Identifying Monetary Policy Shocks in Newspapers using GPT

Felix Betz [a] Jonas Dix[a,b] Leonie Streit[a]

[a] Julius-Maximilians-Universität Würzburg, Germany
[b] Otto-Friedrich-Universität Bamberg, Germany

Research Seminar on the topic Large Language Models (LLMs)
in Monetary Policy Research,
January 22, 2026
at the Banque Centrale du Luxembourg

Outline

The Identification Problem in Monetary Macroeconomics

Presenter: Felix Betz

Motivation: The Endogeneity Problem

Julius-Maximilians-
UNIVERSITÄT
WÜRZBURG

> *"The Federal Reserve employs hundreds of PhD economists [...] to make monetary policy as endogenous as it possibly can be."*
>
> — *Nakamura and Steinsson (2018, p. 59)*

- Identifying causal effects is difficult due to endogeneity.
- Literature often uses HFI and/ or VAR models.
- We propose a scalable, narrative-based approach using LLMs.

Related Literature

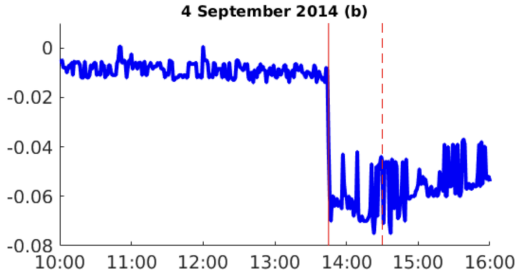
- **Monetary policy shock identification**

Approach	Key References	Share
Cholesky/ SVAR	Sims (1980); Christiano, Eichenbaum, and Evans (1999)	48.3 %
Sign restrictions	Uhlig (2005); Arias, Rubio-Ramírez, and Waggoner (2018)	18.5 %
High-Frequency	Gürkaynak (2005); Gertler and Karadi (2015)	12.8 %
Narrative	Romer and Romer (1989; 2004)	12.3 %
Other Identification	Blanchard and Quah (1989); Smets and Wouters (2007)	8.1 %

Shares according to Enzinger et al. (2025).

- **NLPs in monetary policy analysis**
  - Picault, Pinter, and Renault (2022), Geiger et al. (2025), Hansen and Kazinnik (2024).

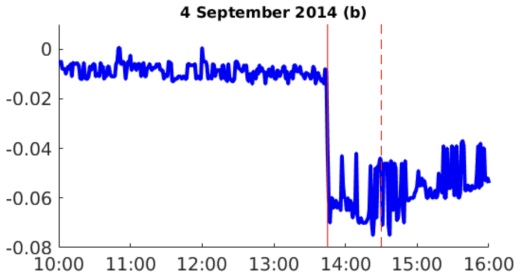
What Exactly is a Monetary Policy Shock?



Altavilla et al. (2019) intraday 2-year OIS rate around event-window of ECB GC-meeting on Sep 4, 2014.

- Sensitive to choice of financial asset (Brennan et al., 2024).
- Information effect (Jarociński and Karadi, 2020).

What Exactly is a Monetary Policy Shock?



Altavilla et al. (2019) intraday 2-year OIS rate around event-window of ECB GC-meeting on Sep 4, 2014.

- Sensitive to choice of financial asset (Brennan et al., 2024).
- Information effect (Jarociński and Karadi, 2020).


FINANCIAL TIMES

EUROPE Friday September 5 2014

Goodbye globalisation

Philip Stephens on why the world has turned, Page 9

Don Corleone missed his chance on Wall St

Notebook, Page 8

● ECB president cites final cut ● Move to prevent deflation ● Euro knocked

Draghi intervention on rates and bonds startles markets

By Claire Jones in Frankfurt and Thomas Hale and Elaine Moore in London

Clear cut

ECB policy interest rates

Mario Draghi, ECB president

Financial Times Front Page Sep 5, 2014.

LLM-Based Narrative Approach

Presenter: Jonas Dix

The Argument for Using LLMs

Julius-Maximilians-
UNIVERSITÄT
WÜRZBURG

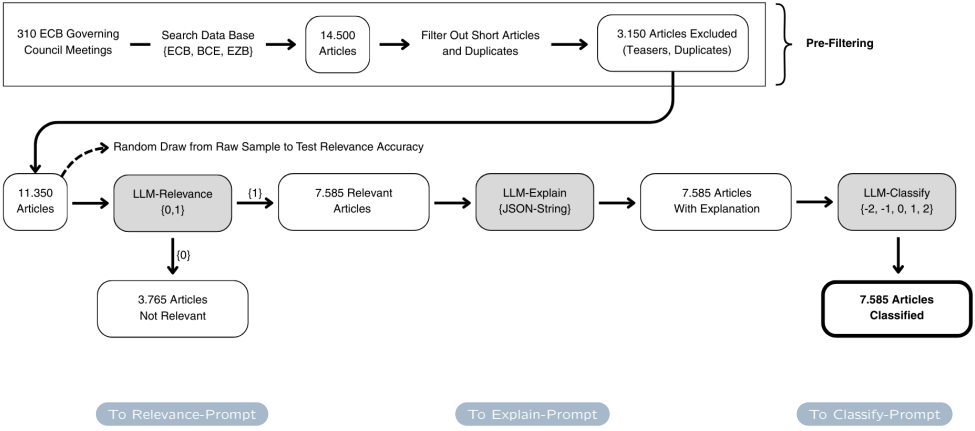
Given our set of newspaper articles ($n_{\text{raw}} = 11{,}350$), why use LLMs?

- Assuming 10 minutes per article for manual classification:

$$11{,}350 \text{ articles} \times 10 \text{ minutes} = 113{,}500 \text{ minutes} \approx 1{,}892 \text{ hours} \approx 236 \text{ work days}.$$

  - Scalability, cost-effectiveness and speed.
  - Consistency.
  - Reproducibility.
- LLMs are **very** good in labeling tasks in which humans have high agreement:
  - AEA: Recent developments in language models for economics by Elliot Ash.
  - Ash and Hansen (2023), Hansen and Kazinnik (2024), Gambacorta et al. (2024).
  - Own evidence (more later).

NLP Pipeline



Calling GPT: OpenAI - API

Julius-Maximilians-
UNIVERSITÄT
WÜRZBURG

```python
def explain_articles(
    articles: pd.DataFrame,
    model: str = "gpt-5.1-2025-11-13",
    temperature: float = 1.0,
    detect_lang: bool = True,
    effort: str = "high",
) -> pd.DataFrame:
    """Adds one column: gpt_explanation (str or None, always English)."""
    # ... (body omitted)
```

- `model` selects GPT version. We employ
  - $\texttt{gpt-4.1}_{\texttt{temp = 0}}$, $\texttt{gpt-4.1}_{\texttt{temp = 1}}$.
  - `gpt-5`, `gpt-5-nano`.
  - $\texttt{gpt-5.1}_{\texttt{noreas}}$, $\texttt{gpt-5.1}_{\texttt{medreas}}$, $\texttt{gpt-5.1}_{\texttt{highreas}}$.
- `temperature` controls randomness (not applicable for `gpt-5.X`-models).
- `effort` controls reasoning depth (not applicable for `gpt-5.X`-models).

Data and Computation of Shock Series

Presenter: Leonie Streit

Sample Coverage of Newspapers

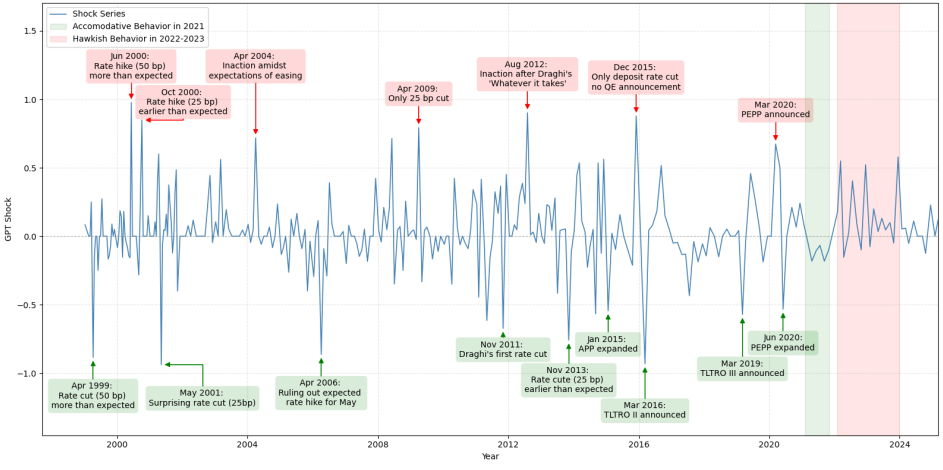
Newspaper	**Coverage**	**Country**	N_{fil}	N_{rel}
Corriere della Sera	2009–	Italy	520	329
Der Standard	2008–	Austria	278	205
El Mundo	2002–	Spain	541	406
El País	Full	Spain	833	645
FAZ	Full	Germany	2,119	938
Financial Times	Full	United Kingdom	1,575	1,223
Handelsblatt	Full	Germany	1,804	1,161
La Stampa	–2019	Italy	569	395
Le Figaro	Full	France	658	476
Les Echos	Full	France	1625	1,160
Neue Zürcher Zeitung	2000–	Switzerland	828	647
Total			11,350	7,585

Notes: N_{fil} = all retrieved articles after pre-filtering; N_{rel} = all articles after LLM-Relevance filtering.

Exemplary Classifications (FAZ)

Date	One-Sentence Summary	Score	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.

Newspaper-Based Shock Series



Shock Distribution

Manual Benchmark - Deviations

Newspaper	Accuracy Relevance	Accuracy Classification	Deviations 0	1	2
Corriere della Sera	92.3 %	93.3 %	14	1	0
Der Standard	80.0 %	91.7 %	11	1	0
El Mundo	89.3 %	80.0 %	16	4	0
El País	89.7 %	90.2 %	37	4	0
FAZ	91.6 %	83.8 %	31	5	1
Financial Times	96.8 %	82.6 %	38	8	0
Handelsblatt	91.6 %	81.4 %	48	10	1
La Stampa	84.4 %	89.5 %	17	2	0
Le Figaro	90.9 %	69.0 %	20	9	0
Les Echos	91.1 %	88.5 %	46	6	0
NZZ	90.2 %	81.5 %	22	5	0
Total	**91.0 %**	**84.0 %**	**300**	**55**	**2**

Model Performance Comparison

Model	Accuracy Relevance	Accuracy Classification	Deviations 0	1	2	Costs (h)	($)
gpt-5.1$_{\text{highreas}}$	91.0 %	84.0 %	300	55	2	34	89
gpt-5.1$_{\text{medreas}}$	91.5 %	82.3 %	293	62	1	23	55
gpt-5.1$_{\text{noreas}}$	88.9 %	79.3 %	260	66	2	3	38
gpt-5	**91.8 %**	**84.6 %**	**296**	**51**	**3**	**35**	**105**
gpt-5 nano	88.9 %	82.7 %	273	54	3	45	1
gpt-4.1$_{\text{temp = 0}}$	91.0 %	78.4 %	272	71	4	11	48
gpt-4.1$_{\text{temp = 1}}$	90.8 %	76.7 %	266	77	4	11	48

Detailed Cost Comparison

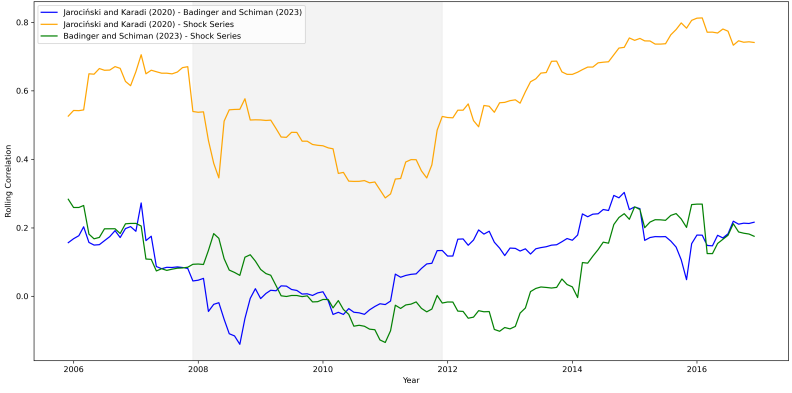
Discussion and Results

Presenter: Jonas Dix

Comparison with Established Shock Series I

Julius-Maximilians-
UNIVERSITÄT
WÜRZBURG

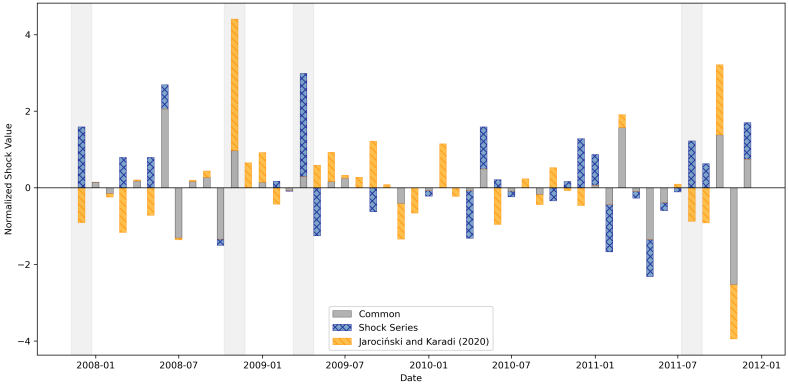
Rolling correlations of shock series from Jarociński and Karadi (2020) and Badinger and Schiman (2023):



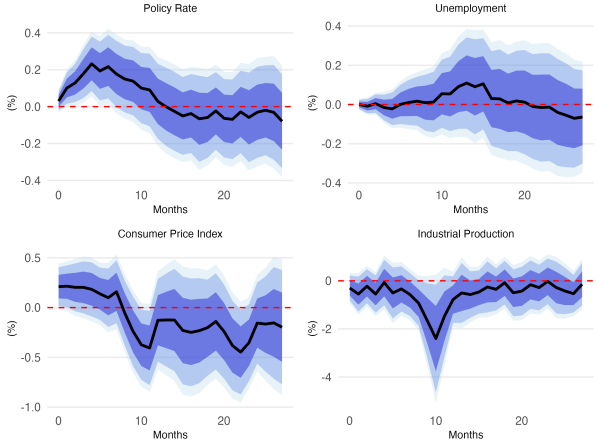
Comparison with Established Shock Series II

Julius-Maximilians-
UNIVERSITÄT
WÜRZBURG

Shock series from Jarociński and Karadi (2020) and from us in agreement/disagreement contributions:



Preliminary Results: Local Projections



Local Projections in response to a 1-unit newspaper-based monetary policy shock (own creation).

Outlook

- Further econometric analysis:
  - Refined local projections.
  - SVAR approaches (sign and magnitude restrictions) according to Badinger and Schiman (2023) with newspaper-based shocks.
  - Proxy SVAR approaches according to Stock and Watson (2018).
- More frequent event dates à la Altavilla et al. (2025) (EA-EAMPD).
- Including multidimensionality of monetary policy decisions/shocks as discussed in Gürkaynak (2005).

References I


Altavilla, Carlo et al. (2019). *Measuring Euro Area Monetary Policy*. Working Paper Series 2281. LU: European Central Bank. (Visited on 11/12/2024).


Altavilla, Carlo et al. (2025). *Monetary Transmission with Frequent Policy Events*. Tech. rep. 3157. DE: European Central Bank. DOI: 10.2866/2500139. (Visited on 01/17/2026).


Arias, Jonas E., Juan F. Rubio-Ramírez, and Daniel F. Waggoner (2018). "Inference Based on Structural Vector Autoregressions Identified With Sign and Zero Restrictions: Theory and Applications". In: *Econometrica* 86.2, pp. 685–720. ISSN: 0012-9682. DOI: 10.3982/ECTA14468. (Visited on 01/15/2026).


Ash, Elliott and Stephen Hansen (Sept. 2023). "Text Algorithms in Economics". In: *Annual Review of Economics* 15.1, pp. 659–688. ISSN: 1941-1383, 1941-1391. DOI: 10.1146/annurev-economics-082222-074352. (Visited on 11/11/2025).


Badinger, Harald and Stefan Schiman (Apr. 2023). "Measuring Monetary Policy in the Euro Area Using SVARs with Residual Restrictions". In: *American Economic Journal: Macroeconomics* 15.2, pp. 279–305. ISSN: 1945-7707. DOI: 10.1257/mac.20210035. (Visited on 12/06/2024).


Blanchard, Olivier Jean and Danny Quah (Sept. 1989). "The Dynamic Effects of Aggregate Demand and Supply Disturbances". In: *The American Economic Review* 79.4, pp. 655–673. DOI: 10.3386/w2737. (Visited on 11/05/2025).


Brennan, Connor M. et al. (Feb. 2024). "Monetary Policy Shocks: Data or Methods?" In: *Finance and Economics Discussion Series* 2024-011, pp. 1–49. ISSN: 1936-2854, 2767-3898. DOI: 10.17016/feds.2024.011. (Visited on 07/02/2025).


Christiano, Lawrence, Martin Eichenbaum, and Charles Evans (1999). *Monetary Policy Shocks: What Have We Learned and to What End?* Tech. rep. w6400. Cambridge, MA: National Bureau of Economic Research, w6400. DOI: 10.3386/w6400. (Visited on 01/15/2025).


Enzinger, Matthias et al. (Aug. 2025). *The Inflated Effects of Conventional Monetary Policy on Output and Prices*. DOI: 10.31219/osf.io/72cen_v2. (Visited on 01/19/2026).

References II

Gambacorta, Leonardo et al. (2024). "CB-LMs: Language Models for Central Banking". In: *BIS Working Papers No 1215*.

Geiger, Felix et al. (2025). "Monetary-Intelligent Language Agent (MILA)". In: *Bundesbank Technical Paper*.

Gertler, Mark and Peter Karadi (Jan. 2015). "Monetary Policy Surprises, Credit Costs, and Economic Activity". In: *American Economic Journal: Macroeconomics* 7.1, pp. 44–76. ISSN: 1945-7707, 1945-7715. DOI: 10.1257/mac.20130329. (Visited on 11/12/2024).

Gürkaynak, Refet S. (Jan. 2005). "Using Federal Funds Futures Contracts for Monetary Policy Analysis". In: (visited on 11/28/2024).

Hansen, Anne Lundgaard and Sophia Kazinnik (Apr. 2024). *Can ChatGPT Decipher Fedspeak?* SSRN Scholarly Paper. Rochester, NY. DOI: 10.2139/ssrn.4399406. (Visited on 04/11/2025).

Jarociński, Marek and Peter Karadi (Apr. 2020). "Deconstructing Monetary Policy Surprises—The Role of Information Shocks". In: *American Economic Journal: Macroeconomics* 12.2, pp. 1–43. ISSN: 1945-7707. DOI: 10.1257/mac.20180090. (Visited on 01/28/2025).

Nakamura, Emi and Jón Steinsson (Aug. 2018). "High-Frequency Identification of Monetary Non-Neutrality: The Information Effect*". In: *The Quarterly Journal of Economics* 133.3, pp. 1283–1330. ISSN: 0033-5533, 1531-4650. DOI: 10.1093/qje/qjy004. (Visited on 11/12/2024).

Picault, Matthieu, Julien Pinter, and Thomas Renault (June 2022). "Media Sentiment on Monetary Policy: Determinants and Relevance for Inflation Expectations". In: *Journal of International Money and Finance* 124, p. 102626. ISSN: 02615606. DOI: 10.1016/j.jimonfin.2022.102626. (Visited on 05/30/2025).

Romer, Christina D. and David H. Romer (1989). "Does Monetary Policy Matter? A New Test in the Spirit of Friedman and Schwartz". In: *NBER Macroeconomics Annual 1989, Volume 4*. MIT Press, pp. 121–184. DOI: 10.1086/654103. (Visited on 07/14/2025).

— (Sept. 2004). "A New Measure of Monetary Shocks: Derivation and Implications". In: *American Economic Review* 94.4, pp. 1055–1084. ISSN: 0002-8282. DOI: 10.1257/0002828042002651. (Visited on 12/10/2024).

References III

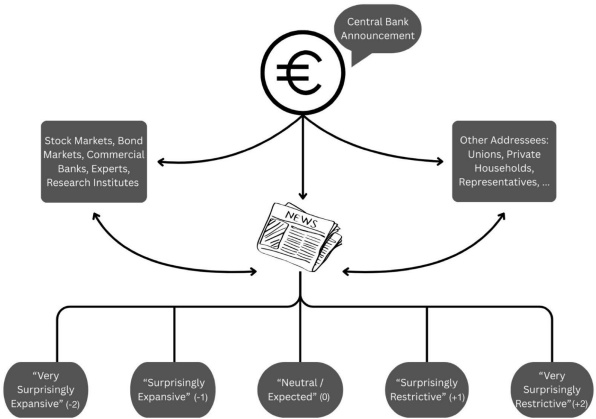

Sims, Christopher A. (Jan. 1980). "Macroeconomics and Reality". In: *Econometrica* 48.1, p. 1. ISSN: 00129682. DOI: 10.2307/1912017. JSTOR: 1912017. (Visited on 11/05/2025).


Smets, Frank and Rafael Wouters (May 2007). "Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach". In: *American Economic Review* 97.3, pp. 586–606. ISSN: 0002-8282. DOI: 10.1257/aer.97.3.586. (Visited on 01/15/2026).


Stock, James H. and Mark W. Watson (2018). "Identification and Estimation of Dynamic Causal Effects in Macroeconomics Using External Instruments". In: *The Economic Journal* 128.610, pp. 917–948. ISSN: 1468-0297. DOI: 10.1111/ecoj.12593. (Visited on 12/04/2024).

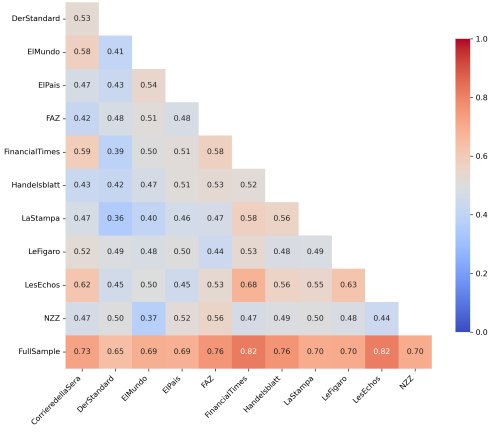

Uhlig, Harald (Mar. 2005). "What Are the Effects of Monetary Policy on Output? Results from an Agnostic Identification Procedure". In: *Journal of Monetary Economics* 52.2, pp. 381–419. ISSN: 0304-3932. DOI: 10.1016/j.jmoneco.2004.05.007. (Visited on 07/11/2025).

Appendix A: Classification Illustration



- Analogy between HFI and our narrative approach.

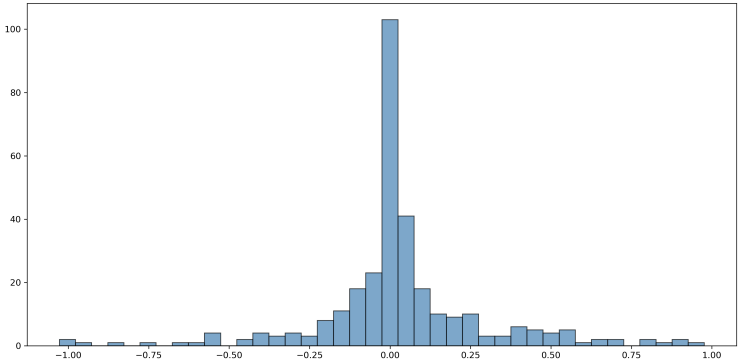
Appendix B: Cross-Newspaper Correlations



back to Classification

- Pairwise correlations mostly in the 0.4-0.6 range.

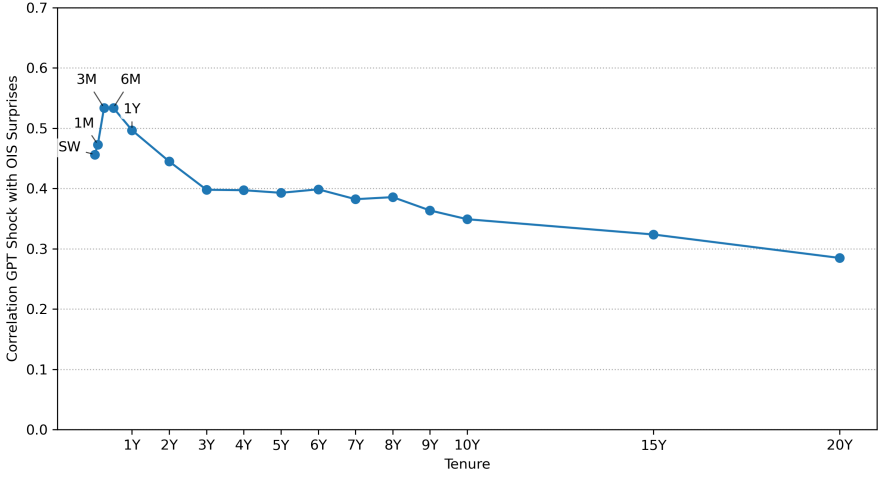
Appendix C: Distribution of Shocks



back to Shock Series

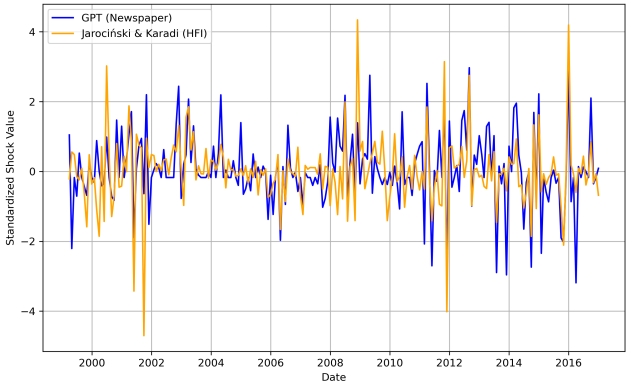
- Cluster around zero (mostly anticipated decisions)
- Most articles: neutral ($\sim$ two-thirds)
- Extreme surprises (± 2): rare ($\sim$1%)

Appendix D: Correlation with Altavilla et al. (2019)



Correlation Table

Appendix E: Validation II: Jarocinski & Karadi (2020)



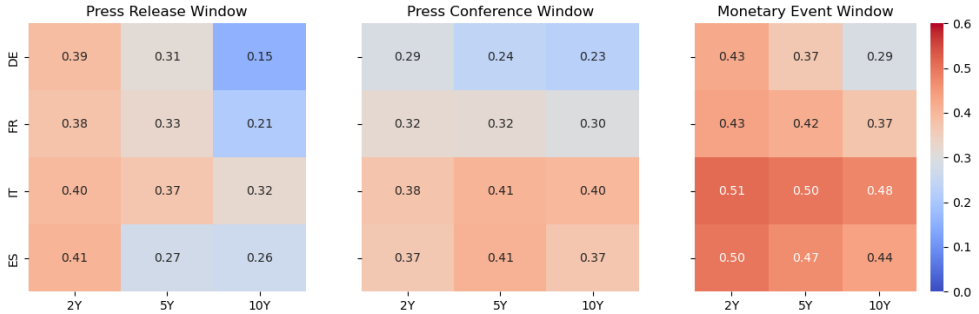
- Correlation: 0.60; Sign agreement: 64%.
- Similarity not driven by information effects.

Appdenix F: OIS Correlation (Altavilla et al., 2019)

Julius-Maximilians-
UNIVERSITÄT
WÜRZBURG

OIS Horizon	Press Release	Press Conf.	Event Window
OIS SW	0.44	0.13	0.46
OIS 1M	0.45	-0.01	0.48
OIS 3M	0.42	0.35	0.54
OIS 6M	0.45	0.34	0.53
OIS 1Y	0.43	0.32	0.48
OIS 2Y	0.36	0.30	0.42
OIS 3Y	0.32	0.25	0.37
OIS 4Y	0.39	0.14	0.37
OIS 5Y	0.40	0.12	0.37
OIS 6Y	0.40	0.12	0.37
OIS 7Y	0.38	0.11	0.36
OIS 8Y	0.38	0.11	0.36
OIS 9Y	0.37	0.10	0.34
OIS 10Y	0.35	0.11	0.32
OIS 15Y	0.30	0.10	0.29
OIS 20Y	0.21	0.13	0.26

Appendix G: Bond Correlation (Altavilla et al., 2019)



- Systematic maturity pattern: strongest alignment at short horizons (2Y, 5Y).
- Germany & France: correlations moderate, close to OIS benchmarks.
- Italy & Spain: stronger responsiveness, esp. at 2Y/5Y maturities.

Appendix H: Prompt (LLM-Relevance)

```
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
}
```

Back to NLP Pipeline

Appendix I: Prompt (LLM-Explain)

```
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": "..."
}
```

Back to NLP Pipeline

Appendix J: Prompt (LLM-Classify)

```
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
}
```

Back to NLP Pipeline

Appendix K: Model Cost Comparison

Model	Computation (hours)	Token usage (Million tokens)	Cost (US-Dollar ($))
gpt-5.1$_{\text{highreas}}$	34	22 (Input)	
		6.1 (Output)	89
gpt-5.1$_{\text{medreas}}$	23	22 (Input)	
		2.7 (Output)	55
gpt-5.1$_{\text{noreas}}$	3	22 (Input)	
		0.7 (Output)	38
gpt-5	35	22 (Input)	
		7.7 (Output)	105
gpt-5 nano	45	22 (Input)	
		0.7 (Output)	1
gpt-4.1$_{\text{temp = 0}}$	11	22 (Input)	
		0,5 (Output)	48
gpt-4.1$_{\text{temp = 1}}$	11	22 (Input)	
		0.5 (Output)	48

Back to model comparison