de Araujo, Douglas Kiarelly Godoy; Bokan, Nikola; Comazzi, Fabio Alberto; Lenza, Michele

Working Paper

Word2Prices: Embedding central bank communications for inflation prediction

ECB Working Paper, No. 3047

Provided in Cooperation with:
European Central Bank (ECB)

Suggested Citation: de Araujo, Douglas Kiarelly Godoy; Bokan, Nikola; Comazzi, Fabio Alberto; Lenza, Michele (2025) : Word2Prices: Embedding central bank communications for inflation prediction, ECB Working Paper, No. 3047, ISBN 978-92-899-7216-1, European Central Bank (ECB), Frankfurt a. M., https://doi.org/10.2866/6638874

This Version is available at:
https://hdl.handle.net/10419/322058

EUROPEAN CENTRAL BANK
EUROSYSTEM

Working Paper Series

Douglas Araujo, Nikola Bokan,
Fabio Alberto Comazzi, Michele Lenza

Word2Prices: embedding central bank communications for inflation prediction

No 3047

Abstract

Word embeddings are vectors of real numbers associated with words, designed to capture semantic and syntactic similarity between the words in a corpus of text. We estimate the word embeddings of the European Central Bank's introductory statements at monetary policy press conferences by using a simple natural language processing model (Word2Vec), only based on the information and model parameters available as of each press conference. We show that a measure based on such embeddings contributes to improve core inflation forecasts multiple quarters ahead. Other common textual analysis techniques, such as dictionary-based metrics or sentiment metrics do not obtain the same results. The information contained in the embeddings remains valuable for out-of-sample forecasting even after controlling for the central bank inflation forecasts, which are an important input for the introductory statements.

JEL classification: E31, E37, E58
Keywords: Embeddings, Inflation, forecasting, central bank texts

Non-Technical Summary

How to best exploit the wide and timely availability of text offering insight about economic dynamics, for the purpose of real-time economic analysis?

The texts of the introductory statements to the press conferences of the European Central Bank provide relevant information about the dynamics of economic variables in the euro area and, hence, are an ideal testing ground for methods devoted to extract information from text.

Our proposal to extract information from text is based on the concept of word embeddings. The latter are vectors of real numbers associated with words, designed to capture semantic and syntactic similarity between the words in a corpus of text. To capture the insight of the introductory statements to the press conferences held in a specific quarter, we compute the average word-embedding across the statements of that quarter. Similar metrics have been successful to distill the informational content of complex text in management science and political economy applications.

The word embeddings are estimated by means of a simple natural language processing (NLP) model, which is re-estimated each quarter on the corpus of introductory statements available at the time of the model re-estimation. The possibility to re-estimate the NLP model allows us to run a proper out-of-sample exercise which does not suffer from the look-ahead bias typical of exercises based on large language models, which need significant resources to be re-estimated.

We find that Vector Autoregressive models augmented with our text variables improve the accuracy of euro area core inflation forecasts multiple quarters ahead, compared to the traditional univariate autoregressive benchmarks for inflation. Embeddings based on large language models also improve forecasts, but their results are potentially affected by look-ahead bias. Commonly used traditional techniques in economics, such as counting specific words or the overall sentiment of each text, lead to a more contained improvement on univariate benchmarks compared to our method.

We also show that the VAR forecasts based on text variables obtain a non-negligible weight in an optimal forecast combination with Eurosystem projections, suggesting that our methodology to transform text in data is able to draw additional insight from the text of the press conferences compared to the Eurosystem projections.

1 Introduction

The wide and ever increasing availability of text in digital format, coupled with the advances in artificial intelligence to decode it, opens new ground for the use of text as data (Gentzkow et al., 2019; Dell, 2024; Ash and Hansen, 2023; Marcucci, 2024). Perhaps the most popular methodology to draw insight for the purpose of macroeconomic analysis, including nowcasting and forecasting (see, for example, Kalamara et al., 2022; Hirschbühl et al., 2021; Barbaglia et al., 2024; de Bandt et al., 2023) is sentiment analysis. The latter is broadly defined as the attempt to extract the "emotional" tone of the message expressed in digital text about some specific topics (for example, inflation dynamics).

In this paper, we suggest an alternative method to extract information from a text that aims to draw deeper insight than is possible with state-of-the-art sentiment analysis. Our proposal relies on the concept of word embeddings and, as a case study, we focus on the usefulness of text for the purpose of the sophisticated economic analysis conducted in a central bank or, in general, in policy institutions.

Word embeddings are vectors of real numbers emerging as an intermediate output of most natural language processing models. Such vectors are designed to capture semantic and syntactic similarity between the words in a corpus of text. For example, proximity of two embedding vectors in the vectoral space should signal that such vectors represent two words expressing related concepts.

Metrics based on word embeddings have been successful to distill the informational content of complex text in management science and political economy applications. For example, Kogan et al. (2021) use the similarity in the average word embedding in patent documents and job descriptions to infer the exposure of different jobs to technological progress. Hansen et al. (2021) use average embeddings of the job descriptions of high level executives as a summary of the underlying large text descriptions. Giavazzi et al. (2020) use the 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 dramatic events, such as terror attacks. More broadly, the concept of embeddings is also used in finance to encode characteristics of different assets and investors (Gabaix et al., 2023).

Our approach to empirically validate text data based on the concept of word embeddings for economic analysis relies on three main building blocks. First, we focus on the text of the introductory statements to the press conferences of the European Central Bank. Well-recognised as an integral part of the policy toolkit (Blinder et al., 2008), these texts convey the central bank's assessments, intentions, forecasts or scenarios, and the implied policy directions. Such content

is aimed at influencing and driving both public and financial market expectations (Jansen and De Haan, 2007; Hansen et al., 2019; Ahrens and McMahon, 2021; Ahrens et al., 2023). We choose the introductory statements because we expect them to have a relevant informational content about the dynamics of economic variables in the euro area. Therefore, we assess the ability of different methods to draw the insight that should be incorporated in that source of information. Second, we run our analysis in a rigorous out-of-sample framework. In particular, the models used to estimate the word embeddings for any given press conference statement are only trained on information available at the date in which the specific introductory statement was delivered. In so doing, we approximately mirror the problem faced by the economists of policy institutions in real-time. For this reason, we cannot fully rely on the recently developed Large Language Models (LLMs) as our central tools to derive word embeddings, given that it is unfeasible or impractical to re-estimate those models several times over expanding text corpora. Carriero et al. (2024) highlight how the complex architecture of LLMs makes their frequent re-estimation cumbersome or unfeasible, imparting a "look-ahead" bias to the analysis based on such models. As our baseline model to estimate the word embeddings, we use Word2Vec (see Bengio et al., 2000; Mikolov et al., 2013), a natural language processing model which relies on a very lean architecture and allows us to fully comply with the out-of-sample logic. At the same time, Word2Vec might not be able to capture the nuances in the text as proficiently as more recent LLMs and, hence, we also use the latter models to evaluate whether the agility of Word2Vec comes at an excessive cost in terms of lost predictive accuracy. Third, we evaluate the usefulness of word embeddings at forecasting inflation, given the relevance of inflation prediction for policy institutions and the known challenges with consistently forecasting inflation (see, for example, Atkeson and Ohanian, 2001; Stock and Watson, 2007; Faust and Wright, 2013; Banbura et al., 2024). Specifically, our measure of the usefulness of the word embeddings consists in the potential improvement in inflation forecast accuracy obtained by models including the embeddings among the predictors, compared to models in which the embeddings are not included.

The first key finding is that our baseline model, i.e. a vector autoregressive (VAR) model with inflation and text data based on Word2Vec embeddings, outperforms autoregressive (AR) models in terms of out-of-sample inflation forecasting accuracy at the forecasting horizons of one to four quarters ahead, over the sample 2008-2023.

In addition, we find that using more sophisticated LLMs to estimate the embeddings is helpful. However, the loss of information incurred with Word2Vec does not appear to be acute and, for the pre-COVID period, Word2Vec performs similarly to models relying on more sophisticated architectures. One caveat on the generality of the results of this comparison between Word2Vec and more sophisticated language models is that we focus on specialized texts of an institution with an explicit inflation target and, hence, our case study may understate the superior ability of

LLMs to capture nuances in the language of more diverse text.

The VAR models with our measure based on word embeddings, achieve a better out-of-sample accuracy than those with sentiment based measures (Gardner et al., 2022), suggesting that the latter may fail to extract some of the useful information incorporated in text data.

Finally, we find that the information in our embeddings measure is not encompassed by the Eurosystem's inflation projections. The projections are a relevant element of the monetary policy briefing and, hence, their insight is integrated in the press statements. Our result suggests that our measure is able to capture nuances in the introductory press statements that go beyond the baseline Eurosystem staff view on the euro area economic outlook such as, for example, the risk assessment of the ECB Governing Council or the different judgement of the latter compared to the staff view

Our paper contributes to the group of manuscripts that use text as data in econometric analyses (see Gentzkow et al., 2019; Ash and Hansen, 2023, for an exhaustive survey of the methods, the applications and the open questions in the literature). More specifically, our paper relates to the growing literature promoting the use of AI-based sentiment indicators for macroeconomic prediction. Kalamara et al. (2022) use machine learning methods to create potentially forward-looking metrics of sentiments and uncertainty. They show that the latter can add predictive content to the standard macroeconomic variables. Barbaglia et al. (2024) employ fine-grained aspect-based sentiment analysis on a large news article text corpus and analyses the predictive power of sentiment indicators for the forecast of four macroeconomic variables namely, GDP, inflation, industrial production index and employment. de Bandt et al. (2023) performs extensive multi-level filtering analysis of French press using Factiva API and constructs indicators similar in spirit to surveys' of balanced opinions. They also show favourable forecasting properties of the indicator in a battery of standard time series models as well and in the context of Phillips curve estimation as in Bańbura et al. (2021). Relative to this line of work, we propose a different way to process text information and we show that it leads to a superior predictive performance compared to the indicators extracted in the context of sentiment analysis.

Our method to extract information from text data can be seen as a rough way to elicit the "average" topic of the introductory press statements. An alternative method to capture the topics in a document is based on the Latent Dirichlet Allocation's (LDA) method of Blei et al. (2003). For example, Larsen and Thorsrud (2019) study the predictive effects of news topics on macroeconomic variables. Larsen et al. (2021) combine LDA based topical analysis with a dictionary-based approach and confirm the out-of-sample predictive power of news topics for both inflation and inflation expectations. Angelico et al. (2022) combine LDA with a pure dictionary-

based approach to construct Twitter-based indicators of inflation expectations which helps to track survey-based inflation expectations. Zheng et al. (2024) show that LDA based news attention data may complement standard indicators for the purpose of nowcasting Chinese GDP and inflation.

While central banks have more generally found a number of uses (such as gauges of inflation expectation) for third-party texts (i.e. from social media) to inform monetary policy (Araujo et al., 2023), our work might drive a reflection in policy circles about the effects of own texts on the inflation path. In parallel, the analysis of central banking official statements is gaining prominence in the recent literature, although not with a focus on the ability of such texts to provide predictive content for inflation. For example, the ECB official communication has been analysed under several different angles in Rosa and Verga (2007), Tobback et al. (2017), Cour-Thimann and Jung (2021), Ehrmann and Talmi (2020), Ehrmann and Wabitsch (2022), Fraccaroli et al. (2022), Gáti and Handlan (2022), Hansen et al. (2019), Mumtaz et al. (2023) and Pavelkova (2022)).

Our empirical application is on inflation forecasting. The literature on the latter is very large and growing. Faust and Wright (2013) and Banbura et al. (2024) are two recent surveys.

The paper is organized as follows. Section 2 describes our data, the models and the features of our out-of-sample forecasting exercise. Section 3 discusses the transformation of the textual data into dense vectors through "embedding" techniques. Section 4 reports and discusses the empirical results. Section 5 concludes.

2 Data, models and features of the out-of-sample evaluation

In this section, we present all the elements of our empirical framework.

2.1 Data

Our dataset includes all the introductory statements to the ECB monetary policy press conferences. Such introductory statements are meticulously written texts read out by the President of the ECB at the beginning of each ECB press conference.[1] Their content follows a regular structure whereby after communicating the monetary policy decision and its rationale, important conjunctural topics are brought up first, followed by commentary and the ECB's outlook on eco-

[1] Available at https://www.ecb.europa.eu/press/press_conference/monetary-policy-statement/html/index.en.html.

nomic activity, inflation, risks to economic activity and financial and monetary conditions. The starting point of our dataset time span is the first quarter of 2002.[2] The policy statements are observed until the second quarter of 2023.

Our measure of consumer prices is the Harmonized Index of Consumer Prices excluding energy and food (HICPex), which is a common proxy for core inflation.[3] The sample for HICPex is 2002Q1-2023Q4.

2.2 Empirical models and forecasting exercise

We aim to extract a simple measure of systematically useful information/concepts for inflation prediction from the ECB press conference introductory statements. The main empirical model we use to assess the predictive power of m_t, which is our text measure based on embeddings (or other alternative text based measures), for quarter-on-quarter inflation (π_t) is the following vector autoregressive (VAR) model:

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

where A is a vector of constants, $B(L)$ a matrix of lag polynomials, and η_t is a multi-variate white noise term. The parameters of the VAR model are estimated by means of Bayesian techniques, imposing informative Normal-Inverse Wishart priors. For the parameterization of the priors, we follow the logic of the Minnesota prior, centering the distributions around values which are appropriate for variables that are considered to be *a priori* stationary. We treat the parameters governing the tightness of the priors as random variables, and we conduct posterior inference on them as suggested in Giannone et al. (2015).[4] The VAR models are estimated with four lags and the empirical results focus on point forecasts, obtained by setting the VAR parameters to the posterior mode.

The out-of-sample exercise is conducted by adopting a recursive scheme. Specifically, first, we estimate the text based measures m_t and the parameters of the VAR model with data until 2007Q4, and we produce a forecast for inflation from one to four quarters ahead. Then, we update the database with one more quarter of data and we repeat the whole exercise. We iterate

[2] This starting point is due to the availability of the text database.
[3] See Ehrmann et al. (2018) for a discussion of the properties of this measure of core inflation.
[4] For more details see Appendix A

this procedure until the exhaustion of the available sample. Notice that, for our baseline forecasts, we also train the language models used to derive the word embeddings only by using information that is available at the time of the production of the forecast, to mimic the problem of forecasters in real-time. We expand on this point in the next section.

We compare our VAR models including text data with an AR benchmark for quarter-on-quarter inflation, that is a special case of the VAR defined above in which the m_t variable is excluded from the inflation equation. If, for instance, the VAR model shows a similar performance compared to the AR benchmark for inflation, we would conclude that m_t does not contain relevant predictive information for inflation. The relative forecasting performance of the VAR and the AR models at different forecast horizons h is measured by computing the ratio of mean squared forecasting error (MSFE) of the VAR with respect to the corresponding AR model's MSFE for each forecast horizon. A value smaller than one indicates that the VAR outperforms the AR model. Notice that the MSFE for a specific forecasting horizon h is computed by looking at the cumulative change in consumer prices over that forecast horizon.

3 Extracting information from text

A key step in our empirical analysis is to extract meaningful information from the introductory statements to the ECB press conferences and to translate it in a continuous real vector space. The result is commonly referred to as an embedding. Words representing similar concepts will be closer neighbors in the embedding space, with directions roughly reflecting human ideas of overall word relatedness, and even including "surprising" relatedness along various salient semantic dimensions (Russell and Norvig, 2022).[5]

More formally we can define the embedding $M : \mathbb{W} \to \mathbb{E}$ which maps words of some text W[6] into some latent space $\mathbb{E} \subseteq \mathbb{R}^n$ with defined distance function $d : \mathbb{E} \times \mathbb{E} \to \mathbb{R}_+$. The dimensionality of the subspace $\mathbb{E}$ is a specification choice: in principle, higher dimensional representations allow for more nuance in differentiating words but they also require bigger and more diverse textual datasets and more sophisticated architectures to learn these concepts satisfactorily.

The mapping is learned through a procedure which minimizes a loss function with respect to the collocation of words in massive textual corpora. Given its inherent non-linearity, M is almost

[5] A typical example of the algebraic manipulation of text concepts enabled by embeddings and illustrating the concepts of semantic and syntactic similarity is the reconstruction of the concept of "queen" by adding the values for "woman" and "king", minus the embedding for "man" (Russell and Norvig, 2022).

[6] Or any other subdivision of texts such as parts of words (commonly referred to as tokens), or sentences, paragraphs, etc.

always modelled as a neural network. Over time, a few architectures - all based on neural networks - have proven to efficiently encode different latent meanings in a way that makes them useful across various applications and domains.

An early such architecture to gain prominence was Word2Vec (Mikolov et al., 2013). The transformer model (Vaswani et al., 2017), unveiled in 2017 and then successfully deployed in 2018 in a flexible end-to-end large language model named the BERT model (Devlin et al., 2018), spawned a newer generation of embedding models that take the deeper context of each text into account. More recently, the GPT-class of models developed by OpenAI further pushed the frontier of embeddings.[7]

3.1 The baseline method to estimate word embeddings: Word2Vec

More sophisticated models might result in more precise quantification of the meaning of a text by taking advantage of broader context windows and leveraging higher-level representations of the concepts. However, especially Large Language Models (LLMs) are very costly and, in practice, unfeasible to train from scratch. Moreover, most large language models are of recent development and are estimated with data which range until very recently. For these reasons, truly out-of-sample forecasting exercises with outcome of such pre-trained models would only be meaningful on very short samples.[8]

At the same time, less recent language models, such as Word2Vec, can easily be trained from scratch, allowing researchers the control over the sample cut-off date and the information set used to train them. This is the route we take to derive the word embeddings used for our baseline forecasts in this paper. Specifically, in our application, we adopt the Word2Vec model described in Bengio et al. (2000) and Mikolov et al. (2013). The model is trained using only the vocabulary μ_T in the documents contained in each sample period. That is, at the cutoff date $T = 2007Q4$, the vocabulary comprises all the unique words appearing in the ECB introductory statements to the Press Conference up to that date: $mu_{2007Q4} = \{w : w \in \cup_{t=2002Q1}^{2007Q4} M_t\}$. Then, for the next period, the cutoff date T moves forward by one quarter and the exercise is repeated. We do not use text pre-cleaning.[9] In our implementation, we use the Continuous Bag of Words (CBOW) variant of

[7] The variety of modern embedding models is best illustrated in open leaderboards such as the Massive Text Embedding Benchmark, available at https://huggingface.co/spaces/mteb/leaderboard (MTEB, Muennighoff et al., 2023).

[8] At the same time, using the off-the-shelf embeddings released by the providers of LLMs could be appropriate for many applications (Araujo, 2023).

[9] We don't pre-clean the text data because such procedure does not appear to be needed in state-of-the-art large language models, which we also use in this paper as alternative methods. We apply the same cleaning procedures to all the methods for comparability purposes. Our results with Word2Vec and pre-cleaning are similar to those

Word2Vec. In this setup, the model is trained to predict a target word based on its surrounding context words. We use a context window C of length 5, enabling the model to consider up to 5 words before and 5 words after the target word. The embedding dimension is set to the default $\phi_{\text{Word2Vec}} = 100$.[10]

A technical exposition of Word2Vec is beyond the scope of this paper.[11] Here we provide only an intuitive description of the method. In very general terms, the model parametrizes the probability of a certain *word* w being v given its *context* C(w) as:

$$P(w = v|C(w)) = f(\rho_c * \rho_v, ...),$$

with ρ_v = embedding vector for word v; ρ_c = context vector (average of embeddings of neighbouring words). The probability is modelled by means of a shallow neural network and ρ_c and ρ_v are estimated by maximizing the predictive accuracy of the model in the whole corpus of text fed to the model. Eg, if $P(w = v|C(w))$ is high, ρ_v tends to have "large" entries where the context vector has also large entries.

As a result, words that appear in similar corpus contexts tend to have similar embeddings, as seen in Fig. 1. This graph shows how the concepts in the press conference data are related to one another, transformed by a dimensionality-reduction technique called t-SNE (Van der Maaten and Hinton, 2008) to represent in two dimensions the ϕ-dimensional embeddings for each word. In Fig. 1, we can see that words related to monetary policy ("refinancing", "purchase", "TLTRO") are close to one another, as are real-economy factors that commonly affect inflation ("wages", "bottlenecks", "production", with "supply-side" also somewhat close). Another group of words includes "pandemic" and "coronavirus", while the words "inflation", "2" and "remain" are also close to one another as a group.

Note that our embedding step does not incorporate any information about inflation beyond what is included in the texts themselves. The embeddings are simply learned from the task of predicting a word given neighbouring words, without explicitly targeting inflation forecasting.

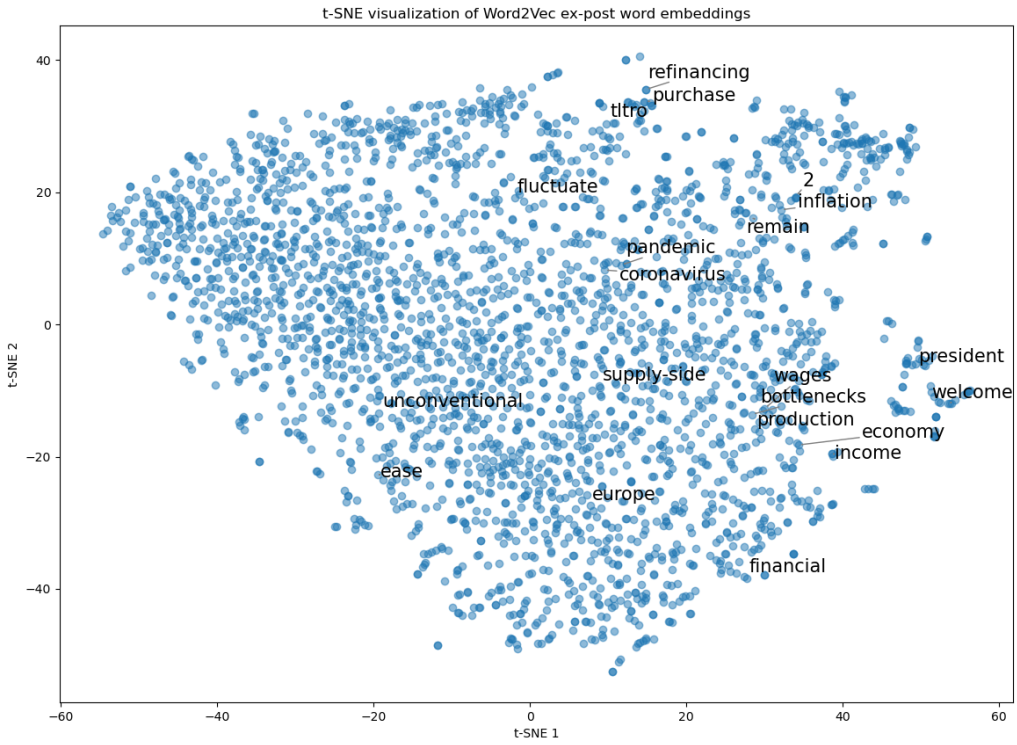
We go from the set of word embeddings related to each individual introductory statement to the quarterly series m_t by averaging the vectors of words twice: first, we average those of each

without pre-cleaning and are available on demand.

[10] For brevity, subsequent mentions of the vector dimensionality ϕ will not indicate in the subscript the technique used unless it is unclear from the context.

[11] The interested reader is referred to Bengio et al. (2000) and Mikolov et al. (2013) for a detailed description of the techniques.

Figure 1: t-SNE representation of the embedding concepts



statement and, then, we take another average across the statements released in each quarter.

3.2 Alternative methods to extract text data

Besides comparing our baseline VAR model including the m_t extracted by using Word2Vec with the simple AR benchmarks, we also conduct comparisons with other VAR models including measured derived by means of alternative methods to extract information from text.

First, we use the pre-trained word embeddings which emerge as a by-product of more sophisticated language models than Word2Vec. Our focus is on two different alternatives, BERT (Devlin et al., 2018) whose embedding vectors are 768-dimensional and OpenAI, with vectors of 1536 elements.[12] As previously discussed, forecasting exercises based on these pre-trained embeddings may suffer

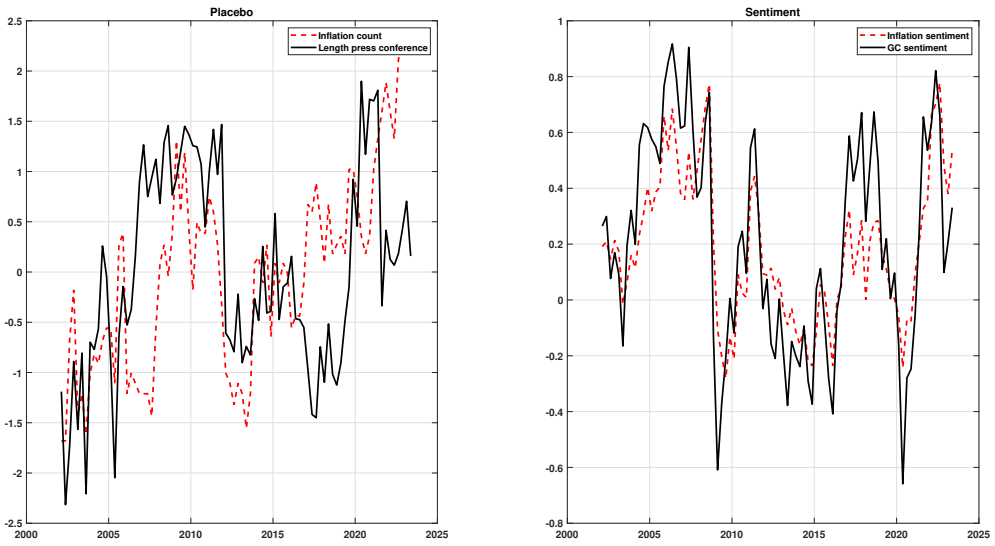
[12]The embeddings publicly released by OpenAI are available at https://openai.com/index/new-and-improved-embedding-model/.

from a "look-ahead" bias. However, comparing the results of our baseline VAR model with those based on these embeddings may reveal how potentially strong such bias might be and whether the agility of Word2Vec comes at an excessive cost in terms of lost predictive accuracy compared to models which can better capture the nuances of language.

Second, we also look at a group of indicators which we define as placebo, because they represent metrics derived from each text but with little or no reason why they should convey meaningful information. The first placebo variable is the count of the word "inflation" in each introductory statement. The second placebo is the length, in words, of each statement.

Finally, we look at sentiment indicators, which are meant to convey an economic signal, albeit extracted in a different way than the embeddings. Specifically, following Gardner et al. (2022), we use dictionary-based sentiment metrics to measure the ECB's sentiment on inflation, as well as the Governing Council's sentiment.[13] Figure 2 plots the placebo (left panel) and the sentiment indicators (right panel).

Figure 2: Placebo and Sentiment indicators.



Note: The placebo indicators enter in log-terms in the VAR. Exclusively to preserve the scaling of this figure, they are centered and standardized.

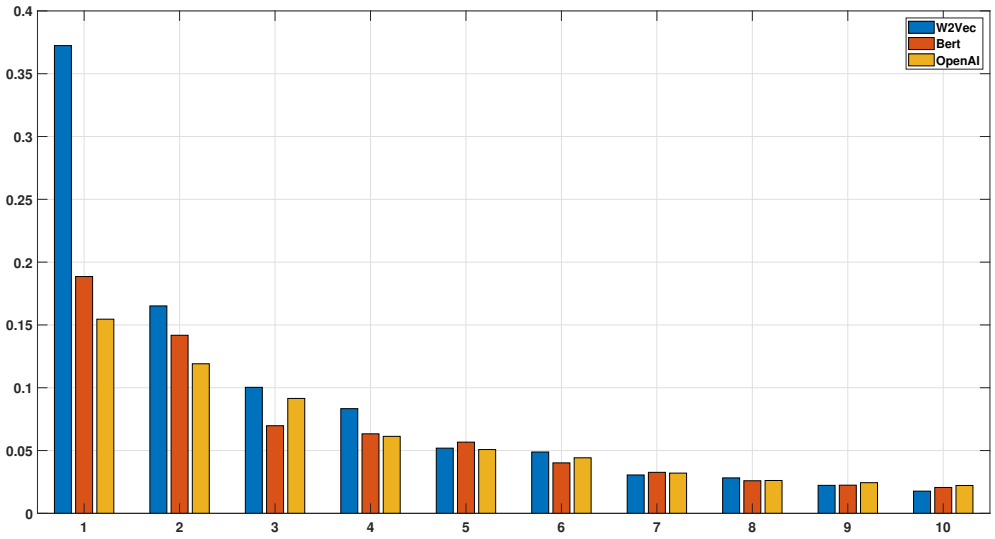
[13] We also considered a similar measure focused on labour markets, monetary policy and output, but they were much less informative and we omitted the results in the interest of brevity.

4 Empirical results

4.1 Principal components of embedding vectors

Figure 3 reports the share of the total panel variance explained by the principal components (PCAs) in the panel of Word2Vec based (blue bars), BERT based (orange bars) and OpenAI based embeddings (yellow).

Figure 3: Variance explained by first ten principal components

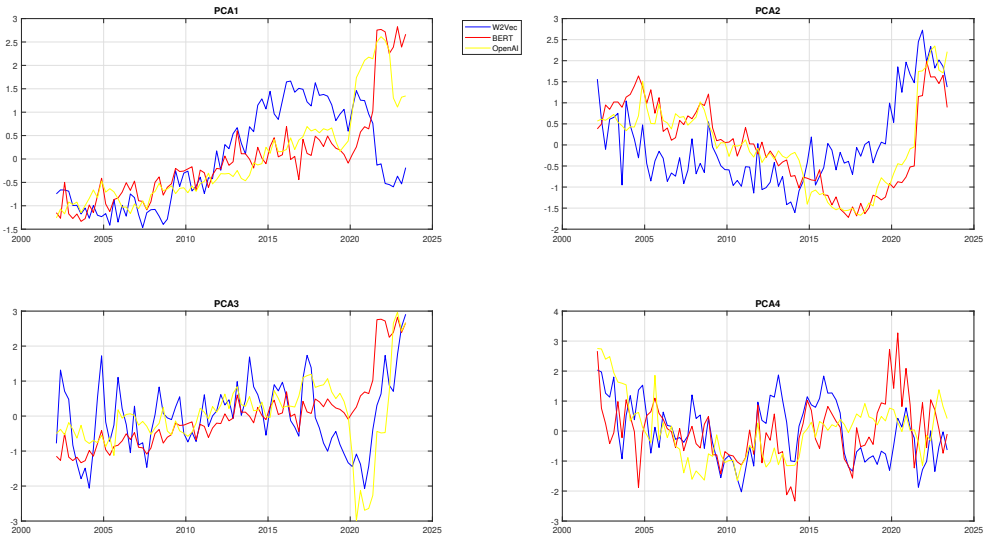


Note: Horizontal Axis: ranked PCAs. Vertical axis: Percentage of total variance of embeddings vector explained. Blue: Word2Vec. Orange: BERT. Yellow: OpenAI.

As it is clear from the figure, taken together, four PCAs explain between about 50% and 70% of the total variance of the time-series of the elements of the embedding vectors, which is clear evidence that the latter display a large extent of common fluctuations. These four PCAs also stand out individually in terms of explained variance, remaining above the threshold of 5% of total panel variance explained, robustly across methods. For these reasons, we summarize the large-dimensional embeddings vectors stemming from the three language models by means of these four PCAs, which we interpret as estimates of the common factors driving the embeddings (on this point, see, in particular, Forni et al., 2000; Stock and Watson, 1998). These four PCAs for the three methods will represent the m_t vectors we include in the three VAR models augmented with word embeddings in our empirical analysis. In the appendix, we also report a comparison with an

alternative inflation forecasting method (based on the Word2Vec word embeddings) which does not reduce the dimensionality of our panel by means of PCAs and shows that focusing exclusively on the four PCAs which explain the bulk of the variance of our embedding based text variables is unlikely to affect the forecasting results. Figure 4 shows the time series of the four PCAs for the three language models we consider, estimated over the full sample at our disposal.

Figure 4: In-sample estimates of first four principal components - Word2Vec/BERT/OpenAI



4.2 Forecasting core inflation by means of text based variables

Table 1 presents the results of our forecasting evaluation over the full sample.

The first result (first row) is that our text variable extracted from the ECB introductory statements and based on Word2Vec improves the accuracy of the inflation forecasts compared to the univariate autoregressive benchmark, despite the notorious difficulty to outperform such benchmarks for inflation forecasting (see Atkeson and Ohanian, 2001; Stock and Watson, 2007; Faust and Wright, 2013; Banbura et al., 2024). The improvement is more relevant at longer horizons. For example, the MSFE at the horizon of four quarters ahead is more than 16% lower than the AR model counterpart.

When more advanced language models are used to interpret the text of the ECB introductory statements (second and third row), the performance of the BVAR model compared to the bench-

Table 1: Results for the full sample

	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

mark further improves, although it is not possible to exactly attribute this gain to the models' sophistication or to the (partial) in-sample nature of the embedding estimation which, as explained in the previous section, could impart a "look ahead" bias to the estimates of m_t.

The alternative measures built using the same original textual data either do not improve inflation forecasts or they do it to a smaller extent than our baseline measure. Precisely, the count of "inflation" mentions in each text and the statement's length (rows four and five) contribute, if anything, to less accurate predictions. As these were "placebo" methods, the result is in line with our expectations. Perhaps more interestingly, the bottom two rows, based on sentiment metrics, present values only slightly lower than one, suggesting that they are able to extract less information from the text of the ECB introductory statements than our baseline measure. This result suggests that embeddings are able to capture sufficient nuance in the texts that sets them apart from other text-based metrics, even when using language models which are not state-of-the-art.

The last five years in our evaluation sample include the COVID pandemic and its aftermath. These years have been characterized by a large macroeconomic volatility and rather unusual economic dynamics. Table 2 reports results only for the pre-COVID sample from 2008 to 2019, to analyze whether the good performance of our text based measures is a general finding or is mainly due to a very strong performance over the most recent turbulent years.

The results in table 2 broadly confirm those over the full sample, with an important qualification related to the results from language models. Namely, now the Word2Vec results are even more competitive and, especially for longer forecasting horizons, they are at least as good as those from more sophisticated language models. Hence, at least for the sample excluding the COVID pandemic and its aftermath, the enhanced sophistication of the BERT and OpenAI underlying

Table 2: Results for the pre-COVID sample

	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

models does not bring much material advantage for the sake of interpreting the ECB introductory statements for inflation forecasting. This result may also tentatively suggest that the "look ahead" bias we have discussed for the large language models may not be always empirically relevant.

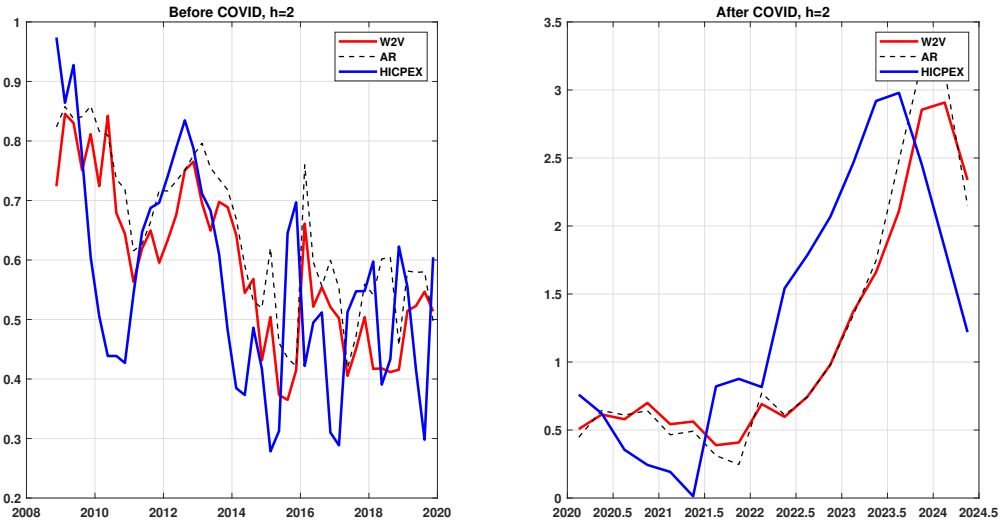
Figure 5 shows the forecasts from the baseline VAR model (red solid line) and those from the autoregressive model (black dashed) together with observed inflation. The forecasts refer to an horizon of two quarters ahead. To isolate the period of high inflation at the end of the sample, we split the figure in two panels. In the left panel, we report forecasts and observed inflation for the 2008Q2-2019Q4 period, and in the right panel we report the same variables for the sample 2020Q1-2023Q4.

The figure shows that the text based variable systematically helps to capture the low inflation environment which characterized the euro area for a large part of the decade preceding the COVID pandemic.

4.3 Eurosystem forecasts and text based measure

The monetary policy briefing which informs the decisions of the ECB Governing Council includes the outcomes of the (Broad) Macroeconomic Projection Exercise (BMPE). The latter is a quarterly exercise in which the staff of the Eurosystem provides its view on the euro area macroeconomic dynamics over the subsequent three years. Given their prominence as an input for the monetary policy decisions, the BMPE outcomes are an important input for the ECB introductory statements. Hence, one possible reason for the proficiency of our text based measure to forecast inflation could be that the word embeddings capture mainly the BMPE input.

Figure 5: 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.

In order to evaluate whether this is the explanation for the predictive power of our text based measure, we test whether, once considering the BMPE inflation projections, the forecasts from our baseline VAR become "redundant". In practical terms, we recur to an encompassing test. More precisely, for each horizon h under analysis, we investigate what would be the weight λ for our text based forecast π^m_{t+h} in an optimal inflation forecast combination with $\pi^{BMPE-based}_{t+h}$, i.e. an inflation forecast based on the BMPE[14]:

$$\pi^{\text{opt}}_{t+h} = \theta + \lambda\pi^m_{t+h} + (1-\lambda)\pi^{BMPE-based}_{t+h}. \tag{2}$$

Values of λ close to zero suggest that our text based forecast is redundant, once we consider the BMPE forecasts, while large values suggest that the information in our text based forecast adds information to what is included in the BMPE projections.[15] Table 3 reports the estimates of λ.

[14] We derive our BMPE based forecast from a VAR model with core inflation and the time-series of the BMPE forecasts at one to four quarters ahead produced in the context of each quarterly exercise in our sample. The cut-off date for each exercise is generally a few days earlier than the ECB press conference. In addition, for both the BMPE and the text based forecasts we assume that we know the inflation reading for the current quarter, which amounts to assume that we are running the VAR estimation at the end of each quarter, when the inflation outcome for the quarter becomes available.

[15] In order to compute the optimal λ, we substract the BMPE based forecast from both sides of equation 2 and, then, we regress the forecast errors of the BMPE-based forecast on a constant and the gap between the text based and the BMPE based forecasts. The estimates for λ are given by the coefficient on the gap between the text based

The top panel refers to the sample 2008-2019, while the bottom panel refers to the full sample.

Table 3: 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)

Considering the sample before the COVID pandemic, the text based forecasts provide signal that are, generally, not encompassed by the BMPE. When the sample includes the turbulent period which started with the COVID pandemic, the result is somewhat attenuated for short-term forecasts but remains robust at longer horizons. This is consistent with Rosa and Verga (2007)'s result that ECB introductory statements and macroeconomic variables act as complementary, not substitutable, information for the prediction of the policy rate.

5 Conclusions

We suggest a simple way to extract information from text, especially (but not exclusively) for the purpose of the economic analysis conducted in a policy institution. Our proposal is based on the concept of "word embeddings" and the measure based on such embeddings is derived by using a relatively simple natural language processing model (Word2Vec), which can be re-estimated at little cost. This feature allows an appropriate out-of-sample analysis, which is paramount to empirically validate our text based measure for the purpose of forecasting in real-time.

Our text based measure helps to predict inflation, adding information to simple univariate benchmarks. Remarkably, using Word2Vec leads to a relatively limited information loss compared to more sophisticated and state-of-the-art language models. Moreover, our proposal to extract information from text produces a measure with superior predictive power than the popular sentiment metrics often adopted to transform text in data.

Despite these promising results, certain challenges and open questions remain. One challenge is related to the interpretation of our embeddings measure. Another important open question, perhaps more relevant for a potential future extension of our method to density forecasting, is how to assess the uncertainty surrounding the predictions in our two stage procedure in which,

and the BMPE based forecasts. We compute HAC standard errors.

first, we derive the text based measure and, then, we use it in a downstream econometric model (Battaglia et al., 2024). Finally, communication varies significantly across central banks (see, for example, Bulíř et al., 2013); assessing whether our results hold for other institutions and economies would allow us to unveil whether some specific policy and/or communication practices are more conducive to extract useful information from text, for the purpose of macroeconomic forecasting.

A Prior distributions and spike-and-slab analysis

In this appendix, we describe the priors we used in our BVAR and then we report an analysis based on a regression model with hyerarchical "spike and slab" priors.

A.1 BVAR priors

Assume that $y_t = [\pi_t \;\; m_t]'$. We adopt the following VAR setting:

$$\begin{aligned} y_t &= C + B_1 y_{t-1} + \cdots + B_p y_{t-p} + \epsilon_t, \\ \epsilon_t &\sim \mathcal{N}(0, \Sigma), \end{aligned}$$

where y_t is an N-dimensional vector of time-series, $B_1, \ldots, B_p$ are $N \times N$ matrices of coefficients on the p lags of the variables, C is an N-dimensional vector of constants and Σ is the covariance matrix of the errors. We have $p = 4$. The largest available estimation sample ranges from 2002Q1 to 2023Q2.

Potentially, this model may be subject to the "curse of dimensionality" due to the large number of parameters to be estimated, relative to the available sample. In such circumstances, the estimation via classical techniques would very likely result in overfitting the data and large estimation uncertainty. De Mol et al. (2008) and Banbura et al. (2010) showed that imposing informative priors which push the parameter values of the model toward those of naïve representations (such as, for example, the random walk model) reduces estimation uncertainty without introducing substantial bias in the estimates, thanks to the tendency for most macroeconomic and financial variables to co-move. In fact, in presence of comovement, the information in the data strongly "conjures" against the prior, so that the parameter estimates reflect sample information even if very tight prior beliefs are enforced.

For this reason, we adopt a Bayesian estimation technique. The prior for the covariance matrix of the residuals Σ is Inverse Wishart, while the prior for the autoregressive coefficients is normal (conditional on Σ). More in details, the prior distributions in our Bayesian VAR are specified as follows. For the prior on the covariance matrix of the errors Σ, we set the degrees of freedom of the Inverse Wishart distribution equal to $N + 2$, the minimum value that guarantees the existence of the prior mean, and we assume a diagonal scaling matrix Ψ, which we parameterize

by setting the diagonal values equal to the variance of the residual from an AR(1) model for each individual variable. The baseline prior on the model coefficients is a version of the Minnesota prior (see Litterman, 1979). This prior is centered on the assumption that each variable follows an independent random walk process in levels, possibly with drift. We difference the inflation data before entering the models and we entertain the prior belief that the text based variables are stationary. For this reason, our prior distribution for all the VAR coefficients is centered on zero. The prior second moments for the VAR coefficients are:

$$\operatorname{cov}\left((B_s)_{ij}, (B_r)_{hm} \middle| \Sigma \right) = \begin{cases} \lambda^2 \frac{1}{s^2} \frac{\Sigma_{ih}}{\psi_j/(d-n-1)} & \text{if } m = j \text{ and } r = s \\ 0 & \text{otherwise} \end{cases} .$$

Notice that the variance of the prior is lower for the coefficients associated with more distant lags and that coefficients associated with the same variable and lag in different equations are allowed to be correlated. The lower prior for longer lags captures the idea that autocorrelation is more likely to die out the longer the lags we are considering. This feature allows us to specify models with long lags. The terms Σ_{ih}/Ψ_j account for the relative scale of the variables. The prior for the intercept C is non-informative.

The setting of the prior distributions depends on the hyperparameter λ, which effectively determines the overall tightness of the prior distributions for the model coefficients because it controls the scale of all variances and covariances. In setting this parameter, we follow the theoretically grounded approach proposed by Giannone et al. (2015), who suggest to treat the hyperparameters as additional parameters, in the spirit of hierarchical modelling. As hyper-prior (i.e., prior distribution for the hyperparameter), we use a proper but almost flat distribution.

A.2 Comparison between BVAR forecasts and forecasts based on a regression model with hyerarchical "spike-and-slab priors"

The BVAR results in the text were derived using four PCAs of the embedding vector extracted by applying Word2Vec. The dimensionality reduction was justified by the consideration that the time-series of the embeddings display a very large extent of commonality and four "common factors" capture the bulk of the variance in the panel of time-series.

This feature should also imply that, intuitively, our results on the predictive ability of our text measure should not depend on the dimensionality-reduction step of our analysis. In this subsec-

tion, we show this by regressing future inflation on all 100 time series $\{\phi_t^{(1)}\}, \{\phi_{t}^{\cdot\cdot}\}, \{\phi_t^{(100)}\}$ of recursively calculated quarterly embeddings. The regression is estimated with bayesian methods, imposing a spike-and-slab prior developed in Mitchell and Beauchamp (1988), with a hierarchical hyper-prior on the degree of shrinkage and model size as in Giannone et al. (2021). This model nests the polar cases of the lasso regression (Tibshirani, 1996) and ridge regression and all the intermediate cases. In other words, our regression estimates a linear model that nests the full spectrum of sparse and dense linear regression models with the full set of embedded dimensions.

More in details, we estimate the following regression:

$$\pi_t = \alpha(L)\pi_{t-1} + \beta_1\phi_t^{(1)} + \cdots + \beta_{100}\phi_t^{(100)} + \epsilon_t, \quad \epsilon_t \sim \text{i.i.d.}\mathcal{N}(0, \sigma^2)$$

The prior for the regression coefficients is:

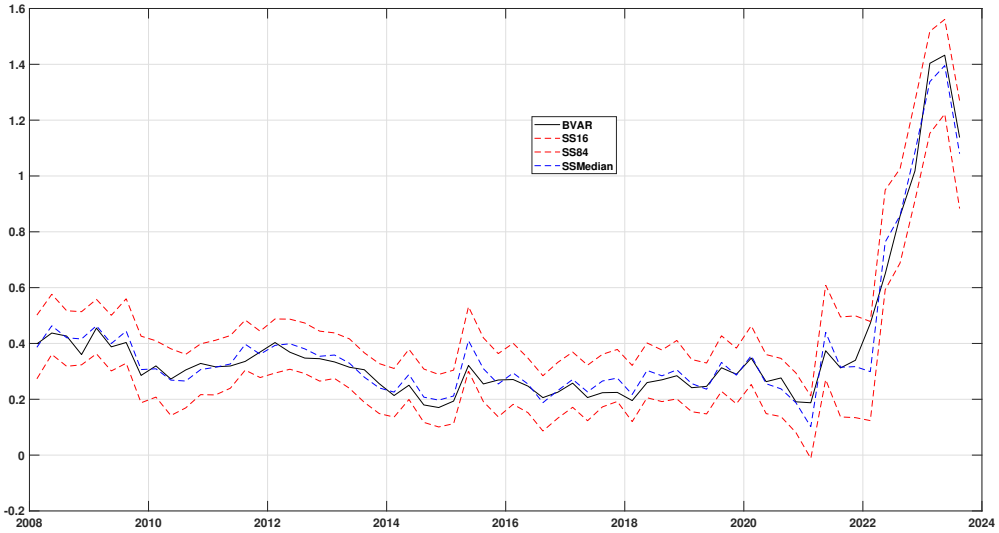
$$\beta_i \mid \sigma^2, \gamma^2, q \sim \text{i.i.d.} \begin{cases} \mathcal{N}(0, \sigma^2\gamma^2) & \text{with probability } q \\ 0 & \text{with probability } 1-q \end{cases}$$

where γ^2 is the degree of shrinkage and q is a measure of model size. As in Giannone et al. (2021) these two crucial parameters are treated as random variables and we conduct posterior inference on them. We specify uniform priors for q and for the R^2 of the model which uniquely identifies also the prior for γ^2.

Figure 6 plots the forecasts we obtained in our BVAR for the quarter-on-quarter inflation rate at the horizon of one quarter ahead (blue solid line), together with the 16^{th} to 84^{th} quantiles (red dashed lines) and the median of the corresponding out-of-sample forecasts (derived by adopting the same recursive scheme as in the VAR) from the regression with spike-and-slab priors (black dashed line).

The BVAR forecasts basically coincide with the median of the spike-and-slab based forecasts, suggesting that focusing on the four most relevant principal components does not lead to any major information loss.

Figure 6: Comparison of bivariate VAR with PCA and spike-and-slab regression



Note: 2008Q1-2023Q3

References

Ahrens, M., D. Erdemlioglu, M. McMahon, C. J. Neely, and X. Yang (2023): "Mind Your Language: Market Responses to Central Bank Speeches," *Available at SSRN 4471242*.

Ahrens, M. and M. McMahon (2021): "Extracting Economic Signals from Central Bank Speeches," in *Proceedings of the Third Workshop on Economics and Natural Language Processing*, ed. by U. Hahn, V. Hoste, and A. Stent, Punta Cana, Dominican Republic: Association for Computational Linguistics, 93–114.

Angelico, C., J. Marcucci, M. Miccoli, and F. Quarta (2022): "Can we measure inflation expectations using Twitter?" *Journal of Econometrics*, 228, 259–277.

Araujo, D. K. G. (2023): "Facilitating the use of machine learning by central banks with transfer learning," in *IFC Bulletin No. 57*, Bank for International Settlements, Box 1.

Araujo, D. K. G., G. Bruno, J. Marcucci, R. Schmidt, and B. Tissot (2023): "Machine learning applications in central banking," *Journal of AI, Robotics & Workplace Automation*, 2, 271–293.

Ash, E. and S. Hansen (2023): "Text algorithms in economics," *Annual Review of Economics*, 15, 659–688.

Atkeson, A. and L. E. Ohanian (2001): "Are Phillips curves useful for forecasting inflation?" *Quarterly Review*, 25, 2–11.

Banbura, M., D. Giannone, and L. Reichlin (2010): "Large Bayesian vector auto regressions," *Journal of Applied Econometrics*, 25, 71–92.

Banbura, M., M. Lenza, and J. Paredes (2024): "Chapter 9: Forecasting inflation in the US and in the euro area," in *Handbook of Research Methods and Applications in Macroeconomic Forecasting*, ed. by M. P. Clements and A. B. Galvao, Edward Elgar Publishing, 218–245.

Barbaglia, L., S. Consoli, and S. Manzan (2024): "Forecasting GDP in Europe with textual data," *Journal of Applied Econometrics*, 39, 338–355.

Battaglia, L., S. Hansen, T. Christensen, and S. Sacher (2024): "Inference for Regression with Variables Generated from Unstructured Data," unpublished manuscript.

Bańbura, M., D. Leiva-Leon, and J.-O. Menz (2021): "Do inflation expectations improve model-based inflation forecasts?" Working Paper Series 2604, European Central Bank.

Bengio, Y., R. Ducharme, and P. Vincent (2000): "A neural probabilistic language model," *Advances in neural information processing systems*, 13.

Blei, D. M., A. Y. Ng, and M. I. Jordan (2003): "Latent dirichlet allocation," *Journal of machine Learning research*, 3, 993–1022.

Blinder, A. S., M. Ehrmann, M. Fratzscher, J. De Haan, and D.-J. Jansen (2008): "Central bank communication and monetary policy: A survey of theory and evidence," *Journal of economic literature*, 46, 910–945.

Bulíř, A., M. Čihák, and D.-J. Jansen (2013): "What drives clarity of central bank communication about inflation?" *Open Economies Review*, 24, 125–145.

Carriero, A., D. Pettenuzzo, and S. Shekhar (2024): "Macroeconomic Forecasting with Large Language Models," *arXiv preprint arXiv:2407.00890*.

Cour-Thimann, P. and A. Jung (2021): "Interest-rate setting and communication at the ECB in its first twenty years," *European Journal of Political Economy*, 70, 102039.

de Bandt, O., J.-C. Bricongne, J. Denes, A. Dhenin, A. D. Gaye, and P.-A. Robert (2023): "Using the Press to Construct a New Indicator of Inflation Perceptions in France," Working Paper 921, Banque de France.

De Mol, C., D. Giannone, and L. Reichlin (2008): "Forecasting using a large number of predictors: Is Bayesian shrinkage a valid alternative to principal components?" *Journal of Econometrics*, 146, 318–328.

Dell, M. (2024): "Deep Learning for Economists," Working Paper 32768, National Bureau of Economic Research.

Devlin, J., M. Chang, K. Lee, and K. Toutanova (2018): "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," *CoRR*, abs/1810.04805.

Ehrmann, M., G. Ferrucci, M. Lenza, and D. O'Brien (2018): "Measures of underlying inflation for the euro area," *Economic Bulletin Articles*, 4.

Ehrmann, M. and J. Talmi (2020): "Starting from a blank page? Semantic similarity in central bank communication and market volatility," *Journal of Monetary Economics*, 111, 48–62.

Ehrmann, M. and A. Wabitsch (2022): "Central bank communication with non-experts–A road to nowhere?" *Journal of Monetary Economics*, 127, 69–85.

Faust, J. and J. H. Wright (2013): "Chapter 1 - Forecasting Inflation," in *Handbook of Economic Forecasting*, ed. by G. Elliott and A. Timmermann, Elsevier, vol. 2 of *Handbook of Economic Forecasting*, 2–56.

Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2000): "The Generalized Dynamic-Factor Model: Identification And Estimation," *The Review of Economics and Statistics*, 82, 540–554.

Fraccaroli, N., A. Giovannini, J.-F. Jamet, and E. Persson (2022): "Ideology and monetary policy. The role of political parties' stances in the European Central Bank's parliamentary hearings," *European Journal of Political Economy*, 74, 102207.

Gabaix, X., R. S. Koijen, R. Richmond, and M. Yogo (2023): "Asset embeddings," *Available at SSRN 4507511*.

Gardner, B., C. Scotti, and C. Vega (2022): "Words speak as loudly as actions: Central bank communication and the response of equity prices to macroeconomic announcements," *Journal of Econometrics*, 231, 387–409.

Gáti, L. and A. Handlan (2022): "Monetary communication rules," Tech. rep., ECB Working Paper.

Gentzkow, M., B. Kelly, and M. Taddy (2019): "Text as data," *Journal of Economic Literature*, 57, 535–574.

Giannone, D., M. Lenza, and G. E. Primiceri (2015): "Prior Selection for Vector Autoregressions," *The Review of Economics and Statistics*, 97, 436–451.

——— (2021): "Economic Predictions With Big Data: The Illusion of Sparsity," *Econometrica*, 89, 2409–2437.

Giavazzi, F., F. Iglhaut, G. Lemoli, and G. Rubera (2020): "Terrorist Attacks, Cultural Incidents and the Vote for Radical Parties: Analyzing Text from Twitter," NBER Working Papers 26825, National Bureau of Economic Research, Inc.

Hansen, S., M. McMahon, and M. Tong (2019): "The long-run information effect of central bank communication," *Journal of Monetary Economics*, 108, 185–202.

Hansen, S., T. Ramdas, R. Sadun, and J. Fuller (2021): "The demand for executive skills," Tech. rep., National Bureau of Economic Research.

Hirschbühl, D., L. Onorante, and L. Saiz (2021): "Using machine learning and big data to analyse the business cycle," *Economic Bulletin Articles*, 5.

Jansen, D.-J. and J. De Haan (2007): "The importance of being vigilant: Has ECB communication influenced Euro area inflation expectations?" Tech. Rep. 2134.

Kalamara, E., A. Turrell, C. Redl, G. Kapetanios, and S. Kapadia (2022): "Making text count: Economic forecasting using newspaper text," *Journal of Applied Econometrics*, 37, 896–919.

Kogan, L., D. Papanikolaou, L. D. Schmidt, and B. Seegmiller (2021): *Technology-skill complementarity and labor displacement: Evidence from linking two centuries of patents with occupations*, National Bureau of Economic Research.

Larsen, V. H. and L. A. Thorsrud (2019): "The value of news for economic developments," *Journal of Econometrics*, 210, 203–218, annals Issue in Honor of John Geweke "Complexity and Big Data in Economics and Finance: Recent Developments from a Bayesian Perspective".

Larsen, V. H., L. A. Thorsrud, and J. Zhulanova (2021): "News-driven inflation expectations and information rigidities," *Journal of Monetary Economics*, 117, 507–520.

Litterman, R. B. (1979): "Techniques of forecasting using vector autoregressions," Tech. rep.

Marcucci, J. (2024): "Macroeconomic forecasting with text-based data," in *Handbook of Research Methods and Applications in Macroeconomic Forecasting*, ed. by M. P. Clements and A. B. GalvÃ£o, Edward Elgar Publishing, Chapters, chap. 16, 425–468.

Mikolov, T., K. Chen, G. Corrado, and J. Dean (2013): "Efficient estimation of word representations in vector space," *arXiv preprint arXiv:1301.3781*.

Mitchell, T. J. and J. J. Beauchamp (1988): "Bayesian variable selection in linear regression," *Journal of the american statistical association*, 83, 1023–1032.

Muennighoff, N., N. Tazi, L. Magne, and N. Reimers (2023): "MTEB: Massive Text Embedding Benchmark," .

Mumtaz, H., J. Saleheen, and R. Spitznagel (2023): "Keep it simple: Central bank communication and asset prices," Tech. rep., Working Paper.

Pavelkova, A. (2022): "The ECB press conference: a textual analysis," .

Rosa, C. and G. Verga (2007): "On the consistency and effectiveness of central bank communication: Evidence from the ECB," *European Journal of Political Economy*, 23, 146–175.

Russell, S. J. and P. Norvig (2022): "Artificial intelligence: A modern approach, global edition," *Harlow: Pearson*.

Stock, J. H. and M. W. Watson (1998): "Diffusion indexes," NBER WP 6702, National Bureau of Economic Research.

——— (2007): "Why has US inflation become harder to forecast?" *Journal of Money, Credit and banking*, 39, 3–33.

Tibshirani, R. (1996): "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 58, 267–288.

Tobback, E., S. Nardelli, and D. Martens (2017): "Between hawks and doves: measuring central bank communication," Tech. rep., ECB Working paper.

Van der Maaten, L. and G. Hinton (2008): "Visualizing data using t-SNE." *Journal of machine learning research*, 9.

Vaswani, A., N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. u. Kaiser, and I. Polosukhin (2017): "Attention is All you Need," in *Advances in Neural Information Processing Systems*, ed. by I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Curran Associates, Inc., vol. 30.

Zheng, T., X. Fan, W. Jin, and K. Fang (2024): "Words or numbers? Macroeconomic nowcasting with textual and macroeconomic data," *International Journal of Forecasting*, 40, 746–761.

Acknowledgements

The authors thank Ben Cohen, Michael Ehrmann, Marek Jarocinski, Sujit Kapadia, Hanno Kase, Juri Marcucci, Joan Paredes, Fernando Pérez-Cruz and several seminar participants for comments. Johannes Damp provided excellent research assistance in the first phase of this project.
The opinions in this paper are those of the authors and do not necessarily reflect the views of the European Central Bank, the Eurosystem, the CEPR, the BIS or the ESM.

Douglas Araujo
Bank for International Settlements, Basel, Switzerland; email: douglas.araujo@bis.org

Nikola Bokan
European Central Bank, Frankfurt am Main, Germany; email: nikola.bokan@ecb.europa.eu

Fabio Alberto Comazzi
European Stability Mechanism, Luxembourg, Luxembourg; email: F.Comazzi@esm.europa.eu

Michele Lenza
European Central Bank, Frankfurt am Main, Germany; Centre for Economic Policy Research, London, United Kingdom; email: michele.lenza@ecb.europa.eu

Postal address 60640 Frankfurt am Main, Germany
Telephone +49 69 1344 0
Website www.ecb.europa.eu

PDF ISBN 978-92-899-7216-1 ISSN 1725-2806 doi:10.2866/6638874 QB-01-25-099-EN-N