of the BCL, the Eurosystem, or their respective policymakers. Furthermore, we thank the participants of the 2025 Forum for Macroeconomics and Macroeconomic Policies Conference, the BCL research seminar on "Large Language Models (LLMs) in Monetary Policy Research" in January 2026, and the Bavarian Macro Day 2026 for their helpful comments and suggestions. All remaining errors are our own.

# Identifying Monetary Policy Shocks in Newspapers using GPT\*

Felix Betz<sup>a</sup>, Peter Bofinger<sup>a</sup>, Jonas Dix<sup>a,b</sup>, and Leonie Streit<sup>a</sup>

<sup>a</sup>University of Würzburg, Germany

<sup>b</sup>University of Bamberg, Germany

April 8, 2026

## Abstract

One of the central challenges in identifying the causal effects of monetary policy is the inherent endogeneity of its conduct. This paper introduces a novel identification strategy that leverages LLMs to detect monetary policy shocks from newspaper coverage following European Central Bank (ECB) policy decisions. Based on a dataset of 7,620 articles from eleven major European newspapers, we classify each policy decision as unexpectedly restrictive, unexpectedly expansionary, or as expected. The resulting narrative-based surprise series captures immediate post-announcement perceptions and shows a close alignment with established High Frequency Identification (HFI) measures with notable exceptions during times of financial turmoil. We subsequently analyze the potential influence of the information effect on our series and find that the majority of identified surprises are unlikely to be driven by information effects.

**Keywords:** Monetary Policy Shocks, NLP, LLM

**JEL Codes:** E52, E58, C88

---

\*The authors gratefully acknowledge financial support from the Banque Centrale du Luxembourg (BCL). The views in this paper are those of the authors and do not necessarily reflect those of the BCL, the Eurosystem, or their respective policymakers. Furthermore, we thank the participants of the 2025 Forum for Macroeconomics and Macroeconomic Policies Conference, the BCL research seminar on "Large Language Models (LLMs) in Monetary Policy Research" in January 2026, and the Bavarian Macro Day 2026 for their helpful comments and suggestions. All remaining errors are our own.

# 1 Introduction

To address the problem of endogeneity in studying the effects of monetary policy, the literature has developed a wide range of identification strategies aimed at isolating exogenous monetary policy shocks. Among contemporary approaches, HFI methods have become particularly popular, as they exploit financial market data to identify unexpected policy actions. In practice, however, the resulting shock series can be quite sensitive to the exact identification strategy applied (Brennan et al., 2024). Additionally, a growing body of literature on the information effect and the response to news channel has raised concerns about the extent to which HFI shocks truly reflect exogenous policy actions (Jarociński and Karadi, 2020; Bauer and Swanson, 2023).

This paper proposes a novel strategy to identify monetary policy shocks from newspaper coverage accompanying policy announcements by the ECB. Although this approach shares with the HFI literature a focus on the reactions of observers of monetary policy rather than the policy actions themselves, it is conceptually more closely related to the tradition of narrative analysis pioneered by Romer and Romer (1989). Advances in Natural Language Processing (NLP) allow us to apply such narrative identification at scale to newspaper coverage. Instead of inferring surprises from asset price movements, we recover them from contemporaneous media accounts of whether a policy decision was perceived as unexpectedly restrictive, unexpectedly expansionary, or in line with prior expectations.

Our identification strategy has several merits. By assessing a large set of reputable newspaper articles, the resulting interpretations are likely to reflect widely held perceptions about the respective policy decision. Similar to the HFI literature, this may help mitigate endogeneity concerns. In addition, despite the rise of social media, traditional media remains the most important channel through which wider audiences receive information about monetary policy (Conrad et al., 2022). While the content of central bank communication and the respective media coverage are related, the latter exerts a much stronger influence on households' perception of key macroeconomic variables (De Fiore et al., 2024). By now, media coverage can be assessed systematically using modern means of automated text analysis, which expands the set of potential material.

When Romer and Romer (1989) paved the way for narrative analysis, they had to personally read tens of thousands of pages. Both the reading itself as well as the documentation and classification can now be conducted more efficiently using advanced NLP methods. In their 2023 update, Romer and Romer express doubts about the feasibility of delegating narrative analyses to computers. However, Hansen and Kazinnik (2024) replicate their study by tasking ChatGPT-4 with an identification prompt detailing the original Romer and Romer

procedure. Similarly, De Fiore et al. (2024) demonstrate that several state-of-the-art LLMs are able to classify hawkish and dovish stances in central bank press conferences at a level comparable to professional human interpreters. These results suggest that LLMs can indeed be used to automate complex analyses, which is essential for our approach. While recent LLM-based applications have mainly focused on classifying central bank communication via sentiment indicators, we use newspaper coverage to identify monetary policy surprises directly. To our knowledge, this is the first study to use LLMs to extract monetary policy shocks directly from media coverage of monetary policy decisions.

In addition to introducing a novel identification strategy, we use our narrative framework to provide new evidence on the empirical relevance of the information effect. Since our dataset consists of contemporaneous newspaper coverage of policy actions, it is particularly well suited to assess whether identified surprises are explicitly linked to the ECB revealing new information about the state of the economy. If such effects are quantitatively important, they should be reflected in the media coverage of policy decisions, as journalists would likely report on the new information revealed by the central bank. However, we find only limited evidence for such a mechanism, suggesting that the quantitative role of the information effect in our setting is modest.

The remainder of the paper is structured as follows: Section 2 reviews the related literature on monetary policy shock identification and narrative analysis. Section 3 describes our identification strategy including the data sources, article selection procedure and LLM classification methodology. We provide a detailed discussion as well as several validation approaches to our identified monetary policy surprises in Section 4. Section 5 assesses the relevance of the information effect in our procedure. Section 6 concludes.

## 2 Related Literature

This paper relates to two main strands of literature. First, it contributes to the literature on identification of monetary policy shocks, where the endogeneity of policy decisions poses a central challenge for empirical macroeconomics. Second, it contributes to the growing literature that applies modern NLP techniques to extract meaningful information from qualitative text sources.

### 2.1 Identification Literature

A large literature has developed different strategies to address the endogeneity of monetary policy (Nakamura and Steinsson, 2018b). Broadly speaking, three approaches have been

particularly influential: *Structural Vector Autoregressions (SVARs)*, *HFI procedures*, and *narrative approaches* (Ramey, 2016).

Historically, SVAR models were the workhorse framework for identifying monetary policy shocks (Kilian and Lütkepohl, 2017). In this framework, shocks are identified from the residual variation that remains after accounting for systematic co-movements implied by the estimated model. Early implementations relied on zero restrictions, imposed either in the short run through recursive Cholesky orderings (Sims, 1980) or in the long run via identifying assumptions on permanent effects (Blanchard and Quah, 1989). Over time, concerns about the plausibility of these assumptions led to more agnostic identification schemes. For instance, sign restrictions may only limit the direction of admissible impulse responses rather than their exact timing or magnitude (Uhlig, 2005), while time-varying parameter Vector Autoregression (VAR)s accommodate the idea that the underlying structure itself may evolve over time (Stock and Watson, 2018). Nevertheless, about half of the early VAR studies yielded a price puzzle, where inflation rose following a monetary tightening (Rusnak et al., 2013). These limitations have motivated a shift toward approaches that identify shocks from external sources while retaining VARs or Local Projections as the estimation tool.

Most prominently, the HFI literature exploits asset-price changes in narrow windows around monetary policy announcements (Gürkaynak, 2005). This approach builds on the assumption that within sufficiently tight time intervals, price movements in highly liquid assets primarily reflect unexpected monetary policy actions rather than other macroeconomic forces (Gertler and Karadi, 2015).

A recurring critique regarding this identification approach is that in many cases, the dominant effect of a central bank policy announcement on financial markets might not be a monetary policy shock, but rather the revelation of new information about the state of the economy (Romer and Romer, 2000). Consistent with this concern, Jarociński and Karadi (2020) show that a large share of ECB policy announcements is accompanied by a theoretically inconsistent positive co-movement of interest rates and stock prices. However, Bauer and Swanson (2023) argue that such patterns may instead reflect a response-to-news channel, in which markets revise their understanding of how the central bank reacts to publicly available information.

Another concern is that different HFI series diverge substantially. Brennan et al. (2024) show that leading U.S. shock measures can correlate as little as 0.5 and coincide in sign only two-thirds of the time. The deviations are mostly driven by the selection of the underlying asset maturities and consequently most pronounced at the effective lower bound, where short-term rates are not informative anymore. Using longer maturities can mitigate this but risks underestimating conventional policy actions (An et al., 2025). So, while HFI offers a

more credible approach, so far, an unquestionably optimal strategy has not yet emerged.

The tradition of narrative analysis can be traced back to Friedman and Schwartz (1963), who identified major monetary policy shocks in the U.S. interwar period through historical interpretation. Romer and Romer (1989) later formalized this approach by systematically assessing the Federal Reserve’s internal meeting records and proposing a more explicit definition of what constitutes a monetary policy shock. Subsequent work has applied similar logic in other historical settings. For example, Petracchi (2022) study historical monographs to identify dates of monetary policy regime shifts, and Monnet (2014) interprets archival policy notes during the Bretton Woods era. The latter example highlights an important advantage of narrative methods: they remain informative even when no single policy instrument exists or when policy frameworks evolve over time.

In a wider sense, any usage of qualitative data can be viewed as narrative analysis, which results in a broad range of different approaches in this category. While these studies still rely on qualitative sources, the human interpretative work is minimized or replaced entirely by algorithmic classifications, which departs from the judgment-driven nature of the original approach. The most frequently adopted "narrative" approach, Romer and Romer (2004), also differs markedly from their 1989 approach. Rather than identifying shocks through qualitative reading, it combines internal Federal Reserve forecasts with a regression framework to isolate the exogenous component of intended policy rate changes. This hybrid strategy is closer to a forward-looking econometric approach than to the manual classification tradition.

## 2.2 Natural Language Processing in Monetary Policy Research

Secondly, our work is rooted in the literature on employing NLP methods in the context of studying monetary policy, which has evolved from simple dictionary-based word counts to more sophisticated models capable of capturing nuanced semantic relationships. Early contributions typically relied on predefined word lists to quantify sentiment or classify qualitative data sources (Baker et al., 2016). For instance, Picault et al. (2022) construct a media sentiment index for ECB policy by applying lexicon-based methods to news coverage. Such approaches are transparent and straightforward, which contributes to their popularity (Gentzkow et al., 2019). However, even basic challenges such as irony, negations, and genre-specific jargon pose serious difficulties (Loughran and McDonald, 2011).

Advances beyond dictionary-based methods have centered on the use of embeddings, where each token is represented as a multi-dimensional vector capturing semantic relationships. Even static embeddings such as Word2Vec or GloVe already outperform dictionaries

by encoding hundreds of latent dimensions rather than a handful of sentiment categories. Studying ECB communication, Baumgärtner and Zahner (2025) report that embeddings can explain up to 70% of policy variation compared to 20-30% for dictionaries. Similar gains arise in forecasting applications, where even simple Word2Vec embeddings of policy statements predict core inflation more accurately than dictionary-based approaches (Araujo et al., 2025).

A further major step was the introduction of transformer architecture, whose contextual embeddings adapt word meaning to the surrounding sentence context. Again, domain-specific finetuning has proved effective in extracting hawkish or dovish sentiment from central bank communication, with Nițoi et al. (2023) showing large accuracy gains over traditional methods. More recently, decoder-based LLMs such as GPT or LLaMA have extended these capabilities to generative and multi-step reasoning tasks and content synthesis. For example, Hansen and Kazinnik (2024) show that GPT models can replicate much of the Romer and Romer (1989) identification procedure by interpreting qualitative records and applying economic judgment in a structured manner. While LLMs can match or exceed the classification accuracy of encoder-based transformers, they introduce additional concerns about reproducibility, cost-efficiency, and transparency. In practice, smaller encoder-based models often remain competitive for narrow classification tasks, whereas LLMs are better suited for tasks requiring richer reasoning or interpretation (Gambacorta et al., 2024).

Our work contributes to both strands of the literature by introducing a novel identification strategy that leverages state-of-the-art LLMs to recover monetary policy shocks from newspaper coverage. While recent advances in NLP have inspired numerous applications, most focus on deriving sentiment-type indicators from central bank communication. For instance, Geiger et al. (2025) finetune an LLM to interpret ECB statements, but the resulting output is aggregated into a sentiment index rather than an exogenous shock measure. A related exception is Ter Ellen et al. (2022), who study media coverage in the context of monetary policy narratives, but do not use LLMs to extract surprise classifications directly from contemporaneous newspaper reporting. The closest LLM-based precedent is Hansen and Kazinnik (2024), who use LLMs to identify shocks from internal Federal Reserve records. By applying a more appropriate shock definition to media sources, our approach extends this line of research to a broader and more market-relevant information channel, constituting, to our knowledge, the first systematic attempt of this kind.

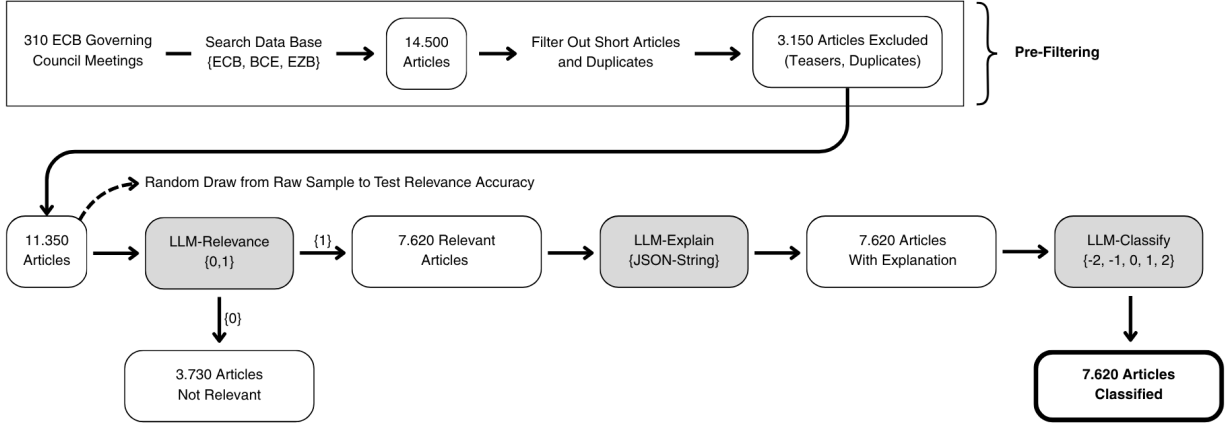


Figure 1: Overview of the identification procedure

### 3 Identification Procedure

While LLMs such as ChatGPT, Claude, or DeepSeek have only recently entered the public spotlight, their ability to process natural language at scale offers promising opportunities for qualitative economic research. Even though the use of language models in the social sciences faces recognizable hurdles, the prospects for new insights and their use for scientific inference are promising Argyle et al. (2025). In particular, transformer-based architectures have demonstrated strong performance in capturing linguistic nuance, even across multiple languages, and are therefore well suited for interpreting qualitative policy communication (Xu et al., 2024). Their usage in the field of economics is thus also highly promising as shown by Ash and Hansen (2023). We put forward a new methodology leveraging a LLM for which we employ a dataset spanning 310 ECB Governing Council (GC) meetings between January 1999 and March 2025, covering 11 major newspapers across 7 countries, and yielding 7,620 relevant articles after filtering. Figure 1 displays the text-processing pipeline used for our procedure which will be explained in more detail in the following sections.

#### 3.1 Selection of Relevant Articles

We select 11 major daily newspapers from core euro area countries as well as Switzerland and the United Kingdom. This selection was guided by (i) national representativeness, (ii) availability, and (iii) relevance for financial and economic reporting. The resulting sample includes both general-interest and business-oriented outlets and thus captures how ECB policy decisions are interpreted for professional financial or economic readership.

For each GC meeting, we retrieve all articles published by the selected newspaper outlets on the next publication day that contain the abbreviation "ECB" in English or its equivalent

in the respective language ("EZB" in German, "BCE" in French, Italian, and Spanish). This keyword filter is simple, transparent, and highly effective, as substantive ECB coverage almost inevitably contains it. If the immediate day after the GC meeting is not a publication day, we shift the search to the next available issue, taking into account weekends and country-specific public holidays. We then apply the following preprocessing steps:

- **Deduplication:** Articles are compared via cosine similarity to discard multiple versions of the same article, stemming from different local editions.
- **Length filter:** To exclude short items that are not considered informative, articles must contain at least 1,000 characters.
- **Relevance filter:** Using an LLM, we exclude articles that mention the ECB but do not contain substantive information about the most recent policy decision.

This last step is particularly important, as the relevance filter accounts for the largest share of discarded observations. We therefore manually validate its performance to assess the reliability of the LLM-based screening.<sup>1</sup> An article is classified as relevant if it contains a discussion of the monetary policy decision taken at the most recent GC meeting. Although the full sample of 14,500 articles is already restricted to articles that mention the ECB on the first publication day after a GC meeting, only about half of them are retained for the core classification steps. As all considered outlets usually have multiple articles mentioning the ECB on any given date, the relevance filter is crucial in ensuring that only articles that explicitly discuss the respective policy decision are considered. Articles that merely mention the ECB without referring to the latest monetary policy decision are excluded. Such discarded articles usually provide a more general assessment of the ECB or other non-monetary-policy-related coverage of the central bank. This process yields the raw count of retrieved articles and the relevant count after filtering. Across the entire sample, we identify 7,620 relevant articles, averaging about 24 articles per GC meeting (see Table 1).

## 3.2 Classification of Articles

Our goal is to measure the shock/surprise component of monetary policy as perceived by the policy's recipients. We therefore classify each relevant article as either **very surprisingly expansive** (−2), **surprisingly expansive** (−1), **neutral/expected** (0), **surprisingly restrictive** (+1), or **very surprisingly restrictive** (+2). This five-point scale offers a

---

<sup>1</sup>See Appendix A for the employed prompt. Results of the respective robustness tests can be found in section 4.1

Table 1: Sample Coverage of Newspapers in the Dataset.

| Newspaper            | Coverage | Country        | $N_{\text{fil}}$ | $N_{\text{rel}}$ |
|----------------------|----------|----------------|------------------|------------------|
| Corriere della Sera  | 2009–    | Italy          | 520              | 329              |
| Der Standard         | 2008–    | Austria        | 278              | 205              |
| El Mundo             | 2002–    | Spain          | 541              | 406              |
| El País              | Full     | Spain          | 833              | 647              |
| FAZ                  | Full     | Germany        | 2,119            | 938              |
| Financial Times      | Full     | United Kingdom | 1,575            | 1,252            |
| Handelsblatt         | Full     | Germany        | 1,804            | 1,161            |
| La Stampa            | –2019    | Italy          | 569              | 395              |
| Le Figaro            | Full     | France         | 658              | 476              |
| Les Echos            | Full     | France         | 1,625            | 1,160            |
| Neue Zürcher Zeitung | 2000–    | Switzerland    | 828              | 647              |
| <b>Total</b>         |          |                | 11,350           | 7,620            |

compromise between the granularity needed to capture meaningful variation and the interpretability required for consistent annotation. It allows for two degrees of magnitude in each direction, providing more nuance than the binary measure in Romer and Romer (1989), while avoiding additional subjectivity concerns that would arise from a larger set of categories.

We implement a two-step prompting procedure for classification, similar in spirit to Geiger et al. (2025). However, in contrast to their objective, we do not aim to create a sentiment (tonality) indicator; instead, we directly target the article’s evaluation of a potential policy shock/surprise. We thus stick to the guidelines of Argyle et al. (2025) by directly acknowledging the ability of the LLM to conduct the narrative inference from the textual data and later provide evidence whether the LLM has succeeded with its given task (see Section 4).

In the first step (see *first-step prompt* in Appendix A), the model independently receives the article and produces a one-sentence English summary of the stated surprise and its direction (restrictive / expansive / neutral). In the second step (see *second-step prompt* in Appendix A), a separate prompt converts this summary into the corresponding numerical category from  $-2$  to  $+2$ . Table 2 illustrates exemplary classifications for selected Frankfurter Allgemeine Zeitung (FAZ) articles along with additional context. The examples provided clearly show that our narrative identification approach goes far beyond the inclusion of conventional monetary policy measures. Monetary policy instruments such as bond purchase programmes or forward guidance are regularly covered in newspaper reports and are explicitly included in our indicator through narrative analysis. We therefore argue that our approach explicitly addresses the multidimensionality of monetary policy communication as discussed

in Altavilla et al. (2019).

Table 2: Exemplary FAZ Classifications and Context

| Date       | One-sentence English summary                                                                                                                                            | Score | Additional context                                                                                                  |
|------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------|---------------------------------------------------------------------------------------------------------------------|
| 2013-11-07 | The article describes the ECB’s rate cut as very unexpectedly expansive, noting that only a tiny minority of analysts had anticipated such an easing at this time.      | −2    | MRO cut to 0.25% (record low at the time).                                                                          |
| 2008-07-03 | The article says markets had expected a series of rate hikes, but Trichet signaled no such path, so the ECB’s decision was more expansive than expected.                | −1    | MRO hike anticipated; Preference for stable interest rate path dampened expectations of further interest rate hikes |
| 2009-03-05 | The article states that the ECB cut its main rate to 1.5 percent ‘as expected’ and does not mention any surprise, so the decision is described as expected/neutral.     | 0     | Rate cut in line with expectations.                                                                                 |
| 2018-06-14 | The article says markets had expected a reduction in ECB bond purchases but not a firm end date, so the decision to terminate the program was unexpectedly restrictive. | +1    | APP reduction schedule with fixed end date.                                                                         |
| 2022-07-21 | The article states that the ECB’s rate hike was twice as large as previously signaled and generally expected, making the decision unexpectedly restrictive.             | +2    | MRO up 50 bps beyond expectations.                                                                                  |

For each GC meeting, we calculate the monetary policy surprise index as the mean classification score across all relevant articles referring to that meeting:

$$MP_t^{surprise} = \frac{1}{N_t} \sum_{i=1}^{N_t} A_{i,t}, \quad (1)$$

where  $MP_t^{surprise}$  is the **Monetary Policy Surprise** at time  $t$ ,  $N_t$  is the number of articles referencing the respective policy decision and  $A_{i,t}$  reflects the classification of an individual article  $i$  at time  $t$ .

The prompts are identical for all articles —regardless of language— leveraging the multi-lingual capabilities of LLMs. This standardization avoids intermediate translation and helps ensure that classification is consistent across countries.

We employ multiple LLM versions which are publicly available from OpenAI, namely `gpt-4.1temp = 0`, `gpt-4.1temp = 1`<sup>2</sup>, `gpt-5`, `gpt-5-nano`, `gpt-5.1noreas`, `gpt-5.1medreas` and `gpt-5.1highreas`<sup>3</sup>. With the above-mentioned selection we include both non-reasoning models (`gpt-4.1`) and reasoning models (`gpt-5` with a `medium` and `high` reasoning effort, `gpt-5 nano` and `gpt-5.1` with a dynamic reasoning effort). In addition, the selected LLMs form the back end of the most widely used chatbot, ChatGPT. We access the individual LLM via the Python API of OpenAI. This setup allows us to control model behavior and ensures that classification is not influenced by prior user interactions.<sup>4</sup> Also when approached via the API several control parameters of the model can be selected that help with reproducibility and stability of the outputs. For the upcoming discussion and analysis we test the latest models from OpenAI and employ `gpt-5.1highreas` as our benchmark model because it is one of the best performers with our internal validation (see Section 4.1). Furthermore, when this study was conducted `gpt-5.1` was the flagship model of OpenAI which itself was by far the most widely used platform for LLMs.

## 4 Discussion of Identification

This section discusses the monetary policy surprise measures derived from the classification procedure described above.<sup>5</sup> As there is some inherent ambiguity in narrative identification of monetary policy shocks (Nakamura and Steinsson, 2018b), we provide several validation exercises to assess the reliability of our approach at both the individual article level and the aggregated surprise series level. Initially, we provide an internal validation of the LLM-based surprise measure by examining the classifications on article level as well as aggregated meeting level. Subsequently, the resulting surprise series is assessed externally by comparing

---

<sup>2</sup>The parameter `temperature` can be selected within the Application Programming Interface (API). `temperature` controls the randomness in the response of LLMs which thus creates more traceable results when `temperature = 0`. However newer versions of LLMs (all `gpts` of `gpt-5X`) do not support this setting due to their reasoning capabilities so that `temperature = default = 1`.

<sup>3</sup>In the `gpt-5X` models the `reasoning effort` can be selected. One can think of reasoning as an iterative thinking process of the LLM. The `default` setting for `gpt-5` is `medium` and for `gpt-5.1` `none`. This is why we employ non-reasoning (`noreas`) and reasoning runs (`medreas` and `highreas`) for `gpt-5.1`

<sup>4</sup>When used via applications like ChatGPT, many LLMs learn from the user behavior or streamline their replies based on previous chats.

<sup>5</sup>Within the following sections we carefully use the term **surprise** to describe our measure thereby noting that it might be subject to limitations that restrict its applicability as pure monetary policy shock measure. We analyze the extent to which our measurement is affected by such limitations, especially the information effect, in section 5 and show that these are rather small.

it to established financial market-based measures.

Figure 2 displays our newspaper-based surprise series along with annotations of prominent monetary policy events in the euro area.<sup>6</sup> The figure reveals that many of the large surprises coincide with well-known monetary policy actions or regime shifts. For instance, the announcements of TLTRO II and TLTRO III in 2016 and 2019 correspond to notable negative spikes, reflecting expansionary monetary policy surprises at the introduction of these new lending facilities. Similar to HFI-based approaches, our series captures the introduction of PEPP in March 2020, as unconventional measures had been anticipated and the actual announcement fell short of expectations, while the expansion of PEPP in June came as a surprise, doubling the volume of asset purchases. Other notable events usually reflect stronger or earlier policy action than anticipated, or inaction when strong actions were expected.

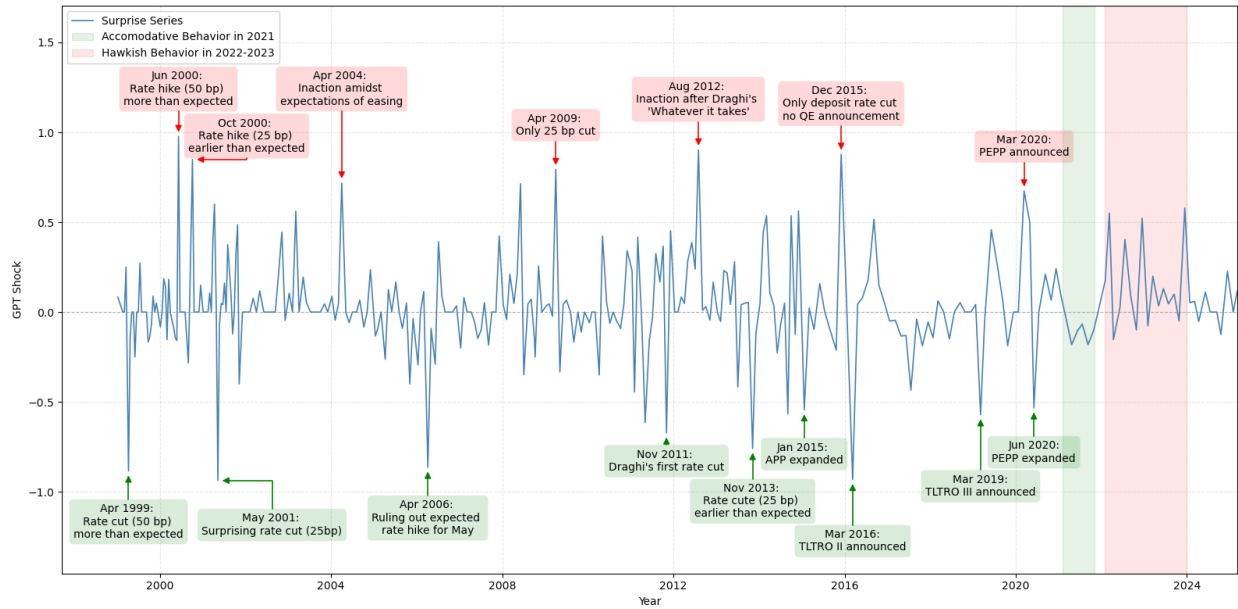


Figure 2: Surprise series with annotations of prominent monetary policy events (Own creation)

## 4.1 Internal Validation

The core of our identification strategy relies on the capability of LLMs to accurately classify articles according to the defined shock categories. To evaluate the LLM-based classification,

<sup>6</sup>The distribution is strongly centered around 0, indicating that most articles classify the monetary policy decision as neutral or expected. This aligns with the ECB's general communication strategy, which aims to communicate policy intentions in a transparent way. Extreme values are rare which is consistent with the characteristics of well-identified monetary policy shocks (Nakamura and Steinsson, 2018b).

we manually conduct a narrative analysis on a randomly drawn subset of articles. The manual coding follows the same conceptual procedure as in the LLM workflow but is carried out independently by research assistants who manually classify 5% of the preprocessed sample ( $n = 568$ ) into six categories  $\{-2, -1, 0, +1, +2, \textit{irrelevant}\}$ , without access to model predictions. This design allows us to jointly assess both steps of the pipeline: relevance selection (Section 3.1) and ordinal shock classification among relevant articles (Section 3.2). Unless otherwise noted, the following results refer to results stemming from `gpt-5.1highreas`, which we use as benchmark model as it was OpenAI’s most advanced frontier model at the time this study was conducted.

Table 3 examines the baseline comparison on newspaper level. Three patterns emerge. First, relevance accuracy is high across most newspapers, typically around 90%, with the lowest value for *Der Standard* (80.0%) and the highest for the *Financial Times* (100%). Second, conditional on relevance, classification accuracy is also high and exceeds 90% for several outlets. Third, disagreements are overwhelmingly small: most classifications are exact matches, and within the remaining cases, the model’s classification is typically only one step away from the manual coding, which indicates that misclassifications typically concern intensity, which can be subject to more ambiguity, rather than the direction of the surprise.

Table 3: Comparison of `gpt-5.1highreas` Classifications with Manual Classifications

| Newspaper           | Accuracy     | Accuracy       | Deviations $ \Delta $ |           |          |
|---------------------|--------------|----------------|-----------------------|-----------|----------|
|                     | Relevance    | Classification | 0                     | 1         | 2        |
| Corriere della Sera | 92.3%        | 93.3%          | 14                    | 1         | 0        |
| Der Standard        | 80.0%        | 100.0%         | 12                    | 0         | 0        |
| El Mundo            | 96.4%        | 81.8%          | 18                    | 4         | 0        |
| El País             | 94.8%        | 95.5%          | 42                    | 2         | 0        |
| FAZ                 | 92.7%        | 84.2%          | 32                    | 5         | 1        |
| Financial Times     | 100.0%       | 91.7%          | 44                    | 4         | 0        |
| Handelsblatt        | 95.3%        | 92.1%          | 58                    | 4         | 1        |
| La Stampa           | 93.8%        | 86.4%          | 19                    | 3         | 0        |
| Le Figaro           | 97.7%        | 80.0%          | 24                    | 6         | 0        |
| Les Echos           | 93.7%        | 94.2%          | 49                    | 3         | 0        |
| NZZ                 | 90.2%        | 85.2%          | 23                    | 4         | 0        |
| <b>Total</b>        | <b>94.4%</b> | <b>89.8%</b>   | <b>335</b>            | <b>36</b> | <b>2</b> |

*Notes:* Relevance accuracy is computed on all manually coded articles. Classification accuracy and absolute deviations are computed on the subset classified as relevant by both manual coders and the model.

Table 4 provides an overview of the performance of all tested LLMs in both the relevance selection and the ordinal classification task. Concerning the agreement rate on whether an article is relevant or not, all models achieve accuracy levels between 91.7% and 95.7%, indicating a consistently high alignment with the manual benchmark across different model versions. Given the size of the benchmark sample, these differences are rather small and do not indicate a clear superiority of one model over the others in terms of relevance selection. However, the classification accuracy among relevant articles varies more substantially across models, ranging from 78.5% to 90.5%. The 90.5% accuracy of the best-performing model `gpt-5` means that the model has classified both the direction and the magnitude of the monetary policy surprise reflected in an article identically to a human researcher in more than 9 out of 10 cases.

Table 4: Model Performance Comparison

| Model                                   | Accuracy     | Accuracy       | Deviations $ \Delta $ |           |          | Relevance    |
|-----------------------------------------|--------------|----------------|-----------------------|-----------|----------|--------------|
|                                         | Relevance    | Classification | 0                     | 1         | 2        | in %         |
| <code>gpt-5.1<sub>highreas</sub></code> | 94.4%        | 89.8%          | 335                   | 36        | 2        | 66.8%        |
| <code>gpt-5.1<sub>medreas</sub></code>  | 95.3%        | 88.2%          | 329                   | 43        | 1        | 66.0%        |
| <code>gpt-5.1<sub>noreas</sub></code>   | 91.7%        | 83.9%          | 287                   | 51        | 4        | 60.3%        |
| <code>gpt-5</code>                      | <b>95.7%</b> | <b>90.5%</b>   | <b>332</b>            | <b>33</b> | <b>2</b> | <b>64.1%</b> |
| <code>gpt-5 nano</code>                 | 92.7%        | 85.6%          | 297                   | 48        | 2        | 62.4%        |
| <code>gpt-4.1<sub>temp = 1</sub></code> | 93.9%        | 78.5%          | 284                   | 73        | 5        | 65.1%        |
| <code>gpt-4.1<sub>temp = 0</sub></code> | 94.8%        | 80.2%          | 292                   | 67        | 5        | 65.0%        |

As expected, across all models the relevance accuracy is higher than the classification accuracy. The relevance selection task is conceptually simpler, as it involves a binary decision about whether an article contains substantive information on the policy decision. In contrast, the ordinal classification task requires interpreting both direction and intensity of the reported surprise, which naturally allows for more ambiguity. Even within the latter task, the agreement rates remain high and almost all deviations are within  $\pm 1$  category in our sample. The mean absolute deviation ranges between 0.10 and 0.12 for frontier models and remains below 0.23 even for the less advanced models. Additionally, the verbal explanations (*first-step prompt*) that determine the numerical classification (*second-step prompt*) can be inspected to comprehend the differing interpretations. These explanations often highlight subtle nuances in how the articles frame the policy decision rather than indicating outright misclassifications. Thus, the deviations typically reflect interpretative ambiguity and do not indicate systematic misclassification. Especially, within the frontier models, the performance is consistently high, clustering around 90%, suggesting that the resulting shock measure is not driven by a particular model specification but remains stable across modern frontier LLMs. Since the empirical shock series aggregates article-level classifications at the meeting level, such small article-level deviations are further attenuated in the final surprise measure. The final column reports the share of articles classified as relevant by each model. These shares range between 60.3% and 66.8% and remain close to the manually observed relevance

share of 63.5%. Importantly, there is no systematic relationship between relevance share and classification accuracy, indicating that higher performance is not achieved through more restrictive relevance filtering.

It should be noted that the classification procedure of the LLMs requires non-negligible computational resources, which is documented in Table 5. Higher-performing frontier models generally involve higher financial costs and token usage. However, even the less costly models achieve satisfactory results at a fraction of the cost of the benchmark model; for instance, `gpt-5` cost more than 100-fold the amount of `gpt-5 nano` while delivering only a modest increase in classification accuracy. The cheaper `gpt-5 nano` came at the cost of the longest computation time, taking more than 45 hours to classify all articles. Depending on the research question and the available resources, those models might be a viable alternative to our benchmark model `gpt-5.1highreas`.

Table 5: Model Cost Comparison

| Model                                   | Computation<br>(hours) | Million Token usage<br>(Output) | Cost<br>(USD) |
|-----------------------------------------|------------------------|---------------------------------|---------------|
| <code>gpt-5.1<sub>highreas</sub></code> | 34                     | 6.1                             | 89            |
| <code>gpt-5.1<sub>medreas</sub></code>  | 23                     | 2.7                             | 55            |
| <code>gpt-5.1<sub>noreas</sub></code>   | 3                      | 0.7                             | 38            |
| <code>gpt-5</code>                      | 35                     | 7.7                             | 105           |
| <code>gpt-5 nano</code>                 | 45                     | 0.7                             | 1             |
| <code>gpt-4.1<sub>temp=0</sub></code>   | 11                     | 0.5                             | 48            |
| <code>gpt-4.1<sub>temp=1</sub></code>   | 11                     | 0.5                             | 48            |

*Notes:* All models use the same amount of input tokens (22 Million Tokens). The token cost for input and output tokens differs across models.

With this internal validation, we put forward evidence that the LLM-based classification is capable of replicating human judgments within narrative analysis, similar to Hansen and Kazinnik (2024).<sup>7</sup> Overall, the LLM classifications align closely with human interpretations, demonstrating that the delegation of interpretational tasks to LLMs is feasible for our application.

<sup>7</sup>We have documented our narrative analysis results in detail. However, as we are not permitted to make the collected newspaper articles freely available for licensing reasons, we can only provide the summary statistics here.

## 4.2 External Validation

We externally validate our newspaper-based monetary policy surprise series by comparing it to two prominent shock/surprise series from the literature: Badinger and Schiman (2023) base their identification on a narrative approach and use a SVAR with sign and magnitude restrictions on four particular events to implement it while Jarociński and Karadi (2020) provide an HFI series that explicitly accounts for the information effects.

Figure 3 displays the rolling correlations between our newspaper-based surprise series and the two benchmark series using a 48-month window on the time period in which all three series are available. Both comparisons yield positive correlations throughout most of the sample period, with a markedly better fit with Jarociński and Karadi (2020) than with Badinger and Schiman (2023). While the correlation with Badinger and Schiman (2023) fluctuates around 0.2 and occasionally turns negative, the correlation between our series and Jarociński and Karadi (2020) is substantially stronger and more stable for large parts of the sample. Even though the four defining (restricted) events of Badinger and Schiman (2023) are in line with our assessment, the overall correlation with their resulting shock series is close to zero and sometimes even negative. However, it is striking that the alignment of the HFI Shock Series from Jarociński and Karadi (2020) to the outcome of Badinger and Schiman (2023) is equally bad. Given the higher alignment of our surprise series with Jarociński and Karadi (2020) and the fact that their approach explicitly accounts for information effects, we subsequently conduct a more detailed comparison with the latter series. Figure 3 also includes the rolling correlation between our series and an updated version of the shock series from Jarociński and Karadi (2020)<sup>8</sup>. The overall correlation is markedly higher with the updated series, that takes into account data up to 2025, suggesting that the updated version aligns better with our measure.

Across the common sample period (March 1999 to December 2016), our newspaper-based series exhibits a correlation of approximately 0.60 with the main HFI measure of Jarociński and Karadi (2020). Moreover, both series agree in sign in about 64% of policy meetings, which lies well within the range documented for comparisons across alternative HFI measures (Brennan et al., 2024). As both shock distributions are tightly centered around zero, sign disagreements predominantly occur for small realizations. Consistent with this interpretation, months with differing signs are characterized by substantially smaller absolute shocks on average. Restricting attention to meetings with an absolute shock exceeding 0.5 (1) standard deviation in either series raises sign agreement to 71.7% (82.2%), indicating a high degree

---

<sup>8</sup>On the personal webpage of Marek Jarociński, updated versions of the shock series are available (see [https://github.com/marekjarocinski/jkshocks\\_update\\_ecb](https://github.com/marekjarocinski/jkshocks_update_ecb)). We also provide a version of Figure 4 with the updated shock series in Appendix D.

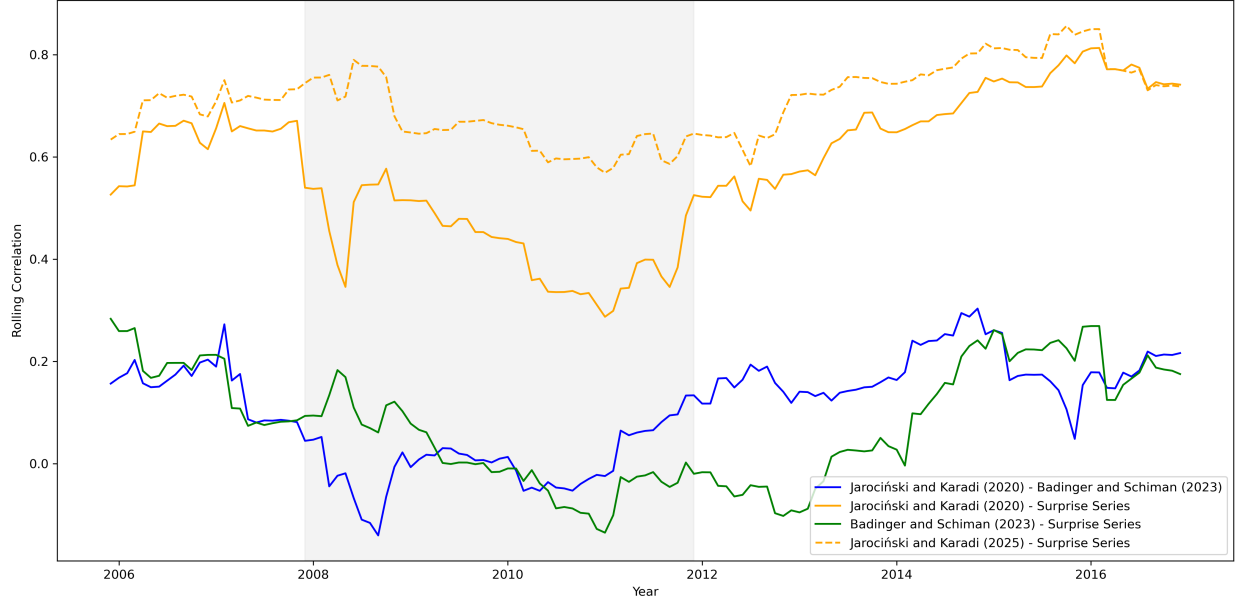


Figure 3: Rolling correlations of shock series from Jarociński and Karadi (2020) and Badinger and Schiman (2023) with our newspaper-based surprise series (Own creation)

of agreement for economically meaningful shocks. Most of the stronger deviations between our shock series and that of Jarociński and Karadi (2020) arise from differences in magnitude rather than direction. In contrast, the lower correlation with Badinger and Schiman (2023) reflects disagreements in both magnitude and direction. While different identification strategies are often evaluated based on the impulse responses they produce, examining the underlying shock series suggests that the different approaches measure fundamentally different dimensions of monetary policy surprises.

In spite of the broad alignment of our newspaper-based surprise series with that of Jarociński and Karadi (2020), there is a notable decline in the rolling correlation in the years following the global financial crisis. Several meetings exhibit clear discrepancies in sign, which account disproportionately for the lower correlation observed during this period. Figure 4 depicts the normalized deviations between our newspaper-based surprise series and that of Jarociński and Karadi (2020) during the period from December 2007 to December 2011. Importantly, many months with opposing signs are characterized by small absolute shock realizations in both series, making it more likely that noise rather than substantive disagreement drives these differences. In the subsequent analysis, we examine four episodes with the largest discrepancies in sign and magnitude in more detail to better understand the underlying sources of disagreement between the two identification approaches. To provide additional institutional context, Appendix C reports the respective statements by the ECB President in the Q&A sessions following the corresponding GC meetings.

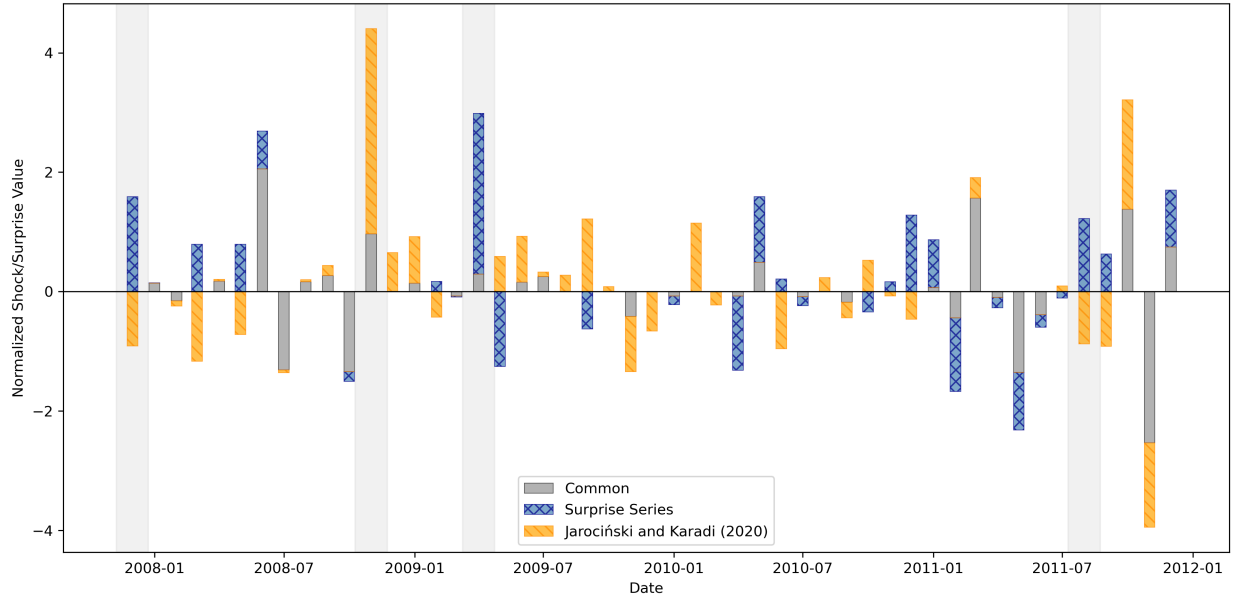


Figure 4: Shock series from Jarociński and Karadi (2020) and from us composed in agreement and disagreement contributions (Own creation)

**December 2007:** In December 2007, our newspaper-based shock series differs in both sign and magnitude from the high-frequency identification of Jarociński and Karadi (2020). While their approach yields a mildly negative (expansionary) monetary policy shock, we identify a strongly restrictive surprise of almost two standard deviations. At its GC meeting on 6 December 2007, the ECB left all key interest rates unchanged. In the accompanying communication, however, the ECB emphasized pronounced upside risks to price stability, noting that Harmonized Index of Consumer Prices (HICP) inflation had reached 3.0% in November and that money and credit growth remained "very vigorous". The GC explicitly signaled its readiness to counter inflationary pressures "in a firm and timely manner" if required, stressing the importance of anchoring inflation expectations, particularly in an environment of heightened financial market uncertainty (Trichet, 2007). The decision was taken at an early stage of the global financial crisis, characterized by rapidly falling US house prices and increasing financial stress. On the same day, the Bank of England (BoE) responded to the turmoil with its first interest rate cut in years, following the Federal Reserve System (Fed), which had already reduced rates in September and October. Against this backdrop, market participants had increasingly expected a more accommodative response also from the ECB, which was not fulfilled. The combination of unchanged policy rates and a strongly inflation-focused communication therefore constituted a restrictive surprise relative to prevailing expectations.

Our procedure identifies 26 relevant articles associated with this monetary policy de-

cision. Eleven articles classify the decision non-zero, all of them in unanimously characterizing it as unexpectedly restrictive. The dominant narrative emphasizes that, despite rising financial stress and easing abroad, the ECB not only refrained from cutting rates but also communicated a willingness to tighten policy should inflationary pressures persist. This narrative-based classification differs from the expansionary shock implied by the high-frequency measure of Jarociński and Karadi (2020) and highlights a divergence in the interpretation of the policy decision at this early stage of the financial crisis. It is also noteworthy that in the updated version of the shock series by Jarociński and Karadi (2020) (see Figure 9 in Appendix D) the directional disagreement regarding this month is resolved.<sup>9</sup>

**November 2008:** In November 2008, the divergence between our newspaper-based shock series and the high-frequency identification of Jarociński and Karadi (2020) is entirely due to differences in magnitude, as both approaches agree on the sign of the shock. While Jarociński and Karadi (2020) identify their strongest positive (contractionary) monetary policy shock, exceeding four standard deviations, our approach yields a considerably smaller positive surprise of around one and a half standard deviations. At its GC meeting on 6 November 2008, the ECB decided to cut all key interest rates by 50 basis points, thereby continuing the easing cycle initiated in October. In its communication, the ECB argued that the outlook for price stability had improved further, while ongoing financial market turmoil was expected to dampen global and euro area demand for a protracted period of time (Trichet, 2008). The policy decision thus largely constituted a continuation of an already anticipated easing path amid extreme financial market stress. For instance, equity market uncertainty surged, as evidenced by exceptionally high implied volatility in European stock markets. Also, broader measures of systemic risk, such as the Composite Indicator of Systemic Stress (CISS), reached their highest levels in that month, reflecting widespread tensions across multiple financial market segments (Garcia-de-Andoain and Kremer, 2018). Thus, this environment was characterized by heightened risk aversion, severe liquidity constraints, and large-scale repricing across asset classes.

Our procedure identifies 25 relevant newspaper articles associated with this monetary policy decision. Of these, 11 classify the decision as unexpectedly restrictive, while the remaining articles interpret it as broadly in line with expectations. The dominant narrative emphasizes that, despite the substantial rate cut, financial markets had anticipated even more decisive action given the severity of the crisis. Several articles stress that the magnitude of the easing was perceived as insufficient relative to prevailing expectations and the scale

---

<sup>9</sup>According to our findings from section 5 December 2007 marks one of the strongest information effect driven surprises in our sample. After purging the information effect from our surprise series the two shock series align almost perfectly while still there is no indication for an expansionary shock.

of financial disruption. This reading is consistent with the assessment of Soenen and Vander Vennet (2021), who argue that early monetary policy responses during the financial crisis were initially perceived as not sufficiently expansionary, and that the ECB’s policy stance only reverted to a clearly expansionary interpretation once policy actions were stepped up more forcefully. This narrative assessment is consistent with the moderate positive surprise identified in our procedure, but contrasts sharply with the extremely large contractionary shock implied by the high-frequency measure of Jarociński and Karadi (2020). Importantly, this divergence does not imply that the strong high-frequency price responses observed in November 2008 were unrelated to monetary policy. Rather, in an environment of acute systemic stress and exceptionally elevated market volatility, announcement-time asset price movements are likely to reflect monetary policy news alongside broader market-wide adjustments that are not specific to the policy decision itself. As a result, the mapping from high-frequency asset price changes to the underlying monetary policy shock may be overstating the degree of contractionary surprise during crisis periods.

**April 2009:** In April 2009, the two shock series agree in sign but differ markedly in magnitude. While Jarociński and Karadi (2020) identify a mild positive (contractionary) shock of around 0.3 standard deviations, our newspaper-based series identifies a large positive surprise of around three standard deviations. Unlike the previous two episodes, this divergence may be partially related to information effects, as Jarociński and Karadi (2020) view this policy decision as being primarily driven by information effects. Their estimated information shock amounts to roughly 1.3 standard deviations and can thus explain part of the deviation. However, even this non-negligible information component does not fully reconcile the difference in magnitudes between the two identification approaches. At its GC meeting on 2 April 2009, the ECB reduced all key interest rates by 25 basis points. Importantly, this move was widely perceived as smaller than expected, which explains the restrictive interpretation in both shock series. Additionally, the GC highlighted the already historically low level of interest rates, signaled caution about moving the deposit rate further and dampened hopes for unconventional measures (Trichet, 2009). Alongside the rate decision, the communication plausibly contributed to a reading that the ECB was approaching the effective lower bound only gradually and was not yet willing to match the perceived aggressiveness of other central banks. The decision came against the backdrop of a deepening global recession, with euro area GDP contracting sharply and unemployment rising. Financial markets had anticipated a more aggressive easing, given the severity of the economic downturn.

Our procedure yields 29 relevant articles, of which 23 classify the decision as unexpectedly restrictive. This unusual degree of consensus on the restrictive nature of the decision

produces the strongest shock in our series in the designated period. The dominant narrative emphasizes that the rate cut was perceived as insufficiently aggressive relative to prevailing expectations and that the communication signaled a cautious approach to further rate reductions or unconventional measures. This narrative-based interpretation of a strong contractionary shock is neither matched by the mildly restrictive pure monetary policy shock nor the information shock identified by Jarociński and Karadi (2020). Again, as we already mentioned for December 2007, the updated version of the shock series by Jarociński and Karadi (2020) (see Figure 9 in Appendix D) dampens the magnitude of disagreement for this month, where their pure monetary policy shock is now close to 1.5 standard deviations, which is more in line with our newspaper-based surprise series.

**August 2011:** In August 2011, the two shock series differ not only in magnitude but also in sign. While Jarociński and Karadi (2020) view this policy decision mainly as an information effect, finding a negative information effect of roughly 3 standard deviations (one of their overall strongest) alongside a still sizable expansionary pure monetary policy shock ( $-0.9$  standard deviations), our newspaper-based surprise series identifies a positive (contractionary) surprise of around 1.3 standard deviations.<sup>10</sup> At its GC meeting on 4 August 2011, the ECB kept key interest rates unchanged, following the 25 basis point increase in July. At the same time, it announced a supplementary six-month LTRO and continued fixed-rate full allotment procedures. While these measures can be interpreted as liquidity-supportive, the overall communication maintained a firm focus on upside risks to price stability. The introductory statement emphasized that the previous rate hike had been warranted, that inflation remained elevated, and that vigilance against second-round effects was essential (Trichet, 2011). Thus, despite renewed financial market tensions, the ECB did not signal a decisive shift toward a more accommodative stance.

Our procedure identifies 40 relevant articles, of which 15 classify the decision as unexpected. Among these, only one article describes it as unexpectedly expansionary, while the other fourteen characterize it as unexpectedly contractionary. The majority of the articles that indicated a surprise reported that the measures, but especially the communication regarding future rate paths and bond purchase programmes were seen as disappointing. In particular, the decision to limit the bond purchase programme to buying Portuguese and Irish bonds, thereby excluding Spanish and Italian bonds, was widely perceived as insufficient given the severity of the sovereign debt crisis and the associated financial market turmoil. The large negative-signed surprise implied by the HFI-scheme on this occasion is

---

<sup>10</sup>It is noteworthy that this event marks the only one among our four designated events in which the disagreement increases when taking into account the updated Jarociński and Karadi series (see Figure 9 in Appendix D).

not corroborated by our narrative assessment of the newspaper coverage.

**General Patterns:** The results of Jarociński and Karadi (2020) allow us to further investigate a potential relationship between our newspaper-based surprise series and information effects, which could potentially explain some of the differences observed above. As they report both the information shock and the pure monetary policy shock separately, we can simply regress our newspaper-based series on both components. While the coefficient on the pure monetary policy shock is positive and highly significant ( $0.63, p < 0.01$ ), the information shock component does not attain statistical significance ( $-0.03, p = 0.61$ ). This exercise suggests that our newspaper-based series predominantly captures the exogenous monetary policy action rather than information effects, supporting its validity as a monetary policy shock measure. We discuss our take on the role of information effects in more detail in section 5.

Beyond the event-specific discrepancies discussed above, several systematic differences emerge. Most notably, their series features more extreme realizations of monetary policy shocks. While our shock series only exceeds 3 standard deviations once, their series does so 7 times. These extreme values likely reflect large market reactions to unexpected policy actions. In all of these extreme cases, the two series do agree in sign, indicating that both measures broadly capture the same underlying phenomenon despite differing magnitudes. Restricting attention to months with opposing signs and normalized absolute deviations exceeding 0.5 standard deviations, we find that these disagreements are concentrated in very specific years. First, the early years of the ECB (1999-2003) exhibit more frequent disagreements, which may reflect the market’s initial uncertainty about the new institution and lower transparency on behalf of the central bank. One third of these months falls into the period before 2001, when the ECB was still holding two meetings per month. The other years with more than one pronounced disagreement are 2008, 2009 and 2012, which were characterized by heightened uncertainty during the global financial crisis and the European sovereign debt crisis.

The particular timing of large deviations raises the question of whether the observed divergence reflects fundamental differences in shock identification or simply amplified market responses during turbulent periods. To investigate the divergence more formally, we conduct several regression analyses reported in Table 6. Financial stress emerges as the dominant predictor of divergence between the two series. The CISS coefficient is highly significant both against the original sample of Jarociński and Karadi (2020) and the updated sample, indicating that heightened financial stress raises the absolute discrepancy between the two series. While the overall explanatory power with an  $R^2$  of around 6% is modest, the relationship is

highly robust across different specifications.

Table 6: Explaining Divergence

| Dependent Variable                                    | Quantile Regression |                   |                     |                     |
|-------------------------------------------------------|---------------------|-------------------|---------------------|---------------------|
|                                                       | OLS                 | Q25               | Q50                 | Q75                 |
| <i>Original Jarociński &amp; Karadi [1999 – 2016]</i> |                     |                   |                     |                     |
| Stress (CISS)                                         | 0.848***<br>(0.203) | 0.034<br>(0.162)  | 0.513**<br>(0.212)  | 1.729***<br>(0.387) |
| $N$                                                   |                     |                   | 204                 |                     |
| $R^2$ / Pseudo $R^2$                                  | 0.058               | 0.001             | 0.018               | 0.058               |
| <i>Updated Sample [1999 – 2025]</i>                   |                     |                   |                     |                     |
| Stress (CISS)                                         | 0.783***<br>(0.167) | 0.244*<br>(0.127) | 0.650***<br>(0.171) | 1.154***<br>(0.360) |
| $N$                                                   |                     |                   | 273                 |                     |
| $R^2$ / Pseudo $R^2$                                  | 0.062               | 0.006             | 0.015               | 0.043               |

Notes: \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.10$ . Standard errors in parentheses. OLS with Newey–West (HAC) standard errors using 12 lags.

Subsequent quantile regressions confirm that the relationship is clearly non-linear and becomes more pronounced in the right tail of the deviation distribution. The CISS coefficient increases monotonically from 0.03 (insignificant) at the 25th percentile to 1.73 ( $p < 0.01$ ) at the 75th percentile in the original sample and exhibits a similar pattern in the updated sample. The documented heterogeneity implies that financial stress significantly contributes to the large deviations while leaving the bulk of the distribution largely unaffected. This pattern is consistent with the interpretation that during periods of elevated financial stress, high-frequency asset prices react more sharply to policy events, producing larger market-based shock realizations. This suggests that our newspaper-based shock series provides a more stable measure of monetary policy surprises, particularly during episodes of extreme financial stress where HFI approaches may conflate the monetary policy signal with broader market volatility.

### 4.3 Empirical Application

To illustrate the empirical application of our newspaper-based surprise series, we estimate the dynamic effects of monetary policy surprises on key macroeconomic variables using local projections (Jordà, 2005). Specifically, we estimate the following equation for each variable of interest  $Y$ :

$$Y_{t+h} - Y_t = \alpha_h + \beta_h MP_t^{surprise} + \gamma_h X_t + \varepsilon_{t+h}, \quad h = 0, 1, 2, \dots, H, \quad (2)$$

where  $Y_{t+h}$  is the value of the variable at horizon  $h$ ,  $MP_t^{surprise}$  is our newspaper-based monetary policy surprise at time  $t$ ,  $X_t$  is a vector of control variables, and  $\varepsilon_{t+h}$  is the error term. The coefficients  $\beta_h$  capture the impulse response of  $Y$  to a one-deviation (contractive) surprise in monetary policy at horizon  $h$ . We include control variables to account for other factors that may influence the dynamics of  $Y$ . It is noteworthy that the interpretation of the impulse responses is different to other standard approaches in the literature as our shock measure is not expressed in a policy rate change but bears its own metric. Therefore, the impulse responses reflect the effect of an exogenous contractive monetary policy change in our newspaper-based surprise measure on the variable of interest. A surprise normalized to 1 can be carefully interpreted as the outcome of a monetary policy decision that was classified by all considered articles as "surprisingly restrictive" so that the impulse responses reflect the effect of such a decision on the variable of interest. Interestingly, such a normalized surprise leads to a response of roughly 25 basis points in our policy rate so that the responses can be interpreted as the effect of an unexpected monetary policy decision that leads to a 25 basis point increase in the policy rate.

As it is standard in the literature, we include lags of the dependent variable and other relevant macroeconomic controls in  $X_t$  to account for potential confounding factors.<sup>11</sup> The specification and selection of our variables  $Y$  follows Holm et al. (2021) and the Local Projection application of Cloyne and Hürtgen (2016). In particular, our dependent variables include a policy rate (we use German one-year government bond yields as policy rate variable), measures for economic activity (unemployment and industrial production) as well as a price indicator (log-HICP). For control variables, we include BBB bond spreads for financial conditions and a stock market index. As industrial production indicates structural breaks during the Great Financial Crisis and during COVID, we further include respective dummies to ensure that our results are not driven by those events. Figure 5 displays the estimated impulse responses. Standard results emerge: A contractionary monetary policy surprise increases the policy rate on impact with the strongest result after about six months. Economic

---

<sup>11</sup>We use 12 lags for the dependent and 3 lags for the control variables.

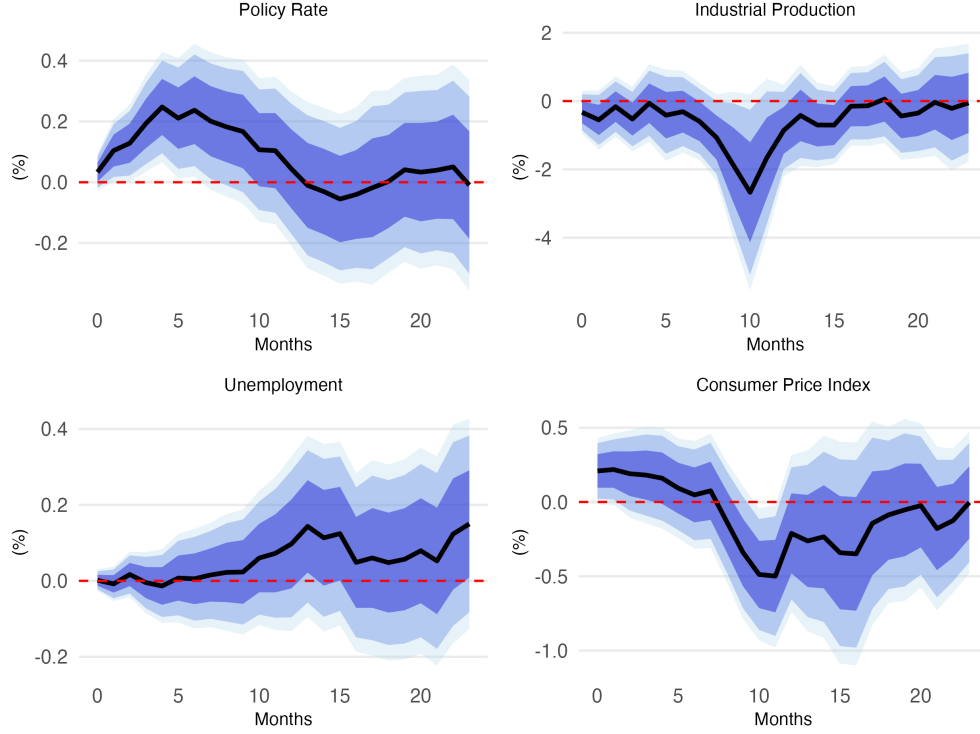


Figure 5: Baseline Local Projection with Newspaper-Based Monetary Policy Shocks

activity reacts delayed with unemployment rising after about ten months while industrial production declines (significant) after about ten months. Prices also respond delayed with a small positive initial reaction before declining significantly with (strongly significant) point estimates of a 50 basis points decrease in HICP after about ten months to one year. These findings are in line with standard monetary policy transmission mechanisms documented in the literature. We take this small empirical application as a first demonstration of the validity and usefulness of our newspaper-based surprise measure, which can be used in a wide range of applications to investigate the effects of monetary policy shocks on various macroeconomic outcomes. However, similar to HFI impulse responses, we observe a small initial price hike, which is often interpreted as a sign of an underlying information effect in the shock/surprise measure. We therefore investigate this issue in more detail in the next section, where we explore the role of information effects in our newspaper-based surprise series.

## 5 The Role of the Information Effect in Newspaper-based Monetary Policy Surprises

Applying a further LLM-based classification step to our identification procedure as presented in Figure 1 allows us to investigate to what extent our newspaper-based monetary policy surprises are affected by the so-called information effect. As described in Romer and Romer (2000) and Nakamura and Steinsson (2018a), a surprise can be polluted by an information effect if the central bank’s policy decision conveys information about the state of the economy that was not previously available to market participants. For instance, an unexpected rate hike may signal that the economy is stronger than previously thought. In such a scenario, the surprise reflects the new information rather than the policy action itself being surprising. Not accounting for this information effect violates the exogeneity assumption of the shock measure and can lead to rather unexpectedly positive reactions in financial markets and the real economy. It is precisely this counter-theoretical reaction of stock markets in response to a policy decision that Jarociński and Karadi (2020) use for the decomposition of their high-frequency shock series into a pure monetary policy shock and an information shock.

As we are able to directly analyze the narrative content of the newspaper articles that signaled the underlying surprise, we should be able to detect indications of such information effects in the articles themselves. If a surprise of a monetary policy announcement is driven by such an information effect, we expect the articles to explicitly mention that the policy decision is surprising because it revealed new information about the state of the economy. The discussion of the information effect in the literature has evolved considerably since the definition proposed by Romer and Romer (2000), which more closely reflects the idea of newly shared *private* information, whereas the definition adopted above from Jarociński and Karadi (2020) better captures the notion that the central bank communicates *new information about its own assessment of the economic outlook*. According to the definition by Miranda-Agrippino and Ricco (2021), an information effect can even arise implicitly through a latent signal in the policy action itself. In our reading, however, this is closely related to the alternative characterization of the ‘Response to News’ effect (Bauer and Swanson, 2023) - that is, the assumption that the surprise in monetary policy action does not stem from a difference in the implied information sets, but rather from a *misperception of how the central bank weights that information set*, which is equally accessible to both the public and the central bank. For the purposes of this simple decomposition exercise, we therefore retain the term ‘information effect’ which, particularly in the case of our narrative data source, cannot be completely distinguished from the ‘Response to News’ effect.

We therefore extend our narrative pipeline by analyzing the articles that signaled a

surprise throughout the main procedure. We apply a further LLM-based classification step to these 1,722 articles, which is based on a similar prompt architecture as within the main procedure, but now specifically asks the LLM to analyze the article and determine whether the surprise was due to the information effect. Within the prompt we further provide the LLM with the following definition, which is based on the description of the information effect in Jarociński and Karadi (2020):

A monetary policy surprise that arises because the central bank reveals new information about its assessment of the economic outlook.

The LLM is then asked to return a binary classification of whether the surprise was due to an information effect or not accompanied by a short explanation of its assessment. The full prompt is documented in Appendix A. We thus follow the "poor-man's proxy" approach of Jarociński and Karadi (2020) in which a newspaper-indicated surprise either reflects a pure monetary policy shock or an information shock. Since we then aggregate the article-level classifications to the meeting level, we get a continuous decomposition of our newspaper-based surprise series into a pure monetary policy shock and an information shock (see Figure 6), which is in the spirit of Figure 9 in Jarociński and Karadi (2020).

We re-run our empirical application from section 4.3 (see Figure 7 in Appendix B) but do not find significant differences when using the 'pure' monetary policy shock series instead of the original surprise measure. Running the same local projection specification as in equation 2 but replacing the original shock measure with the information effect labeled shock series (see Figure 8 in Appendix B) yields largely insignificant results, where only partial components of the price puzzle from Figure 5 can be explained. Since the 'Response to News' channel may compress the role of an information effect (Bauer and Swanson, 2023), our estimate of the presence of an information effect is particularly conservative. Nevertheless, we document rather small contributions of the information effect regarding the magnitude of the underlying surprise. Given our initial sample of 1,722 articles that signaled surprising monetary policy behavior, only 245 articles were labeled as being driven by an information effect. As visible in Figure 6, only a small proportion (14.1%) of the underlying surprises were driven by an information effect. We thus document much smaller information effect contributions compared to Jarociński and Karadi (2020), who find that the information effect accounts for almost 50% of their overall surprises.

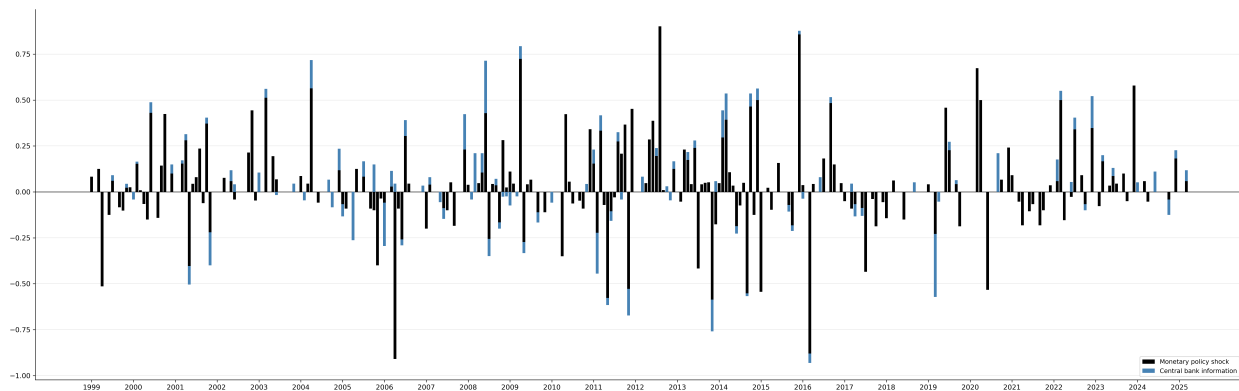


Figure 6: Decomposition of Information Effect in Newspaper-Based Monetary Policy Shocks

## 6 Conclusion

This paper proposes an alternative approach to constructing a comprehensive and historically consistent shock series from contemporaneous newspaper coverage of monetary policy decisions. A central strength of our approach lies in its transparency. Even though HFI monetary policy shocks—the benchmark in monetary economics—are theoretically appealing, they often lack retrospective interpretation and are sensitive to the financial market indicators and methodology used. By focusing on public and media reception rather than solely on financial markets, our approach yields surprisingly similar results while still being fully traceable to the underlying narrative content.

We present evidence that LLMs are capable of replicating human narrative analysis even if the task requires the complexity of interpreting nuanced economic coverage. We leverage these capabilities by building a newspaper-based shock series which aligns particularly well with Jarociński and Karadi (2020). However, the alignment is notably lower during periods of financial turmoil, where we find our newspaper-based series to provide a more plausible assessment. The empirical application of our newspaper-based series aligns with standard findings in the literature. We further investigate the role of information effects in our series and in contrast to Jarociński and Karadi (2020) find that they only account for a small proportion of the surprises.

Several limitations require further consideration. Firstly, newspaper coverage may be subject to editorial bias or country-specific framing. Secondly, while LLM-based classification performed accurately in our validation exercises, it remains sensitive to prompt design. Furthermore, our decomposition exercise was limited to a simple binary classification of the information effect. A more nuanced classification might distinguish between different types of information effects or take into account the response to news channel. Such extensions could provide deeper insights into the underlying drivers of monetary policy shocks.

## References

- Altavilla, C., Brugnolini, L., Gürkaynak, R. S., Motto, R., and Ragusa, G. (2019). Measuring euro area monetary policy. Journal of Monetary Economics, 108:162–179.
- An, P., Dilts Stedman, K., and Lusompa, A. (2025). How High Does High Frequency Need to Be? A Comparison of Daily and Intradaily Monetary Policy Surprises. Research Working Paper no. 25-03, Federal Reserve Bank of Kansas City.
- Araujo, D. K. G., Bokan, N., Comazzi, F. A., and Lenza, M. (2025). Word2prices: Embedding Central Bank Communications for Inflation Prediction. BIS Working Paper Series No 3047, Bank for International Settlements.
- Argyle, L. P., Busby, E. C., Gubler, J. R., Hepner, B., Lyman, A., and Wingate, D. (2025). Arti-‘fickle’ intelligence: Using LLMs as a tool for inference in the political and social sciences. Nature Computational Science, 5(9):737–744.
- Ash, E. and Hansen, S. (2023). Text Algorithms in Economics. Annual Review of Economics, 15:659–688.
- Badinger, H. and Schiman, S. (2023). Measuring Monetary Policy in the Euro Area Using SVARs with Residual Restrictions. American Economic Journal: Macroeconomics, 15(2):279–305.
- Baker, S. R., Bloom, N., and Davis, S. J. (2016). Measuring Economic Policy Uncertainty\*. The Quarterly Journal of Economics, 131(4):1593–1636.
- Bauer, M. D. and Swanson, E. T. (2023). An Alternative Explanation for the “Fed Information Effect”. American Economic Review, 113(3):664–700.
- Baumgärtner, M. and Zahner, J. (2025). Whatever it takes to understand a central banker — Embedding their words using neural networks. Journal of International Economics, 157:1–20.
- Blanchard, O. J. and Quah, D. (1989). The Dynamic Effects of Aggregate Demand and Supply Disturbances. The American Economic Review, 79(4):655–673.
- Brennan, C. M., Jacobson, M. M., Matthes, C., and Walker, T. B. (2024). Monetary Policy Shocks: Data or Methods? Finance and Economics Discussion Series 2024-011, Federal Reserve Board, Washington, DC.

- Cloyne, J. and Hürtgen, P. (2016). The Macroeconomic Effects of Monetary Policy: A New Measure for the United Kingdom. American Economic Journal: Macroeconomics, 8(4):75–102.
- Conrad, C., Enders, Z., and Glas, A. (2022). The role of information and experience for households’ inflation expectations. European Economic Review, 143:1–22.
- De Fiore, F., Maurin, A., Mijakovic, A., and Sandri, D. (2024). Monetary policy in the news: Communication pass-through and inflation expectations. BIS Working Papers No 1231, Bank for International Settlements.
- Friedman, M. and Schwartz, A. J. (1963). A Monetary History of the United States, 1867-1960. Studies in Business Cycles. Princeton University Press, Princeton.
- Gambacorta, L., Kwon, B., Park, T., Patelli, P., and Zhu, S. (2024). CB-LMs: Language models for central banking. BIS Working Paper Series No 1215, Bank for International Settlements.
- Garcia-de-Andoain, C. and Kremer, M. (2018). Beyond spreads: Measuring sovereign market stress in the euro area. Working Paper Series No 2185, European Central Bank, LU.
- Geiger, F., Kanelis, D., Lieberknecht, P., and Sola, D. (2025). Monetary-Intelligent Language Agent (MILA). Technical Paper, Deutsche Bundesbank.
- Gentzkow, M., Kelly, B., and Taddy, M. (2019). Text as Data. Journal of Economic Literature, 57(3):535–574.
- Gertler, M. and Karadi, P. (2015). Monetary Policy Surprises, Credit Costs, and Economic Activity. American Economic Journal: Macroeconomics, 7(1):44–76.
- Gürkaynak, R. S. (2005). Using Federal Funds Futures Contracts for Monetary Policy Analysis. Finance and Economics Discussion Series, Federal Reserve System, Washington, DC.
- Hansen, A. L. and Kazinnik, S. (2024). Can ChatGPT Decipher FedSpeak?
- Holm, M. B., Paul, P., and Tischbirek, A. (2021). The Transmission of Monetary Policy under the Microscope. Journal of Political Economy, 129(10):2861–2904.
- Jarociński, M. and Karadi, P. (2020). Deconstructing Monetary Policy Surprises—The Role of Information Shocks. American Economic Journal: Macroeconomics, 12(2):1–43.

- Jordà, Ò. (2005). Estimation and Inference of Impulse Responses by Local Projections. American Economic Review, 95(1):161–182.
- Kilian, L. and Lütkepohl, H. (2017). Structural Vector Autoregressive Analysis. University Press, Cambridge.
- Loughran, T. and McDonald, B. (2011). When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks. The Journal of Finance, 66(1):35–65.
- Miranda-Agrippino, S. and Ricco, G. (2021). The Transmission of Monetary Policy Shocks. American Economic Journal: Macroeconomics, 13(3):74–107.
- Monnet, E. (2014). Monetary Policy without Interest Rates: Evidence from France’s Golden Age (1948 to 1973) Using a Narrative Approach. American Economic Journal: Macroeconomics, 6(4):137–169.
- Nakamura, E. and Steinsson, J. (2018a). High-Frequency Identification of Monetary Non-Neutrality: The Information Effect\*. The Quarterly Journal of Economics, 133(3):1283–1330.
- Nakamura, E. and Steinsson, J. (2018b). Identification in Macroeconomics. Journal of Economic Perspectives, 32(3):59–86.
- Nițoi, M., Pochea, M.-M., and Radu, Ș.-C. (2023). Unveiling the sentiment behind central bank narratives: A novel deep learning index. Journal of Behavioral and Experimental Finance, 38:1–15.
- Petracchi, C. (2022). The Mussa puzzle: A generalization. European Economic Review, 149:1–17.
- Picault, M., Pinter, J., and Renault, T. (2022). Media sentiment on monetary policy: Determinants and relevance for inflation expectations. Journal of International Money and Finance, 124:1–24.
- Ramey, V. (2016). Macroeconomic Shocks and Their Propagation. In John B. Taylor and Harald Uhlig, editors, Handbook of Macroeconomics, pages 71–162. Elsevier.
- Romer, C. D. and Romer, D. H. (1989). Does Monetary Policy Matter? A New Test in the Spirit of Friedman and Schwartz. NBER Macroeconomics Annual, (4):121–170.
- Romer, C. D. and Romer, D. H. (2000). Federal Reserve Information and the Behavior of Interest Rates. American Economic Review, 90(3):429–457.

- Romer, C. D. and Romer, D. H. (2004). A New Measure of Monetary Shocks: Derivation and Implications. American Economic Review, 94(4):1055–1084.
- Romer, C. D. and Romer, D. H. (2023). Presidential Address: Does Monetary Policy Matter? The Narrative Approach after 35 Years. American Economic Review, 113(6):1395–1423.
- Rusnak, M., Havranek, T., and Horvath, R. (2013). How to Solve the Price Puzzle? A Meta-Analysis. Journal of Money, Credit and Banking, 45(1):37–70.
- Sims, C. A. (1980). Macroeconomics and Reality. Econometrica, 48(1):1–48.
- Soenen, N. and Vander Venet, R. (2021). ECB monetary policy and bank default risk. Journal of International Money and Finance, 122:1–20.
- Stock, J. H. and Watson, M. W. (2018). Identification and Estimation of Dynamic Causal Effects in Macroeconomics Using External Instruments. The Economic Journal, 128(610):917–948.
- Ter Ellen, S., Larsen, V. H., and Thorsrud, L. A. (2022). Narrative Monetary Policy Surprises and the Media. Journal of Money, Credit and Banking, 54(5):1525–1549.
- Trichet, J.-C. (2007). Introductory statement with Q&A. Technical report, European Central Bank, Frankfurt am Main.
- Trichet, J.-C. (2008). Introductory statement with Q&A. Technical report, European Central Bank, Frankfurt am Main.
- Trichet, J.-C. (2009). Introductory statement with Q&A. Technical report, European Central Bank, Frankfurt am Main.
- Trichet, J.-C. (2011). Introductory statement with Q&A. Technical report, European Central Bank, Frankfurt am Main.
- Uhlig, H. (2005). What are the effects of monetary policy on output? Results from an agnostic identification procedure. Journal of Monetary Economics, 52(2):381–419.
- Xu, R., Sun, Y., Ren, M., Guo, S., Pan, R., Lin, H., Sun, L., and Han, X. (2024). AI for social science and social science of AI: A survey. Information Processing & Management, 61(3):103665.

# A LLM Prompts for Article Classification

## Prompt 1: Relevance Detection

You are an expert in European monetary policy and economic journalism.

Task:

Analyze the following newspaper article from an economic perspective. The goal is to assess whether the article explicitly discusses the most recent monetary policy decision by the European Central Bank (ECB). An article that only focuses on a monetary policy decision by another central bank (e.g. Fed, BoE, SNB) is not relevant. Monetary policy decisions refer in particular to interest rate decisions and ECB press conferences, during which the central bank communicates its monetary policy assessments, forecasts, and actions.

An article is considered relevant if it reports on a specific monetary policy decision (e.g., an interest rate hike, cut, or hold or other monetary policy decisions like asset purchase programs or forward guidance).

Articles are not considered relevant if they only mention terms like "ECB" or "interest rates" without referring to a specific and recent ECB decision or its monetary policy communication (e.g., general commentary on monetary policy, retrospective analysis without reference to a specific event, passing mentions).

Return exactly this JSON object:

```
{  
  "relevant": 1 | 0  
}
```

## Prompt 2: Surprise Explanation - (*first-step prompt*)

You are an expert in European monetary policy and economic journalism. Only analyze the article with respect to the monetary-policy decision of the ECB. Ignore mentions referring to other central banks (e.g. Fed, BoE, SNB).

Task:

Read the article and decide whether the article itself claims the decision was:

very unexpectedly restrictive

unexpectedly restrictive

expected / neutral

unexpectedly expansive

very unexpectedly expansive

Definition of "unexpectedly":

- Compare each statement only with the expectations explicitly mentioned in the article (analysts, market participants, author ...).
- A policy decision shall be judged as "unexpectedly restrictive" if the article says the policy decision fell short of expectations (e.g. if the rate cut is smaller than expected) or comes later than expected, and vice-versa for a restrictive measure (e.g. if the rate hike is smaller than expected or comes later than expected).
- If the article gives no indication of surprise in either direction, return "expected / neutral".

Only return the explanation. The explanation must be ONE short English sentence summarizing the article's stated surprise and the direction of the surprise (restrictive/expansive).

Return exactly this JSON object:

```
{  
  "explanation": "..."  
}
```

### Prompt 3: Numerical Mapping - (*second-step prompt*)

You are an expert in European monetary policy and economic journalism. You receive the summary of another expert who evaluated an article regarding a recent policy decision by the ECB.

Task:

Based on this summary, characterize the decision using the following categories:

- 2 very unexpectedly restrictive
- 1 unexpectedly restrictive
- 0 expected / neutral
- 1 unexpectedly expansive
- 2 very unexpectedly expansive

Definition of "unexpectedly"

- The ECB's decision was more expansive than expected -> -1.
- The ECB's decision was way more expansive than expected -> -2.
- The ECB's decision was expected -> 0.
- The ECB's decision was more restrictive than expected -> 1.
- The ECB's decision was way more restrictive than expected -> 2.

Return exactly this JSON object:

```
{  
  "score":  -2 | -1 | 0 | 1 | 2  
}
```

## Prompt 4: Detection of Information Effects

You are an expert in European monetary policy and economic journalism.

Task:

The following article contains information that the recent monetary policy decision by the European Central Bank (ECB) was surprising. Analyze the article and determine whether the surprise was due to the "information effect". A surprise due to the "information effect" is defined as follows:

A monetary policy surprise that not reflects an unexpected policy action but the central bank revealing new information about its assessment of the economic outlook.

Output format:

Return exactly this JSON object:

```
{  
  "information_effect": 1 | 0,  
  "reasoning": "One short English sentence explaining the key  
    evidence from the text."  
}
```

## B Local Projections with Pure Monetary Policy Shocks

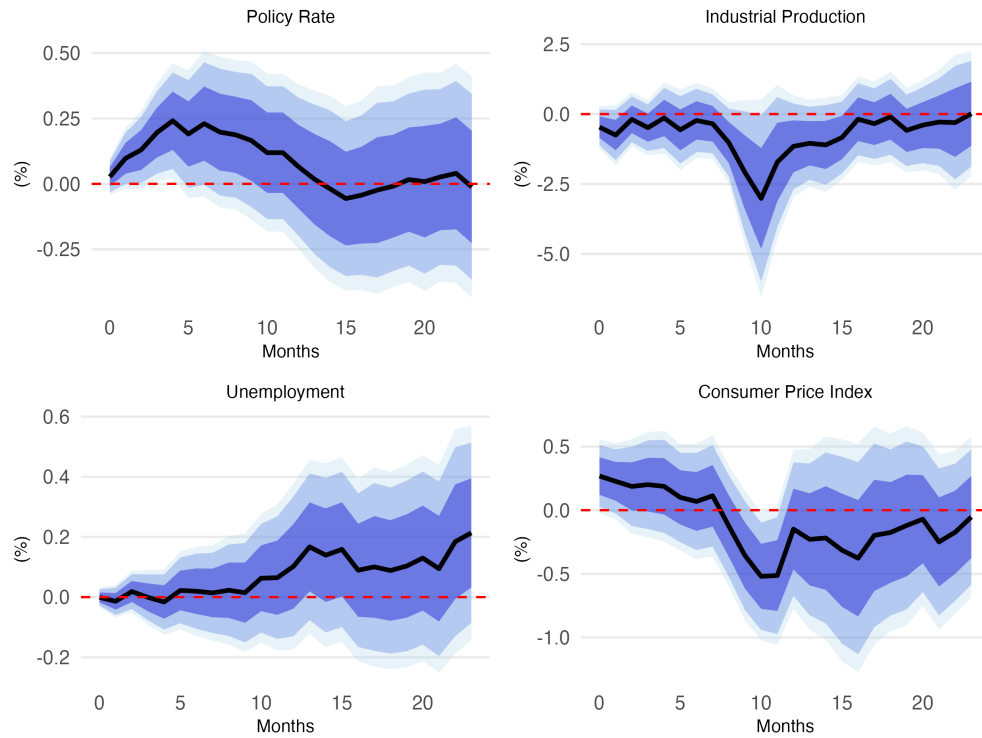


Figure 7: Baseline Local Projection with pure Newspaper-Based Monetary Policy Shocks (without information effect component)

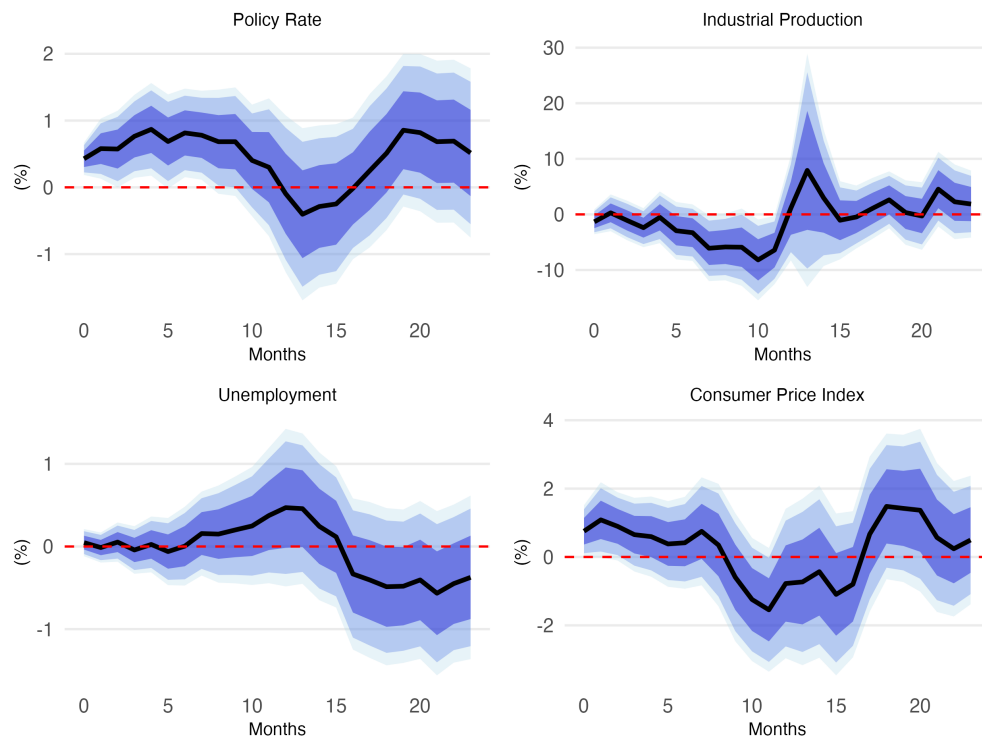


Figure 8: Baseline Local Projection in response to the Information Effect component of Newspaper-Based Monetary Policy Shocks

## C Discussion of Designated Period

The ECB, in its press conference following the monetary policy decision on 6 December 2007, reports:

”On the basis of our regular economic and monetary analyses, we decided at today’s meeting to leave the key ECB interest rates unchanged. The latest information has confirmed the existence of strong short-term upward pressure on inflation, with the HICP inflation rate reaching 3.0% in November. It has also fully confirmed our assessment that there are upside risks to price stability over the medium term. Against this background, and with money and credit growth remaining very vigorous in the euro area, the Governing Council stands ready to counter upside risks to price stability, as required by its mandate. The economic fundamentals of the euro area remain sound. However, the reappraisal of risk in financial markets is still evolving and is accompanied by continued uncertainty about the potential impact on the real economy. We will therefore monitor very closely all developments. By acting in a firm and timely manner on the basis of our assessment, we will ensure that second-round effects and risks to price stability over the medium term do not materialise. Firmly anchoring medium and long-term inflation expectations in line with price stability is all the more important at times of financial market volatility and increased uncertainty. As regards the financial markets, we will continue to pay great attention to developments over the coming weeks.”

(Trichet, 2007)

The ECB, in its press conference following the monetary policy decision on 6 November 2008, reports:

”On the basis of its regular economic and monetary analyses, at today’s meeting the Governing Council decided to reduce the key ECB interest rates further by 50 basis points, following the previous coordinated interest rate cut on 8 October 2008. The outlook for price stability has improved further. Inflation rates are expected to continue to decline in the coming months, reaching a level in line with price stability during the course of 2009. The intensification and broadening of the financial market turmoil is likely to dampen global and euro area demand for a rather protracted period of time. In such an environment, taking into account the strong fall in commodity prices over recent months, price, cost

and wage pressures in the euro area should also moderate. At the same time, the underlying pace of monetary expansion has remained strong but has continued to show further signs of deceleration. All in all, the information available and our current analysis indicate a further alleviation of upside risks to price stability at the policy-relevant medium-term horizon, even though they have not disappeared completely. At this juncture, it is therefore crucial that all parties, including public authorities, price-setters and social partners, fully live up to their responsibilities. The level of uncertainty stemming from financial market developments remains extraordinarily high and exceptional challenges lie ahead. We expect the banking sector to make its contribution to restore confidence. The Governing Council will continue to keep inflation expectations firmly anchored in line with its medium-term objective. In so doing, it supports sustainable growth and employment and contributes to financial stability. Accordingly, we will continue to monitor very closely all developments over the period ahead.”

(Trichet, 2008)

The ECB, in its press conference following the monetary policy decision on 2 April 2009, reports:

”On the basis of its regular economic and monetary analyses, the Governing Council decided to reduce the key ECB interest rates by a further 25 basis points. This decision brings the total reduction in the interest rate on the main refinancing operations of the Eurosystem since 8 October 2008 to 300 basis points. Today’s decision takes into account the expectation that price pressures will remain subdued, reflecting the substantial past fall in commodity prices and the marked weakening of economic activity in the euro area and globally. The latest economic data and survey information confirm that the world economy, including the euro area, is undergoing a severe downturn. Both global and euro area demand are likely to remain very weak over 2009, before gradually recovering in the course of 2010. Available indicators of inflation expectations over the medium to longer term remain firmly anchored in line with the Governing Council’s aim of keeping inflation rates at levels below, but close to, 2% over the medium term. A cross-check with the outcome of the monetary analysis confirms that inflationary pressure has been diminishing. After today’s decision we expect price stability to be maintained over the medium term, thereby supporting the purchasing power of euro area households. The Governing Council will continue to ensure a firm anchoring of medium-term inflation expectations. Such anchoring is indispensable

to supporting sustainable growth and employment and contributes to financial stability. Accordingly, we will continue to monitor very closely all developments over the period ahead.”

(Trichet, 2009)

The ECB, in its press conference following the monetary policy decision on 4 August 2011, reports:

”Based on its regular economic and monetary analyses, the Governing Council decided to keep the key ECB interest rates unchanged, following the 25 basis point increase on 7 July 2011. The information that has become available since then confirms our assessment that an adjustment of the accommodative monetary policy stance was warranted in the light of upside risks to price stability. While the monetary analysis indicates that the underlying pace of monetary expansion is still moderate, monetary liquidity remains ample and may facilitate the accommodation of price pressures. As expected, recent economic data indicate a deceleration in the pace of economic growth in the past few months, following the strong growth rate in the first quarter. Continued moderate expansion is expected in the period ahead. However, uncertainty is particularly high. For monetary policy, it is essential that recent price developments do not give rise to broad-based inflationary pressures. Inflation expectations in the euro area must remain firmly anchored in line with our aim of maintaining inflation rates below, but close to, 2% over the medium term. Such anchoring is a prerequisite for monetary policy to make an ongoing contribution towards supporting economic growth and job creation in the euro area. At the same time, short-term interest rates remain low and financing conditions are favourable. Thus, our monetary policy stance remains accommodative. We will continue to monitor very closely all developments with respect to upside risks to price stability. Given the renewed tensions in some financial markets in the euro area, the Governing Council today also decided to conduct a liquidity-providing supplementary longer-term refinancing operation (LTRO) with a maturity of approximately six months. The operation will be conducted as a fixed rate tender procedure with full allotment. The rate in this operation will be fixed at the average rate of the main refinancing operations (MROs) over the life of the supplementary LTRO. The operation will be announced on 9 August 2011, with allotment on 10 August 2011 and settlement on 11 August 2011, and will mature on 1 March 2012.

The Governing Council also decided to continue conducting its MROs as fixed rate tender procedures with full allotment for as long as necessary, and at least until the end of the last maintenance period of 2011 on 17 January 2012. This procedure will also remain in use for the Eurosystem's special-term refinancing operations with a maturity of one maintenance period, which will continue to be conducted for as long as needed, and at least until the end of the last quarter of 2011. The fixed rate in these special-term refinancing operations will be the same as the MRO rate prevailing at the time. Furthermore, the Governing Council has decided to conduct the three-month LTROs to be allotted on 26 October, 30 November and 21 December 2011 as fixed rate tender procedures with full allotment. The rates in these three-month operations will be fixed at the average rate of the MROs over the life of the respective LTRO."

(Trichet, 2011)

## D Comparison with updated Jarociński and Karadi Shock Series

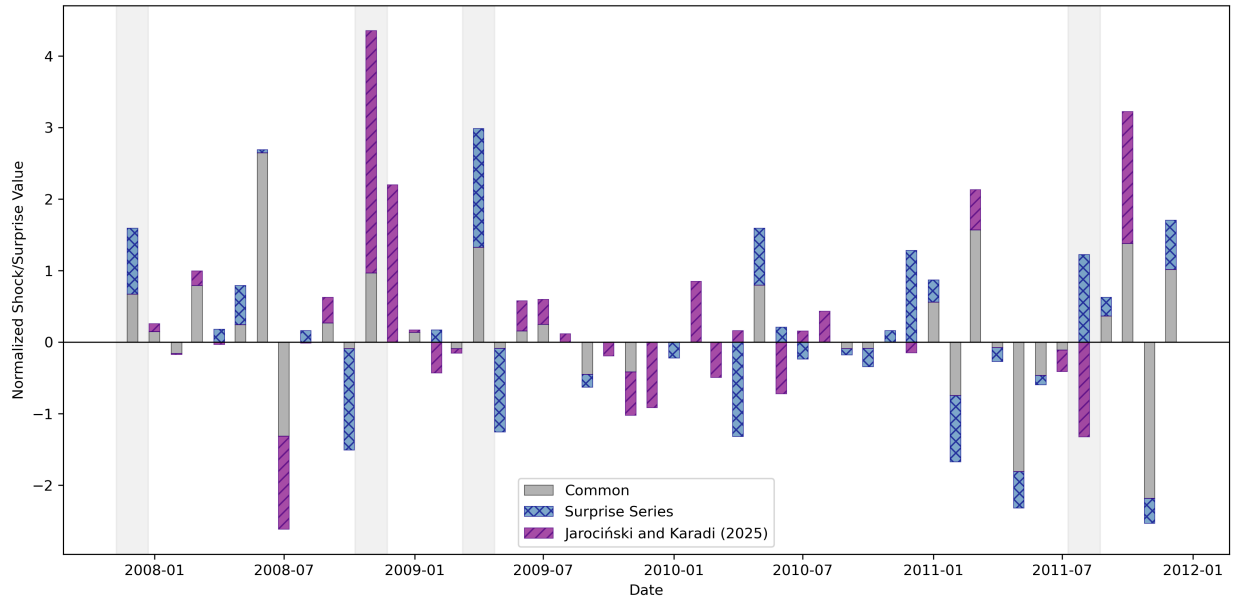


Figure 9: Shock series from Jarociński and Karadi (2025) available at [https://github.com/marekjarocinski/jkshocks\\_update\\_ecb](https://github.com/marekjarocinski/jkshocks_update_ecb) and from us composed in agreement and disagreement contributions (Own creation)