# Word2Prices: Embedding central bank communications for inflation prediction

Douglas Araujo (BCB), Nikola Bokan (ECB), Fabio Alberto Comazzi (ESM) and Michele Lenza (ECB and CEPR)

$8^{th}$ Annual Research Conference
Bank of Spain, Madrid
14 November 2025

# The appeal of AI

AI bears promise for the analysis in support of monetary policy

In my view, two main relevant (and related) features:

- The ability to capture very general forms of non-linearity
- The ability to tap into non-traditional sources of data

$\Rightarrow$ **An illustration of the second aspect, today**

# The aim of this paper

Suggest a method to draw insight for the purpose of real-time economic analysis

- Controlled environment, meaningful and informative text
- Design a method that is feasible in real-time and robust over time

# Implementation

- Can we extract systematically useful information about future inflation from Central Bank texts?
- Focus on the transcripts of ECB Monetary Policy Statements to extract fully in *real-time* a very simple measure of the "concepts" expressed in the press conference.
- Does this measure help to predict inflation in a rigorous out-of-sample set-up?

$\Rightarrow$ **Do Central Bank texts Granger-cause inflation?**

# Preview of the main results

- General premise: important to conduct the analysis in a rigorous out-of-sample set-up
- The information extracted from Central Bank texts helps to predict inflation
- We improve on standard sentiment analysis and other methods, and we capture information beyond what is contained in the Eurosystem forecasts
- We highlight some remaining challenges (interpretability and measurement of uncertainty)

# Literature and contribution

1. Text as data: Gentzkow, Kelly and Taddy (2019); Hirschbühl, Onorante and Saiz (2021); Ash and Hansen (2023)
2. Information content of central bank texts: Cieslak and Schrimpf (2019), Hansen, McMahon and Prat (2017), Jarocinski and Karadi (2020), Ahrens and McMahon (2021), Ahrens, Erdemlioglu, McMahon, Neely and Yang (2023)
3. Inflation forecasting: Faust and Wright (2013); Banbura, Lenza and Paredes (2024). Notice recent paper by Carriero, Pettenuzzo and Shekhar (2024).

$\Rightarrow$ **We suggest a way to process text information, and we assess it in a "policy-relevant" environment**

# Features of the out-of-sample forecasting exercise

**Benchmark AR model** for quarter-on-quarter inflation ($\pi_t$)

$$\pi_t = a + b(L)\pi_{t-1} + \varepsilon_t$$

**VAR model** for quarter-on-quarter inflation ($\pi_t$) and text-based measure ($m_t$) or alternative predictors

$$y_t = \begin{bmatrix} \pi_t \\ m_t \end{bmatrix} = A + B(L)y_{t-1} + \eta_t \tag{1}$$

**⇒ Which of the two models provides the best forecast for inflation at different horizons?**

Two steps to go from a time-series of monetary policy statements to several time-series of figures

1. Transform each word in the Monetary Policy Statements in a vector of $N$ numbers (embeddings)
2. Take average of all embeddings in a quarter $\Rightarrow$ $N$ quarterly time-series

# The concept of word embeddings

- An embedding function $M$ maps the texts into a dense $\phi$-dimensional real vector: $M : |W| \rightarrow \mathbb{R}^{\phi}$
- Intuition: start with one-hot encoding (no relationship among words) and then use forecasting logic/language structure to define the embedding mapping
- Measured vector proximity should in principle capture semantic and syntactic similarity between the words (Russel and Norvig, 2022).

# Intuition, see e.g. Ash and Hansen, 2023, Word2Vec

NLP model parameterizes the probability of a certain *word* being $v$ given its *context* C(w) as

$$P(w = v|C(w)) = f(\alpha * \rho_v, \ldots)$$

$\rho_v$ = embedding vector for word $v$; $\alpha$ = context vector (average of embeddings of neighbouring words)

$\alpha$ and $\rho_v$ estimated by maximizing the predictive accuracy of the model in the whole corpus of text fed to the model.

E. g. if $P(w = v|C(w))$ is high, $\rho_v$ tends to have "large" entries where context vector has large entries (similar to the context).

$\Rightarrow$ Similar embeddings for words that appear in similar corpus contexts

# NLP Models

- Embeddings can be learnt from scratch, starting from some random vectors, to optimise some criteria
  - This is what we do in our baseline application (Word2Vec, N=100)
  - Our embedding represents the “average word” which summarizes the press conference of each quarter in the sample
  - Not anymore state-of-the-art, but good enough (see results) and it allows us to cast the problem in an appropriate out-of-sample framework

- One can also use pre-trained word embeddings
  - This is what we do with OpenAI (N=1536) and BERT (N=768), for which we use the pre-trained embeddings - mostly because it is too costly or impossible to train the model ourselves.

# Interesting applications of this idea from “related” worlds

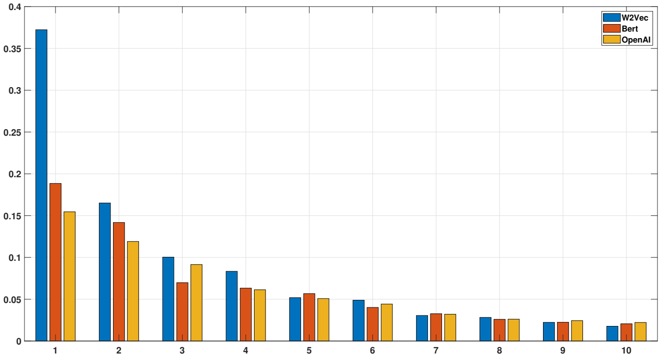
Kogan, Papanikolaou, Schmidt and Seegmiller (2021): measure of workers’exposure to technological innovation from text analisis of job descriptions

Hansen, Ramdas, Sadun and Fuller (2021): summarize the skills required of executives in job descriptions

Giavazzi, Iglhaut, Lemoli and Rubera (2020): evolution in similarity between the average embeddings of tweets of extreme right parties in Germany and the electorate to analyze how the electorate preferences evolve in response to terror attacks.

# Principal components: explained variance

Figure: Variance explained by first ten principal components




Note: Horizontal Axis: ranked PCAs. Vertical axis: Percentage of total variance of embeddings vector explained. Blue: Word2Vec. Orange: BERT. Yellow: OpenAI.

# Models

1. NLP models: i) Word2Vec, ii) BERT and iii) OpenAI
2. Placebo: i) Count of word inflation and ii) length of statements
3. Dictionary based sentiment analysis (Gardner et al. 2022): i) Inflation sentiment and ii) Governing Council Sentiment
4. LDA based forecasting: Add the time-series of the probability of (N-1) topics (Ellingsen et al. 2022)

# Features of out-of-sample

Speech database: 2000Q1 - 2023Q2

Inflation data: 2000Q1 - 2023Q4 (first estimation sample 2007Q4, recursive)

Bayesian estimation of parameters (Minnesota priors calibrated for stationary data; hierarchical model as in Giannone, Lenza and Primiceri, 2015)

# Results for the full sample, Relative MSE

| | H=1 | H=2 | H=3 | H=4 |
|---|---|---|---|---|
| | **Language Models** | | | |
| Word2Vec | 0.9685 | 0.9687 | 0.8593 | 0.8318 |
| Bert | 0.8075 | 0.7728 | 0.6756 | 0.6440 |
| OpenAI | 0.7746 | 0.7479 | 0.6714 | 0.7425 |
| | **Placebo** | | | |
| Count Inflation | 1.0336 | 1.0835 | 1.1016 | 1.1091 |
| Statement length | 1.0157 | 1.0195 | 1.0049 | 1.0030 |
| | **Sentiment** | | | |
| Sent. Inflation | 0.9408 | 0.9639 | 0.9389 | 0.9627 |
| Sent. GC | 0.9820 | 0.9805 | 0.9621 | 0.9695 |
| | **Topic Model** | | | |
| LDA four topics | 1.03 | 1.09 | 0.97 | 0.95 |
| LDA ensemble mean | 1.0454 | 0.9842 | 0.8435 | 0.8219 |

# Results for the pre-COVID sample, Relative MSE

| | H=1 | H=2 | H=3 | H=4 |
|---|---|---|---|---|
| | **Language Models** | | | |
| Word2Vec | 0.9012 | 0.7698 | 0.7304 | 0.6969 |
| Bert | 0.8799 | 0.8781 | 0.8905 | 0.9098 |
| OpenAI | 0.8623 | 0.7950 | 0.7637 | 0.7595 |
| | **Placebo** | | | |
| Count Inflation | 1.0280 | 1.0583 | 1.0916 | 1.1345 |
| Statement length | 1.0238 | 1.0597 | 1.0764 | 1.1048 |
| | **Sentiment** | | | |
| Sent. Inflation | 0.9527 | 0.8862 | 0.8901 | 0.8987 |
| Sent. GC | 0.9799 | 0.9259 | 0.9251 | 0.9378 |
| | **Topic Model** | | | |
| LDA four topics | 0.96 | 0.80 | 0.81 | 0.77 |
| LDA ensemble mean | 0.975 | 1.011 | 1.2038 | 1.2296 |

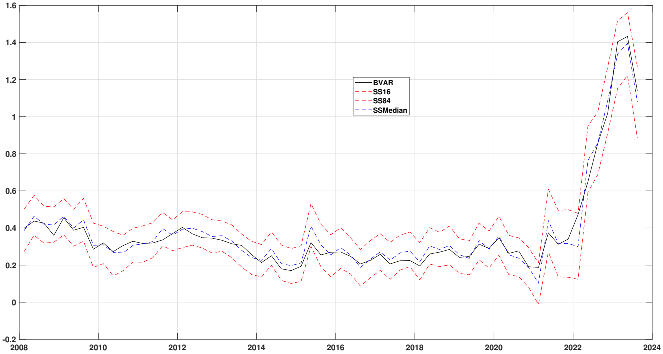
# Encompassing tests with BMPE

Are we only reflecting the information content in the economic analysis briefing, presumably incorporated in the economic projections?

Table: Optimal weight of text based forecast in an optimal combination with BMPE based inflation forecasts (Note: HAC standard errors in parenthesis.)

| | **H=1** | **H=2** | **H=3** | **H=4** |
|---|---|---|---|---|
| Pre-COVID | 0.37 | 0.55 | 0.56 | 0.52 |
| | (0.22) | (0.21) | (0.23) | (0.21) |
| Full Sample | 0.22 | 0.04 | 0.26 | 0.68 |
| | (0.24) | (0.26) | (0.32) | (0.31) |

# Comparison of h=1 forecast PC-BVAR and SS-based




Note: 2008Q1-2023Q3

# Open challenges

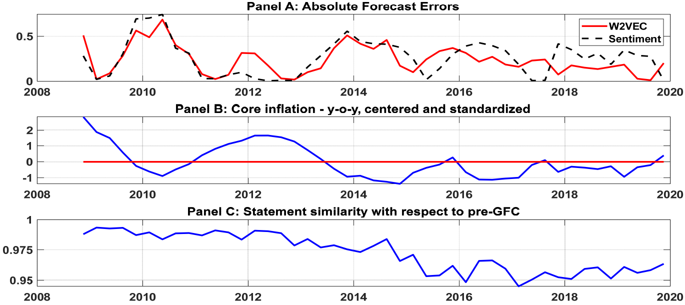
1. Interpretation: what are we capturing?
    - Ongoing work. Strong performance in periods in which MPS signals intense downside risks
    - Proximity of embeddings with words related to downside risks

2. Two-stage method, complicated to assess uncertainty

# Background Slides

# Comparison errors

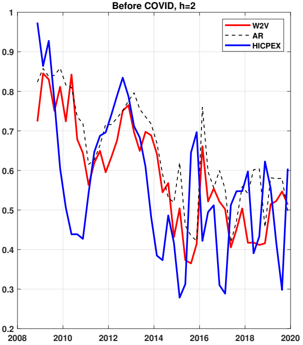
## Comparison of the absolute errors and regimes

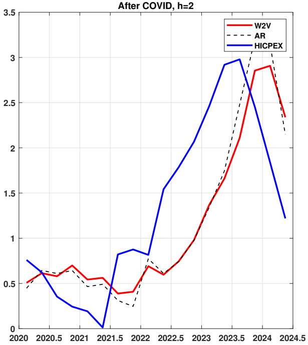



14 www.ecb.europa.eu ©

# Forecasts before and after COVID, H=2

Figure: VAR and AR Forecasts, core inflation, two quarters ahead




Note: Left panel: 2008Q2-2019Q4. Right panel: 2020Q1-2023Q4. Vertical line: change in core price levels over two-quarters.