Nowcasting the euro area with social interaction

Konstantin Boss[1], Luigi Longo[1], and Luca Onorante[1]

[1]JRC - European Commission

2026-05-06

Abstract

Social interaction matters. We show how discussions among social media users can be used to improve our understanding of inflation and unemployment dynamics in Europe. We extract forward-looking and context-sensitive sentiment from millions of Reddit posts and, crucially, from their associated comments threads and user feedback. Our empirical results show that taking discussions into account delivers consistent gains in out-of-sample accuracy relative to indicators built without consideration for social interaction, as well as relative to daily newspaper sentiment and financial variables. Moreover, we find that Reddit-based signals which incorporate comments significantly enhance explanatory power for business and consumer expectations. Our results suggest that the interaction among social media users can describe general expectations and sentiment better than indicators without social interaction. Overall, our paper highlights the role of online social discussion in selecting appropriate economic information.

JEL: E31, C32, C53, C55

Keywords: Social Media, Nowcasting, Natural Language Processing, Sentiment Analysis, Big Data, Large Language Models, Online Social Interaction.

[1]At this link, we provide a live dashboard with updated real-time results based on this paper. We thank our consultants, Florian Huber and Michael Pfarrhofer for valuable support. We also thank Gary Koop, Niko Hauzenberger, Luigi Gifuni, Luca Barbaglia, participants of the Conference ECONDAT 2025 spring meeting: "Economics with nontraditional data and analytical tools", and 13th ECB Conference on Forecasting Techniques: "Artificial intelligence in the analysis of economic narratives, forecasting, and risk assessment", for useful input. The views expressed herein are those of the authors and do not necessarily represent the views of the European Commission. All remaining errors are our own. Corresponding author: Luigi Longo, Via E. Fermi, 2749, 21027 Ispra (VA). Email: luigi.longo@ec.europa.eu.

1 Introduction

In Europe, the share of citizens who use social media as their primary source of news and information has jumped by 11 percentage points from 2022 to 2024, making it the fastest-growing category of news sources, particularly among young Europeans (European Commission, 2025). A recent Eurobarometer shows that more than two-thirds of users actively seek out social and political information on social media and over a third engage with the content through commenting or liking (European Parliament, 2025). Researchers have started to make increasing use of the vast amounts of unstructured data that are thus generated on social media, to study, for example, house prices (Bailey et al., 2018), central bank communication (Ehrmann and Wabitsch, 2022), or climate change Arteaga Garavito et al. (2025).

The main contribution of this paper is the use of social network data from Reddit to assess whether information extracted from peer-to-peer social interactions is more informative than top-down communications, such as newspaper articles. We focus on social media discussions around the recent history of inflation and labour market conditions in Europe.

To do this, we collect millions of posts - together with the related comments - from r/europe, the largest Reddit community focusing on European economic and political matters. With the help of state-of-the-art large language models (LLMs), we extract forward-looking and context-sensitive time series signals at daily frequency. Notably, we treat comments as a source of auxiliary information and use them to "regularize" the signal extracted from each initial post. We first assign a signal to the submission, which is often a piece of news, and subsequently to all associated comments. We then update the submission's initial signal by combining it with the average signal obtained from its associated comments. This procedure allows us to integrate the information contained in both submissions and comments into a single, unified measure.

The empirical evaluation of the indicators' quality is carried out in two steps. First, we set up an out-of-sample application to empirically show that the information content of the Reddit signals significantly improves nowcasting performance for inflation and unemployment in the euro area over benchmark models, particularly when discussions are taken into consideration. We also compare the Reddit series against other daily indicators obtained from financial markets or newspapers and show the superiority of the former. As a second validation, we link the Reddit-based series to surveys of expectations. In this application we show that Reddit sentiment is significantly correlated with business and consumer expectations, even when controlling for optimally-selected macro-financial variables. Again, signals which include social interaction through comments have the largest explanatory power.

Our results indicate that, the inclusion of the discussion is essential. This happens because in many cases the initial submission consists of a piece of *information*, for example taken from the news. The ensuing discussion, on the other hand, represents *information* *processing*, which translates more readily to beliefs and actions. We find that often, online discussants disagree with the messages transmitted by the original posts and overturn the submission's signal, thus acting as an important regularization device. These benefits from incorporating the peer-to-peer discussions into the signals are larger for unemployment than for inflation. In the case of inflation, initial posts trigger less discussion in relative terms, as inflation-related threads are mostly limited to posting news selected by the users on the dedicated community. For unemployment-related threads, there is a more lively discussion around each post, and the sentiment score often changes when comments are included. As a result, considering social interaction notably improves the performance of unemployment-related signals for both nowcasting and linking the signals with expectations.

Relation to the literature: We contribute to the fast-growing literature that uses social media data to construct signals for the macro economy. A widely used data source for this has been X/Twitter. Antenucci et al. (2014) use it to create a US job market tracker. Gorodnichenko et al. (2024) use data from the platform to construct measures of US inflation expectations and to study how Federal Reserve communication is perceived by users. Macaulay and Song (2023) study how different X/Twitter narratives feed through to sentiment and how this affects economic cycles. Heikkinen and Heimonen (2025) use data from Refinitiv to understand how social media tone affects inflation expectations as measured by the Michigan Survey of Consumers and the Survey of Professional Forecasters (SPF). Evidence on the usefulness of Reddit data for inflation forecasting in the US is provided in Del Monaco et al. (2025). Research output for Europe is significantly sparser. For example, Angelico et al. (2022) use X/Twitter data for inflation expectations in Italy and Benatti et al. (2020) leverage LinkedIn data to track European unemployment. In a contribution closely related to ours, Born et al. (2024) use Tweets from Germany in combination with artificial intelligence (AI) to construct an index of inflation expectations. We contribute to this literature by considering Reddit data for Europe, extending the comparison to unemployment, and showing how to add social interaction to signal construction. Furthermore, we compare our social media indicators to other high frequency measures that have been considered for monitoring in Europe such as newspaper articles (Barbaglia et al., 2024), or financial market indicators (Aliaj et al., 2023).

On the methodological side, the literature has recently started to evolve away from traditional natural language processing (NLP) towards generative AI. Early NLP methodologies focused on structured extraction and classification of text patterns. For example, Baker et al. (2016) use NLP to construct the Economic Policy Uncertainty (EPU) index from newspaper data. More recently, Barbaglia et al. (2024) forecast European GDP with newspaper sentiment extracted with a dedicated economic dictionary. On the side of AI tools, Carriero et al. (2024) apply LLMs directly to forecast US time series data, while Faria-e Castro and Leibovici (2024) and Hansen et al. (2024) show how prompting enables LLMs to act as forecasters, mimicking survey expectations without structured inputs. Similar to our paper, Bybee (2023) use LLMs to extract sentiment from US newspaper articles, while Gueta et al. (2025) apply them to social media data, demonstrating their ability to go beyond simple pattern recognition. We add to this branch of research by demonstrating that LLMs perform better than dedicated dictionaries on the complex classification task posed by social media data, since they can easily handle colloquial language and are more suitable to understand whether text is forward looking or not.

The remainder of the paper is structured as follows: section 2 explains the raw data extraction and transformation into time series signals using the LLM; section 3 presents the setup of the out-of-sample evaluation and discusses the results in detail; section 4 shows the link between Reddit series and business and consumers expectations; section 5 concludes. In the Appendix we add details and further results on the data construction and evaluation steps. Additionally we provide a description of a live and interactive dashboard that we update on a monthly basis with social media data and real-time forecasts of macroeconomic variables. The dashboard can be found at inflation-nowcast-socialplatform.

2 Data and methodology

2.1 Collecting social media data

We use data from Reddit, a social media platform composed of communities, known as "subreddits", dedicated to specific topics such as politics, technology, and economics. Users can share news items, opinions, and experiences related to economic trends, market behavior, or technological developments, thus producing a rich and up-to-date data collection. A key difference between Reddit and traditional media outlets is the fact that users on Reddit make an active selection of the posts they submit, thus acting as an important filter of the available wealth of information. Moreover, users can engage with the posted content through commenting and voting, which provides an additional layer of information, not typically found in print media or news broadcasts. Subreddits come with a set of rules that are curated by moderators who can, for example, remove off-topic or incoherent submissions and hate speech.

To manage the different levels of information that Reddit has to offer, we proceed in steps, as shown in Figure 2.1. First, we select the community of interest for our study, the subreddit *r/europe*, which is the largest community discussing European-related news and economic developments. Second, within the selected community, we identify (1) submissions and (2) comments attached to them. The first row of Table 2.1 shows the total number of comments and submissions recorded for the time period between 2012 and 2023. Given the very large number of items, we use a keyword-filtering approach to pre-select the submissions that are relevant for the macroeconomic concepts inflation and unemployment.[1] This yields 4,825 inflation-related and 1,934 unemployment-related submissions to be analyzed. Each submission can trigger a subsequent discussion, which spans on average more than 30 comments (see total numbers in Table 2.1). This highlights the substantial additional information that comments provide on top of the pure submissions.

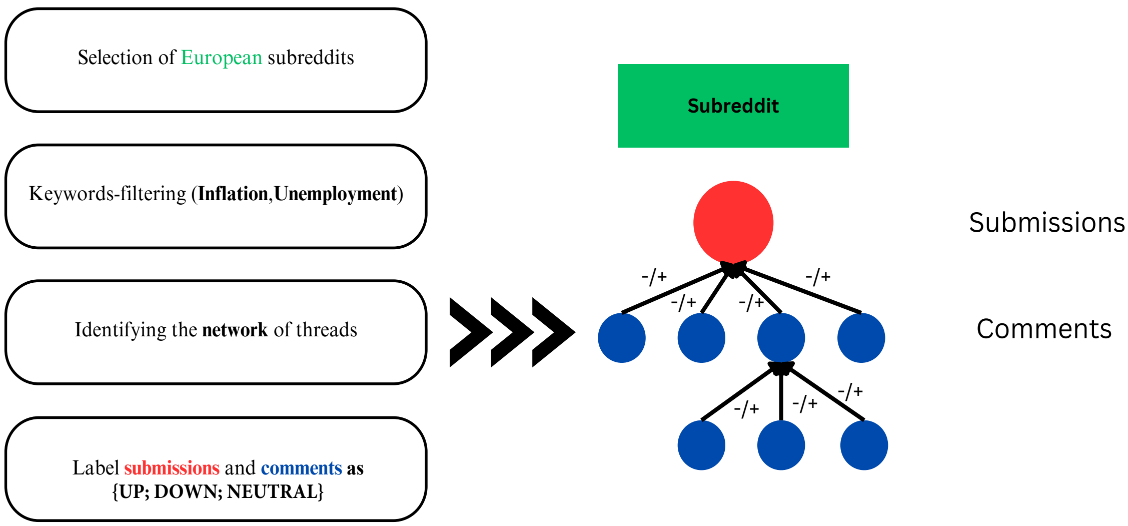


Figure 2.1: To extract signals from Reddit data, a four-step approach is employed. First, discussions are sourced from the subreddit *r/europe*. Second, keyword filtering is applied to identify submissions related to inflation and/or unemployment. Third, comments on each submission are stored. They can be made directly on the submission or on previous comments. We also record the difference between upvotes and downvotes on each comment as a possible weighting device. Fourth, submissions and comments are scored as either signaling inflation/unemployment going UP, DOWN, or NEUTRAL by the LLM.

[1]Submissions should contain at least one of the following keywords to be selected: **inflation-related**: inflation, deflation, hyperinflation, price; **unemployment-related**: unemployment, employment, unemployed, job. We notice that changing the set of keywords used to filter does not significantly alter the results. Indeed, the most important selecting words turn out to be the economic concept we wish to target, i.e. *inflation* or *unemployment*.

Third, for the associated comments we make a distinction between those that immediately follow the original submission and the full collection of comments. While the direct comments are likely to be the most relevant ones for the topic raised in the submission, additional layers of comments may still contain valuable information. To make this distinction, we recover the full network structure of comments following an initial submission. We call the comments that are directly attached to the submission *first-level*. The set of comments across all levels, on the other hand, is very large. Therefore, we make use of the keyword filtering approach discussed above to subset it again. This collection is labeled *keyword*. The final numbers of first level and keyword-filtered comments are summarized in Table 2.1. In addition, each comment and submission can be upvoted and downvoted by the users and we record the net scores for all posts as the difference between upvotes and downvotes. Following these steps, the selected submissions and comments are fed to the LLM, to extract a signal that we use for evaluation. We explain this step in detail below.

Summary statistics for r/europe

	Submissions	Comments (*first level*)	Comments (*keyword*)
r/europe	759,179	Total: 25,088,378	
Inflation	4,825	31,938	24,932
Unemployment	1,934	14,464	12,349

Table 2.1: Summary statistics for submissions and comments for each concept in the subreddit *r/europe*. The time range is 2012m1-2023m12.

2.2 Extracting indicators with the LLM

We give the job of assigning a sentiment classification to each Reddit post to an LLM. Due to their generality, these models are particularly well-suited for handling the challenges posed by the linguistic characteristics of Reddit communities, as noted in Long et al. (2023). Unlike the language found on conventional news outlets, Reddit users often write in community-specific slang, use abbreviations, and generally informal language that strongly deviates from standard dictionaries commonly found in economic research, such as the seminal one of Loughran and McDonald (2011). By using an LLM, we can capture the nuanced, community-specific signals present in Reddit data without the need of extensive fine-tuning or adjustment of existing dictionaries. We use the model *LLaMa-3-70b-instruct* on local servers of the European Commission. For the present research, we instruct the model to classify each submission and comment into one of three categories (UP, DOWN, NEUTRAL) based on the forward-looking signal contained in each of the provided items. The exact prompt is reported below:

```
You are a forecaster and want to predict the future {concept} in Europe
from textual documents.  The following document will report sentences
potentially referring to Europe:  df[‘title’][i].  You have to print
a signal that can be UP (if the document is signalling {concept} going
up in the short/long-term run), DOWN (if the sentence is signalling
{concept} going down in the short/long-term run) or NEUTRAL (if the
sentence is neutral or does not signal a particular direction on
{concept}).  Print only the results of the signal, do not summarize the
sentence nor give any reason on your choice.  Even if more sentences or
paragraphs are provided to you, you only have to print one signal that
can be UP, DOWN or NEUTRAL.
```

Figure 2.2: LLM prompt used with LLaMa-3-70b-instruct. *concept* is either unemployment rate or inflation rate, `df['title'][i]` contains the submission or comment.

2.3 Assessing the accuracy of the LLM

To understand whether the LLM does a good job at classifying posts, we adopt two dictionary-based approaches as benchmarks to compute the three-outcome signal taking into account positive and negative words appearing in each Reddit submission. In Appendix A.1 we provide the dictionary based on the work of Granziera et al. (2025). The second dictionary of Consoli et al. (2022) is larger and therefore not reported in the Appendix. Both works have been shown to be successful in extracting meaningful signals from economic texts such as Fed speeches (Granziera et al., 2025) and newspapers Barbaglia et al. (2024).

We calculate the accuracy of the LLM and the dictionaries in terms of the F1 score. Since our classification yields three possible outcomes for each piece of text, the F1 score is computed as the unweighted average of the F1 score for each of the three categories. The F1 score per category is computed as

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{1}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{2}$$

$$\text{F1} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{3}$$

where TP is the category's true positive rate, FP is the false positive rate and FN is the false negative rate. The score takes values between 0 and 1 and larger values imply better performance. Evidently, there is no objective ground truth for the sentiment attached to Reddit submissions or comments that could be used to compute and assess error statistics. Instead, we rely on a human-labeling approach to gauge the accuracy of the models. We randomly select a subset of 483 inflation and 194 unemployment related submissions (10% of the total number of submissions for the two targets) which we manually label using the UP, DOWN, NEUTRAL scale. We ensure that the time coverage over the selected submissions is even and not concentrated in a few specific years. We then compute the F1 scores for the LLM and the two dictionaries explained above. The dictionary approach of Granziera et al. (2025) achieves a score of 0.340 for inflation and 0.333 for unemployment. The Consoli et al. (2022) dictionary approach reaches 0.290 for inflation and 0.339 for unemployment. The results for the LLM are visualized in Figure 2.3, which shows that the generative model significantly outperforms the scores of the dictionary approaches for both variables. Unemployment scores are slightly higher than those for inflation. Overall, the scores lie around 0.71 (inflation) and 0.75 (unemployment).

A common concern around the use of LLMs is replicability, as different models can produce different outcomes even when prompted the same way several times. To assess whether this may be an issue in our application, we vary the temperature parameter of the LLM on a grid between 0.1 and 0.9 in steps of 0.2. This nonnegative parameter controls the "creativity" of LLM responses. Higher values typically produce more volatile responses and 0 produces deterministic answers. For each temperature, the LLM is asked to score the same 483/194 submissions 100 times to obtain a measure of variability for each configuration. In nearly all cases, we observe that increasing the temperature widens the range of possible accuracy scores, but the median stays essentially stable. We use a temperature of 0.5 throughout the following analysis.

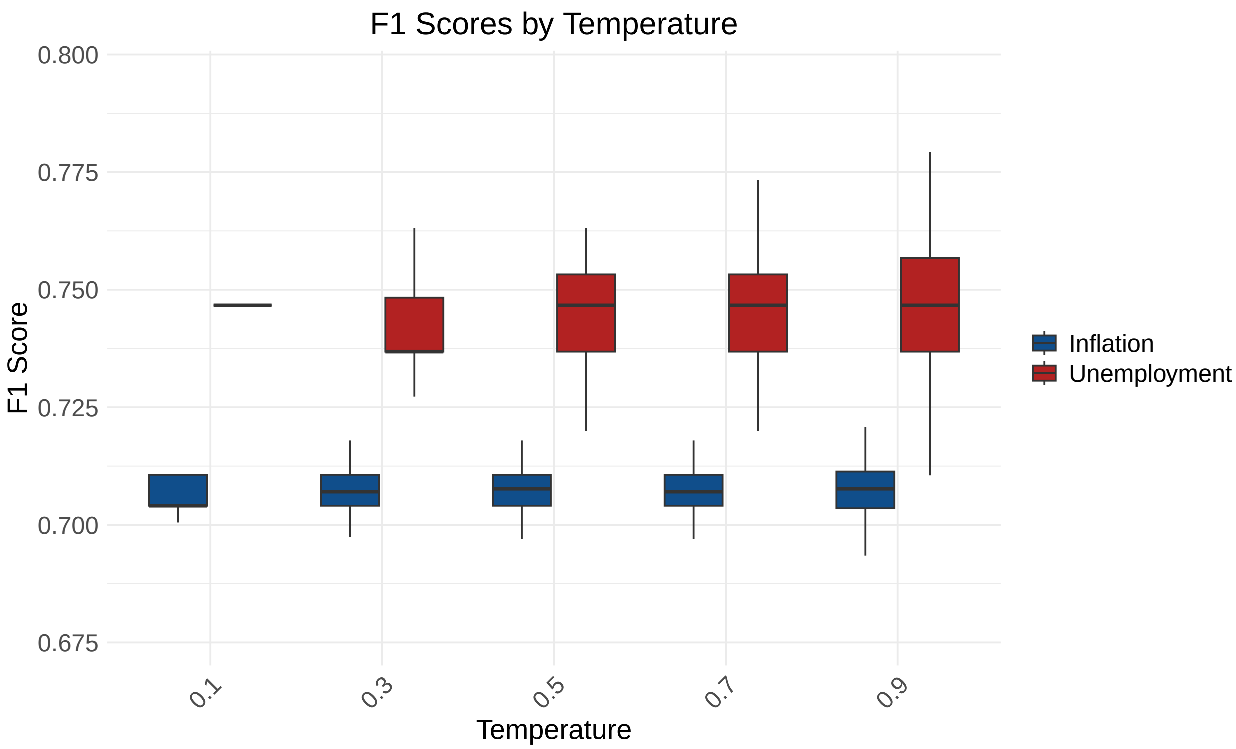


Figure 2.3: LLM accuracy F1 scores for inflation and unemployment submissions. The model is run 100 times for each temperature value. Boxes represent the interquartile range (Q25 and Q75). Whiskers extend to a 95% interval for the median (thick horizontal line).

2.4 Constructing time series indicators

Building on the classifications obtained via the LLM, we construct the time series indicators X_t for submissions only and $\bar{X}_t$ for signals, which combine both submissions and comments. We begin with the description of the submissions-only signal, that is constructed by aggregating all signals $S_i \in \{\text{UP, DOWN, NEUTRAL}\}$ for submissions posted on day t, computing the daily sum, where UP takes value 1, DOWN takes value -1, and NEUTRAL is treated as 0:

$$X_t = \sum_{i=1}^{N_t} S_i \tag{4}$$

N_t represents the total number of submissions recorded on a given day. If no submissions are available in t, the indicator is assigned a value of zero. X_t is then smoothed using a backward looking moving average (MA) filter of varying window sizes to test the importance of noise reduction.

Next, we turn to the construction of time series signals $\bar{X}_t$, which also incorporate social interaction, a key novelty of this paper. The steps are illustrated in Figure 2.4.

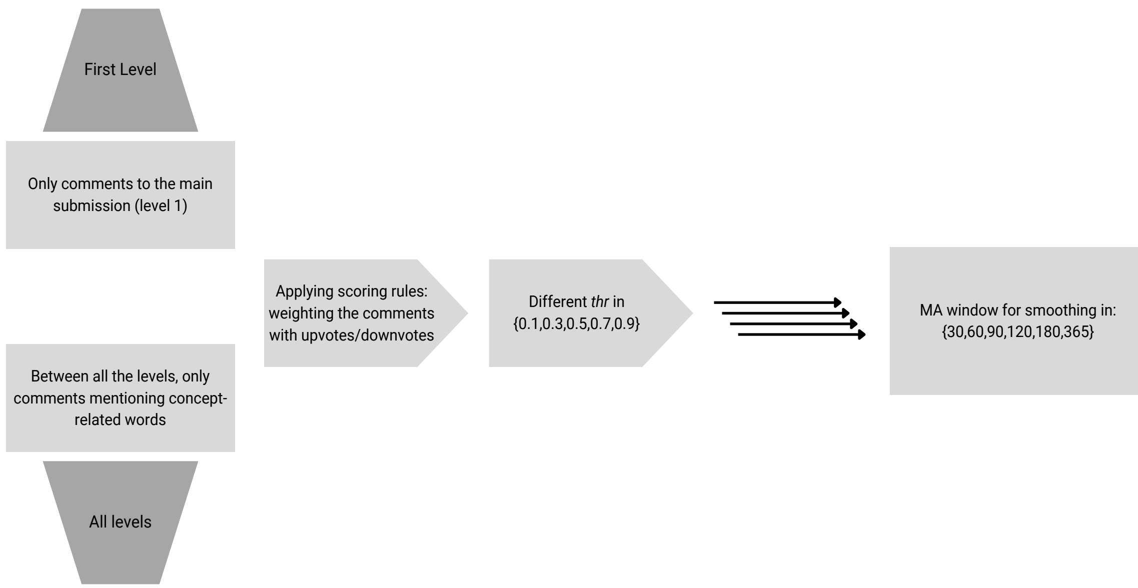


Figure 2.4: 2 sets of comments a) all first level comments; b) considering all the levels, only keeping comments containing keywords related to the concept. To both types of comments the following additional transformation are applied: i) different scoring rules for aggregation based on up- downvotes, ii) different thresholds, iii) transformation with MA smoothing windows.

For each submission i we compute a new score $L_i \in [-1, 1]$, which combines not only the submission's original signal (S_i), but also the signals derived from the J comments $C_{i,j} \in \{\text{UP, DOWN, NEUTRAL}\}$ associated to it. If a submission receives no comments, the original signal remains unchanged.[2]

[2]A valid concern is that comments could arrive much later than the submission, invalidating their use for a real-time monitoring application. We checked the timing of all the comments and found that they are posted at most one week after the main submission. This ensures a significant reduction of any possible look-ahead bias coming from comments arriving at later points in time. The effects of information flow are discussed in detail in Appendix B.4.

$$L_i = \frac{S_i + \sum_{j=1}^{J} C_{i,j}}{J+1}. \tag{5}$$

Comments $C_{i,j}$ can either be those that are directly attached to the submission, but unfiltered by keywords (*first level*), or the set of all keyword-filtered comments (*keyword*), as explained in section 2.

We treat each comment as a vote. If the LLM assigns UP to the comment, the vote is counted as 1. If it assigns DOWN, the vote is -1 and 0 for NEUTRAL classifications. As noted in section 2, comments can also receive upvotes and downvotes by users. The net score of these can be used as a weight on the vote that each comment casts. Taking the average of the (possibly weighted) votes of all comments in addition to that of the original submission yields the score L_i in (5). Finally, we need a decision rule to transform the score L_i back to the UP, DOWN, NEUTRAL scale. We achieve this through the threshold parameter τ. This parameter is varied on a discrete grid, to evaluate its impact on the quality of the signal. Based on the voting outcome, the new label $\bar{S}_i$ for submission i is obtained:

$$\bar{S}_i = \begin{cases} \text{UP} & \text{if } L_i > \tau \\ \text{DOWN} & \text{if } L_i < -\tau \\ \text{NEUTRAL} & \text{if } -\tau \leq L_i \leq \tau \end{cases} \tag{6}$$

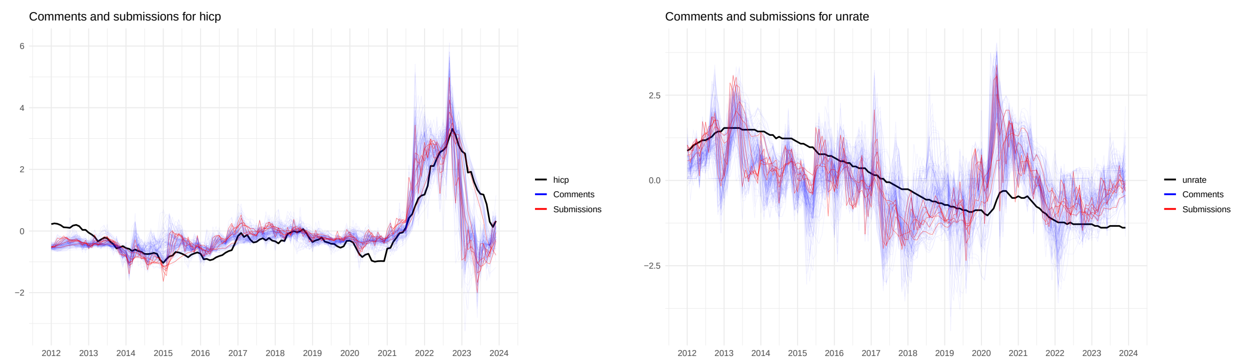
Once this process is finished and all new submission labels $\bar{S}_i$ are obtained, the daily scores are re-computed as $\bar{X}_t$ according to:

$$\bar{X}_t = \sum_{i=1}^{N_t} \bar{S}_i \tag{7}$$

The resulting signals $\bar{X}_t$ are also smoothed with an MA filter of different window lengths. Crucially, this voting scheme allows users to overturn the originally assigned score of the submission S_i, if they disagree with the original sentiment.

In sum, the time series involving not only submissions, but also social interaction through comments differ along four dimensions: (1) unfiltered first level or keyword-filtered comments, (2) usage or not of up- and downvotes, (3) thresholds used for voting outcomes, and (4) MA smoothing windows. At the end we have a total of 120 series for each of the two target variables, inflation and unemployment. Since Reddit is a largely unexplored data base in the context of academic economic analysis for Europe and social interaction has not been widely discussed either, we are deliberately exploratory in the construction of the comments signals. Our aim is to understand what filters and transformations have an impact on signal quality to be able to report approaches that are promising and approaches which are not. In the empirical section, we subject all of them to rigorous out-of-sample nowcasting exercises, designed to assess their informational content relative to standard benchmarks. Despite the multiple specifications, a consistent subset of series emerges as systematically outperforming the benchmarks.

The collection of time series from Reddit for inflation and unemployment are reported in Figure 2.5. Our Reddit signals track HICP inflation and the unemployment rate well and appear to anticipate some key turning points related to the COVID pandemic and the subsequent high inflation regime.



(a) HICP inflation

(b) Unemployment Rate

Figure 2.5: Time series of observed HICP inflation and unemployment data at monthly frequency against Reddit signals aggregated from daily to monthly frequency using averages. All series are z-scored purely for comparability in the plots. Z-scoring is not used in the evaluation steps reported below.

3 Out-of-sample evaluation

3.1 Setup

To assess the validity of our Reddit signals and, in particular, which forms of social interaction are useful for real-time monitoring of macroeconomic conditions in the euro area, we set up a nowcasting application. The target variables are year-on-year HICP inflation and its components for energy, services, food and core, as well as the overall unemployment rate plus those for people over and under the age of 25 years in the euro area. The target variables are obtained from the EA-MD data base created by Barigozzi and Lissona (2024), with the exception of food price inflation, which comes from the European Commission (Eurostat mnemonic teicp010). Since Reddit users tend to be relatively young non-experts, we expect our signals to be particularly useful for monitoring indicators whose components are more frequently observable by consumers such as food prices, energy prices (see Anesti et al. (2024) and references therein) or youth unemployment. The out-of-sample period ranges from January 2018 to December 2023, leaving an initial estimation sample from January 2012 to December 2017, which we expand recursively.

Our set of candidate indicators includes all 120 Reddit indicators whose construction is detailed in section 2.4, taken one by one. To this set of social media variables we add for comparison the daily newspaper sentiment indices computed by Barbaglia et al. (2024), inflation swaps for 1 to 5 years and the Brent oil price (FRED mnemonic DCOILBRENTEU), reflecting the fact that commodity prices have been shown to improve forecasting performance for inflation (Breitung and Roling, 2015). The newspaper sentiment indicators of Barbaglia et al. (2024) are available for Germany, France, Italy, and Spain. To obtain an aggregate index for inflation, unemployment and monetary policy sentiment, we take the first principal components of the available data across countries.

To combine monthly data with daily data, we estimate the following MIDAS-AR regression (Ghysels et al., 2016):

$$\alpha(L)y_t = c + \beta(L)\sum_{i=0}^{k} w_i x_{s(t)-i} + \epsilon_t, \quad \forall i, \quad w_i = h(\gamma, i) \text{ and } \sum_{i=0}^{k} w_i = 1 \tag{8}$$

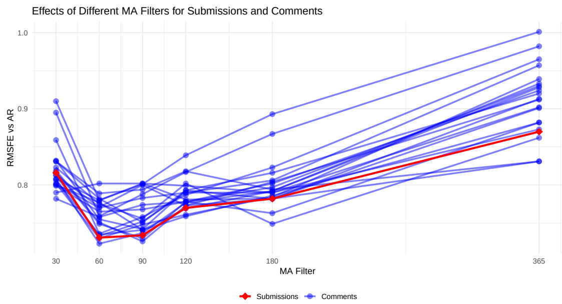
where we choose standard second-degree Almon polynomials for the weights:

$$w_i = \frac{\exp\left(\sum_{j=1}^{d} \gamma_j i^j\right)}{\sum_{l=0}^{k} \exp\left(\sum_{j=1}^{d} \gamma_j l^j\right)}, \quad d = 2 \tag{9}$$

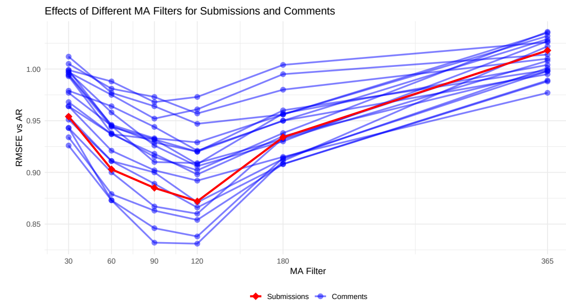
For each target-predictor combination, the lag order q in $\beta(L)$ for the daily data is determined based on the Akaike information criterion (AIC) each time a new nowcast is produced. The lag order p in $\alpha(L)$ is fixed at 1. We choose this simple setup because the emphasis of our exercise is on comparing information sets rather than models or optimizing nowcasts. Therefore, we fix the model and $\alpha(L)$ and vary only the variable x and its associated lag order to isolate information differences. We note that the exercise is not fully real-time, as we do not use data vintages. However, price indexes and unemployment rates are rarely subject to revisions. For Reddit data (and our other daily indicators) revisions are also negligible and arise only if certain submissions or comments are removed by the moderators or if users delete them themselves.

3.2 Results

We begin by discussing the differences between Reddit signals based only on submissions and signals which incorporate social interaction through comments. Figure 3.6 clearly shows that virtually all Reddit specifications improve upon the AR(1) benchmark as indicated by relative error statistics below 1. Importantly, for all target variables, there is a specification with Reddit comments which is preferable to one that neglects social interaction. Both results are uniform across (1) target variables and (2) smoothing windows. The benefit of social interaction is more pronounced for unemployment, as there are more comments specifications that beat pure submissions. Across smoothing windows, the relative out-of-sample scores are almost all hump-shaped, except for HICP core and HICP services (the main part of core inflation), which may even benefit from wider smoothing windows, mirroring the persistence of the data. In sum, while only using Reddit submissions already produces noticeable information improvements over a naive benchmark, adding social interaction is beneficial, especially for tracking unemployment.

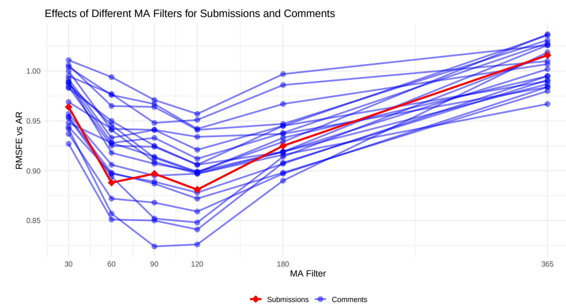


(a) HICP inflation

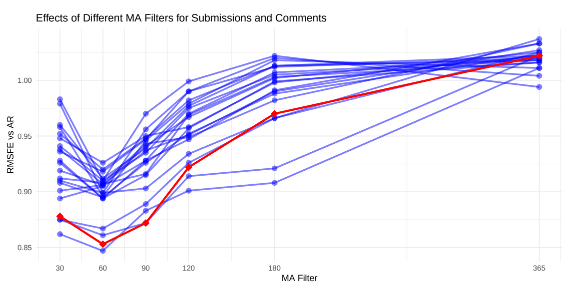


(b) Unemployment rate

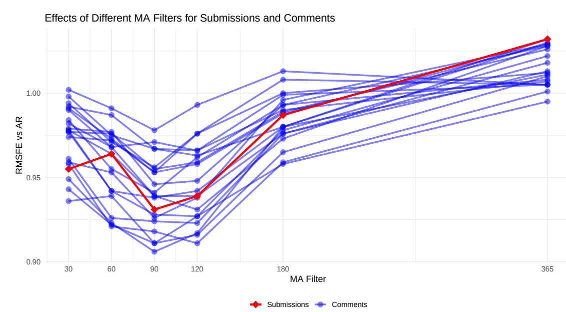
(c) HICP Core



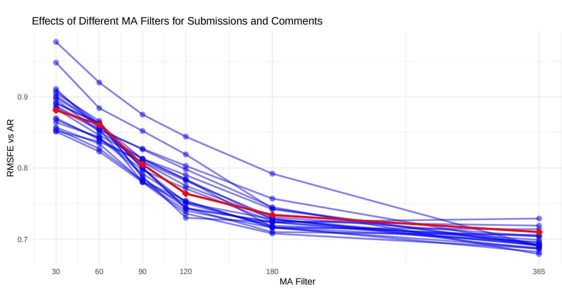
(d) Unemployment >25



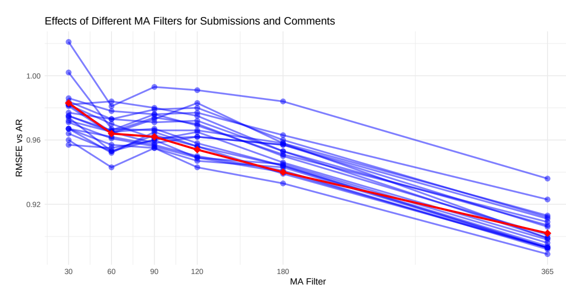
(e) HICP Energy



(f) Unemployment <25



(g) HICP Food



(h) HICP Services

Figure 3.6: RMSFE scores for Reddit signals against AR(1) benchmark compared for different sizes of the MA filter window in days.

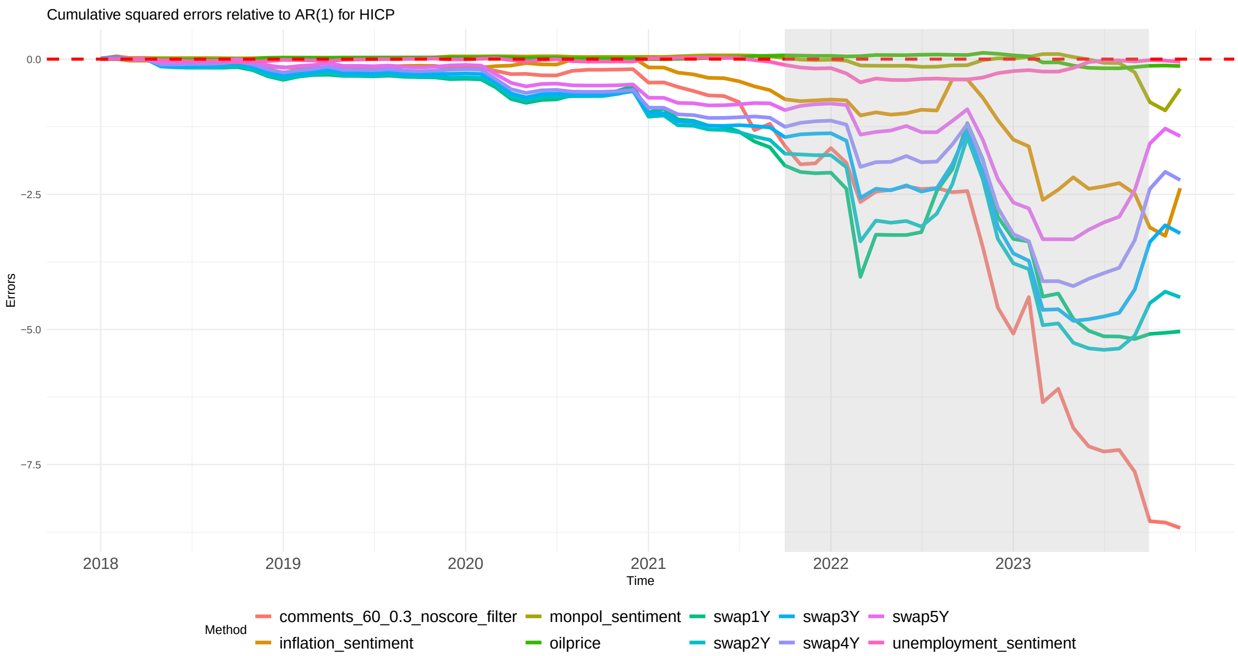
Next, we turn to the comparison of the Reddit series to the other daily indicators we have considered. Table 3.2 shows that the Reddit indicator generally outperforms the other variables in terms of point and density accuracy. The nowcasting gains using Reddit range from 13 percentage points (food price inflation) to at least 5 percentage points (youth unemployment) relative to the next best specification. It significantly beats the AR(1) benchmark in most cases and presents particularly strong performance gains for food price inflation and the overall HICP index. Gains for unemployment nowcasting are sizable but statistically significant only for the MAFE. Somewhat surprisingly, the nowcast improvements are slightly larger for the unemployment rate for people *over* the age of 25. Inflation-related newspaper sentiment and swaps also improve frequently over the AR(1), but to a smaller extent than the Reddit indicator. The sentiment index for unemployment proves a solid competitor as well. Models involving oil prices are on par with the AR(1). We note that the results for Reddit are quite encompassing, performing well even for difficult-to-predict series such as energy prices. Our positive results suggest that online discussions go beyond the simple oil price pass-through, possibly incorporating other related up-to-date information on geopolitical and economic events.

Out-of-sample results

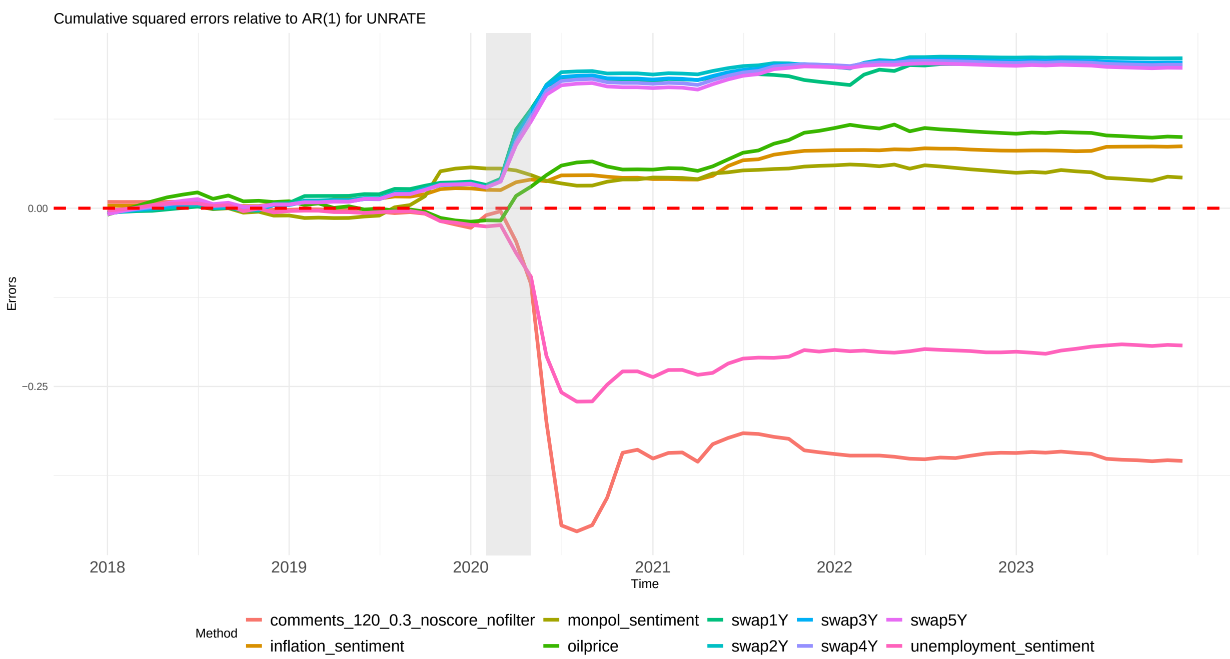
Target	*Metric*	*Best Reddit*	*Best Newspaper*	*Best Swap*	*Oil Price*
HICP		*com_60_0.3_0_1*	*inflation*	*swap1Y*	
	RMSFE	0.723**	0.932	0.85*	0.996
	MAFE	0.776**	0.952	0.866	1.000
	CRPS	0.751**	0.966	0.869	1.016**
HICP Core		*com_365_0.1_0_0*	*inflation*	*swap2Y*	
	RMSFE	0.844***	0.941*	0.918*	1.000
	MAFE	0.801***	0.923*	0.938	1.037***
	CRPS	0.815**	0.949	0.951	0.996
HICP Energy		*com_60_0.9_1_1*	*monetary policy*	*swap5Y*	
	RMSFE	0.847**	0.997	1.028	1.011
	MAFE	0.814***	0.998	0.982	1.003
	CRPS	0.832***	1.014	1.011	1.006
HICP Food		*com_365_0.7_1_1*	*inflation*	*swap1Y*	
	RMSFE	0.679***	0.809**	0.871**	1.022**
	MAFE	0.661***	0.825**	0.891*	0.999
	CRPS	0.662***	0.827**	0.89*	0.999
HICP Services		*com_365_0.1_0_0*	*inflation*	*swap2Y*	
	RMSFE	0.889*	0.978	0.924	1.014
	MAFE	0.922	0.987	0.925	1.043*
	CRPS	0.906	0.995	0.934	1.061
Unemployment		*com_90_0.3_0_0*	*unemployment*	*swap5Y*	
	RMSFE	0.832	0.912	1.028	1.041
	MAFE	0.856*	0.977	0.982	1.048*
	CRPS	0.875	0.945	1.011	1.033
Unemployment > 25		*com_90_0.3_0_0*	*unemployment*	*swap1Y*	
	RMSFE	0.824	0.888	1.092**	1.039
	MAFE	0.826**	0.902*	1.076*	1.024
	CRPS	0.854	0.897*	1.088**	0.996
Unemployment < 25		*com_90_0.3_0_0*	*unemployment*	*swap5Y*	
	RMSFE	0.906	0.962	1.040*	1.034
	MAFE	0.872**	0.957	1.016	1.000
	CRPS	0.909	0.974	1.060**	1.000

Table 3.2: Numbers below 1 imply that the benchmark monthly AR(1) is worse than the MIDAS model. Asterisks refer to significance levels 0.10 (*), 0.05 (**), and 0.01 (***) as calculated with the one-sided Diebold-Mariano test. "Best" indicators in each category are chosen on the basis of RMSFE. The best performing indicator per category and per target is in the first row of each target. The structure of the winning specification for the Best Reddit column is "com_MA_threshold_scoring_firstlevel'. E.g. "com_60_0.3_0_1" means that the best indicator uses comments, not pure submissions, an MA window of 60 days, a threshold of 0.3 to determine the outcome of the comments' inclusion, does not use up/down-votes and considers only the first level of comments, not the full comment structure after the submission.

Lastly, we check whether there are specific episodes in the out-of-sample period where the information benefits from using Reddit and social interaction are more pronounced. Figure 3.7 shows the cumulated squared errors for all competitors over time. In terms of HICP inflation, we see that the inflation swaps performed better than the Reddit indicators after the onset of the COVID-19 recession and through the first part of the high inflation period between 2021 and 2023. As the inflation rate started to drop, this picture reverses and the performance of the Reddit indicator improves markedly. This suggests that the sentiment among social media users reversed faster than on financial markets. For the case of the unemployment rate, we see a single significant improvement in the nowcasting performance after the COVID-19 recession for both the newspaper sentiment indicator and the Reddit signal. Otherwise performances are quite stable and the results seem to be driven by a few large shocks that were missed badly by the AR(1). These results appear in line with the literature which has argued that high frequency information is particularly useful in times of turmoil, but less so in tranquil times (Barbaglia et al., 2023).



(a) HICP inflation



(b) Unemployment rate

Figure 3.7: Cumulative squared errors of MIDAS models minus cumulative squared errors of AR(1) over time. Values below 0 mean better performance than the AR(1). Shaded areas are periods when HICP inflation was above 4% and the COVID-19 recession months, respectively. Among the Reddit indicators, we provide the cumulative loss for the best performing one according to their target (HICP/Unemployment rate), as shown in Table 3.2.

In summary, Reddit signals provide valuable information about price and labour market dynamics out-of-sample, improving over other indicators that have been proposed for the purpose of short term monitoring such as financial time series or newspaper sentiment. They are particularly useful when social interaction is added to the information set and in times of severe economic turmoil such as the recent hyperinflation period in Europe.

4 Reddit's connection with survey expectations

4.1 Setup

The previous section has shown that the Reddit-derived indicators contain daily information that is valuable for real-time monitoring of economic conditions. In this section, we examine their connection to survey expectations. Our approach is motivated by two key results of the literature: first, expectations are known to play an important role in nowcasting (Cascaldi-Garcia et al., 2024), policy transmission, and business-cycle analysis Coibion and Gorodnichenko (2025). Second, social media signals have been shown to capture relevant features of survey-based expectations (Daas and Puts, 2014; Angelico et al., 2022; Born et al., 2024; Gorodnichenko et al., 2024), but without substantial publication lags and at higher frequencies, providing policy-makers with timely information on current perceptions of the economy's state, at potentially lower costs than surveys.

Our working hypothesis is that signals with discussions are a better high-frequency proxy for expectations than signals based only on submissions. This is because discussions include an additional layer of information processing, beyond the news reporting that constitutes an important share of submissions. For this purpose, we use a regression framework to assess whether the Reddit series explain variation in several survey-based measures of expectations coming from the *Business and Consumer Surveys* (BCS) collected by the European Commission. The BCS surveys are conducted monthly across the Member States of the EU and five candidate countries. We use euro area aggregates in our analysis. We have two series of price change expectations: one comes from a survey conducted on consumers, and the other is conducted on firms[3], where the survey asks for the selling price over the next three months. The survey question on the unemployment level is available for consumers only. For the firm surveys, the question targets changes in the employment level. As an example, the survey question on price change expectations is reported below, the remaining questions are reported in the Appendix[4]:

[3]For firms, we use a weighted average of the responses for industry, services, retail, and construction surveys.

[4]Questions are taken from the Irish survey to avoid translation to English.

By comparison with the past 12 months, how do you expect that consumer prices will develop in the next 12 months? They will...

Possible responses include: *increase more rapidly (PP), increase at the same rate (P), increase at a slower rate (M), stay about the same (=), fall (MM), don't know.* The final indicator is constructed using the formula: $E^{BCS} = PP + \frac{P}{2} - \frac{M}{2} - MM$, where each term represents the cross-sectional share of respondents selecting the corresponding answer category. The "don't know" responses are excluded from the calculation.

We estimate variations of the following regression using OLS:

$$E_t^{BCS} = \alpha + \beta(L)' X_t^R + \gamma' X_t^* + \varepsilon_t \tag{10}$$

Our focus is on (1) the additional variation explained by Reddit (measured in terms of additional adjusted R-squared) beyond the information contained in the current state of the economy, summarized by the control set X_t^*, and (2) the significance of the coefficients attached to the Reddit indicators. The sample used for the regression coincides with the one used for nowcasting: 2012m1-2023m12 for a total of 144 observations. Standard Errors are heteroskedasticity and autocorrelation robust (HAC).

X_t^R represents a vector containing a Reddit indicator for either inflation (if the target variable is price change expectations) or unemployment (if the target variable is either unemployment or employment change expectations) measured at time t. We also include two lags of this vector. Importantly, due to the design of our signal extraction approach (section 2.4), we have a total of 120 different Reddit series for both inflation and unemployment. In this application, we do not pre-select a single variable, but test all of them one-by-one in regression (10). The Reddit series are aggregated to monthly frequency using averages and transformed to year-on-year differences.

In order to select the set of controls X_t^*, we follow the two-step approach of Belloni and Chernozhukov (2013), by using a Least Absolute Shrinkage and Selection Operator (LASSO) on the EA-MD dataset (Barigozzi and Lissona, 2024). In the first-stage, the LASSO is used to select the variables X_t^* that present coefficients different from zero using 5-fold cross-validation to determine how fast coefficients are shrunk. Since many series are released with substantial delay and so cannot affect survey expectations contemporaneously, we include lagged versions of them in the regressions. For instance, industrial production is released with two months, the unemployment rate with a one-month delay. Financial variables and interest rates instead are treated as available in real-time.[5] The selected variables act as controls in the second-stage OLS regression (10).

[5]The same is true for prices, where the flash estimate is available to agents a few days after the beginning of the next month. In Appendix D we show the variables used from the EA-MD dataset, with associated mnemonics as well as the lag used for each predictor.

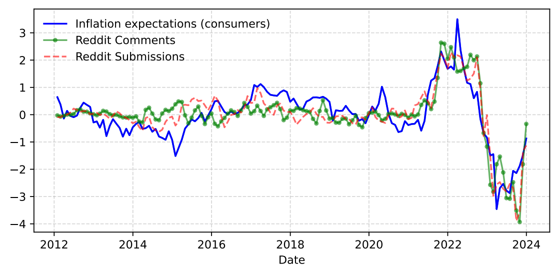
4.2 Results

The variables selected by the LASSO are presetned in Table 4.3 and suggest that survey expectations are persistent (the lag is always included), and systematically co-move with real activity, price/cost dynamics, but also with rates and share prices.

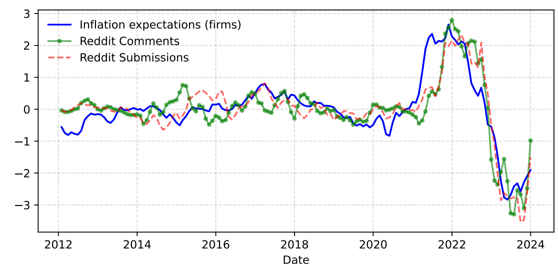
Table 4.3: Optimal regressors selected with LASSO

Indicator Name	**Regressors**
Price Expectations (Consumers)	PriceTrendFut_lag, IPNRG_EA, HPRC_EA, IPING_EA
Price Expectations (Firms)	SellPriceExp_lag, PPINRG_EA, SHIX_EA, CAREG_EA, IPING_EA, IP-NRG_EA
Unemployment Expectations (Consumers)	UnempExp_lag, IPDCOG_EA, TRNCAG_EA, IRT6M_EACC, IP-NRG_EA, HICPSV_EA, PPICAG_EA
Employment Expectations (Firms)	EmpExp_lag, CAREG_EA, SHIX_EA, PPIING_EA, IPING_EA, HPRC_EA, IPNDCOG_EA

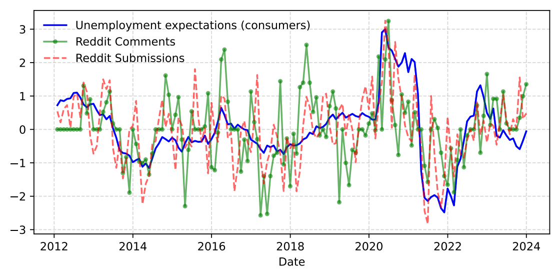
In Figure 4.8 we show the time series of survey expectations and the respectively best-performing Reddit indicators using only submissions or incorporating comments according to in-sample fit. We observe that our social media indices for inflation expectations have similar dynamics to consumers' price change expectations, but trail firm selling price expectations. For the unemployment and employment expectations, we see that the Reddit measures closely track both measures from the consumer survey and are fairly well-aligned with important spikes in the firm survey around the COVID-19 pandemic and thereafter.



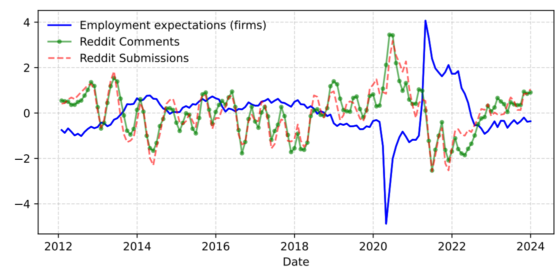
(a) Consumer price expectations



(b) Industry selling price expectations



(c) Consumer unemployment expectations

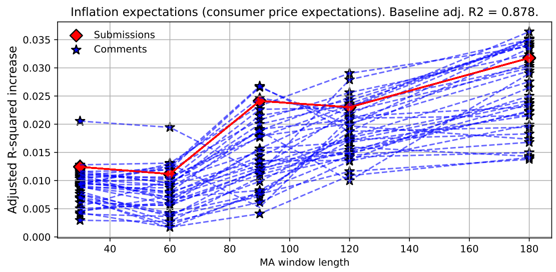


(d) Firm employment expectations

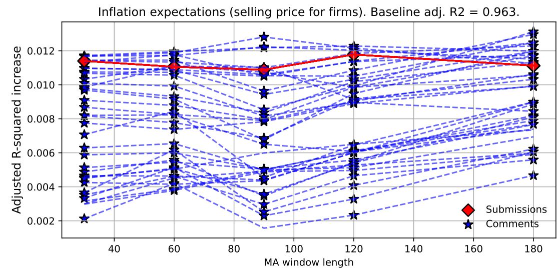
Figure 4.8: Time series of BCS expectations and Reddit series (comments and submissions). Each series is transformed in y-o-y difference, according on how it enters in the regression. The MA selected differs per each series and corresponds with MA giving the best in-sample fit in the regression with expectations.

In Figure 4.9 we group the results for all the expectation series, considering all 120 Reddit indicators. Each of the plots has the *additional* adjusted R-squared on the y-axis when X^R is added on top of the controls, and shows results for all the MA windows on the x-axis. The points with stars show a significant F-test statistic for joint significance of the Reddit indicator and its lags at the 5% confidence level. The figure shows a clear result: Reddit data always help in additionally explaining expectation dynamics, for both firms and consumers. While the submission-only signal is already relevant for inflation, unemployment/employment variables benefit from social discussion. These findings are in line with the results from the nowcasting application.

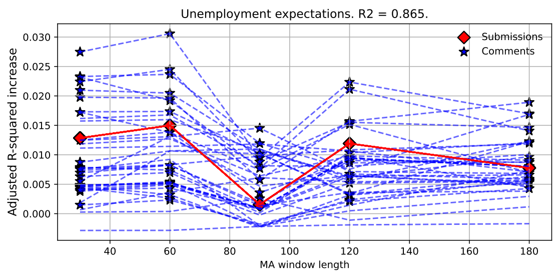
The in sample-results can plausibly be explained by the composition of Reddit submissions (presented in detail in Appendix E), which are for the most part based on newspaper headlines. This means that the submissions act as a filter on the news that Reddit users select and choose to post in the community. We note that news-based mechanisms are typically effective for inflation, but economic literature has not found equally strong media-based channels for labour market expectations. In fact, submissions alone act as a news filter that



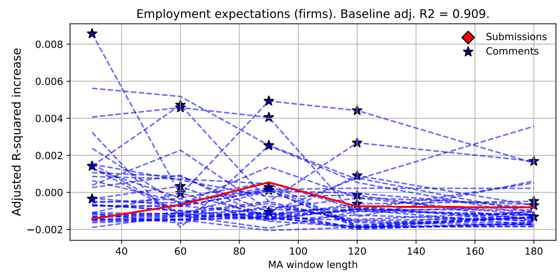
(a) Price trend future expectations



(b) Sell price expectations



(c) Unemployment expectations



(d) Firm employment expectations

Figure 4.9: Incremental change in the adjusted R-squared when the Reddit indicator X^R is added to the regression of survey expectations onto the optimally selected set of controls alone. The star (comments) and dashed (submissions) marker refers to significant f-test statistics.

captures inflation dynamics, but they are insufficient to explain unemployment or employment. On the other hand, labour market discussions on Reddit generate substantially more interaction across users.

As shown in Figure 4.10, unemployment-related submissions receive, on average, significantly more comments per post than inflation-related submissions. This suggests that labour market expectations are more strongly predicted by social interaction rather than news exposure, highlighting the distinct information structures underlying the two types of expectations[6].

[6]The additional discussion is also explained by the different timeliness of publication of official figures. While provisional inflation figures are published on the first day after the end of the month, the publication lag for the European Union (EU) and euro area monthly unemployment figures is approximately one month for monthly unemployment rates and around 75 days for detailed quarterly employment rates.

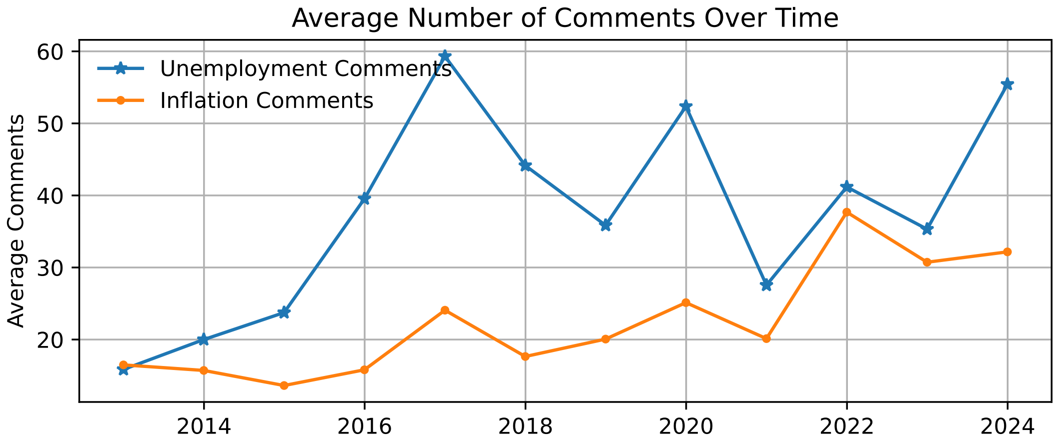


Figure 4.10: Yearly number of comments for inflation and unemployment-related submissions.

5 Conclusion

The main contribution of this paper is to show how social interaction on Reddit can be an important component in the construction of macroeconomic signals for the euro area. Rather than treating posts as isolated observations, we explicitly exploit the discussion they generate. Through a voting scheme that aggregates the sentiment contained in submissions and their associated comment networks, we account for how users confirm, refine, or overturn the initial signal. In this way, discussion threads act as a regularization device that amplifies information collectively endorsed and attenuates noisy or misleading content. We document that this mechanism is particularly useful for labour market conditions: unemployment-related submissions trigger richer interaction, and incorporating comments substantially improves both nowcasting performance and the ability of Reddit indicators to explain survey-based expectations. For inflation, which is released more timely, submissions alone already carry substantial information and social interaction plays a smaller role.

Our empirical evaluation shows that Reddit-based indicators provide valuable high-frequency information for real-time monitoring of euro area prices and labour market conditions. In a nowcasting framework, the Reddit signals significantly improve upon a simple AR(1) benchmark and outperform other daily indicators such as newspaper sentiment and financial variables. Gains are especially pronounced for food price inflation and for unemployment measures, with improvements of up to 13 percentage points for food price inflation and at least 5 percentage points for youth unemployment relative to the next best specification. In a complementary regression framework linking Reddit indicators to business and consumer expectations, we find that these signals significantly increase the explanatory power of models even if they already include a rich set of macro-financial controls. The added value of Reddit is again strongest for unemployment and employment expectations, consistent with the idea that labour market perceptions are shaped not only by news exposure but also by collective discussion.

We conclude that systematically modelling social interaction on platforms such as Reddit constitutes a useful addition to the toolkit available to economic forecasters and policymakers. Social media do not merely transmit information. Indeed, users collectively process and regularize signals in ways that are highly relevant for short-term macroeconomic surveillance. Our framework is flexible and can be embedded in more sophisticated multivariate models such as Bayesian VARs, Dynamic Factor Models or variable selection approaches, and extended to country-specific Reddit communities to obtain geographically granular indicators. Finally, our analysis shows that it is preferable to use LLMs rather than traditional dictionary-based methods for extracting economic signals from social media: in our application, they deliver substantially higher classification accuracy and thereby improve the quality of the resulting indicators.

References

Aliaj, T., Ciganovic, M., and Tancioni, M. (2023). Nowcasting inflation with lasso-regularized vector autoregressions and mixed frequency data. *Journal of Forecasting*, 42(3):464–480. 3

Anesti, N., Esady, V., and Naylor, M. (2024). *Food prices matter most: Sensitive household inflation expectations*. London School of Economics and Political Science. 13

Angelico, C., Marcucci, J., Miccoli, M., and Quarta, F. (2022). Can we measure inflation expectations using twitter? *Journal of Econometrics*, 228(2):259–277. 3, 20

Antenucci, D., Cafarella, M., Levenstein, M., Ré, C., and Shapiro, M. D. (2014). Using social media to measure labor market flows. Technical report, National Bureau of Economic Research. 3

Arteaga Garavito, M. J., Colacito, R., Croce, M. M. M., and Yang, B. (2025). International climate news. *NBER Working Paper*, (w34084). 2

Bailey, M., Cao, R., Kuchler, T., and Stroebel, J. (2018). The economic effects of social networks: Evidence from the housing market. *Journal of Political Economy*, 126(6):2224–2276. 2

Baker, S. R., Bloom, N., and Davis, S. J. (2016). Measuring economic policy uncertainty. *The quarterly journal of economics*, 131(4):1593–1636. 3

Bańbura, M., Giannone, D., Modugno, M., and Reichlin, L. (2013). Now-casting and the real-time data flow. In *Handbook of economic forecasting*, volume 2, pages 195–237. Elsevier. 39

Barbaglia, L., Consoli, S., and Manzan, S. (2024). Forecasting gdp in europe with textual data. *Journal of Applied Econometrics*, 39(2):338–355. 3, 7, 13, 35

Barbaglia, L., Frattarolo, L., Onorante, L., Pericoli, F. M., Ratto, M., and Pezzoli, L. T. (2023). Testing big data in a big crisis: Nowcasting under covid-19. *International Journal of Forecasting*, 39(4):1548–1563. 18

Barigozzi, M. and Lissona, C. (2024). Ea-md-qd: Large euro area and euro member countries datasets for macroeconomic research. 12, 21, 35, 46, 47

Belloni, A. and Chernozhukov, V. (2013). Least squares after model selection in high-dimensional sparse models. 21

Benatti, N., Botelho, V., Consolo, A., Da Silva, A. D., and Osiewicz, M. (2020). High-frequency data developments in the euro area labour market. *Economic Bulletin Boxes*, 5. 3

Born, B., Dalal, H., Lamersdorf, N., Schuster, J.-L., and Steffen, S. (2024). Inflation expectations in the social media age. Manuscript. 3, 20

Breitung, J. and Roling, C. (2015). Forecasting inflation rates using daily data: A nonparametric midas approach. *Journal of Forecasting*, 34(7):588–603. 13

Bybee, L. (2023). Surveying generative ai's economic expectations. *arXiv preprint arXiv:2305.02823*. 4

Carriero, A., Pettenuzzo, D., and Shekhar, S. (2024). Macroeconomic forecasting with large language models. *arXiv preprint arXiv:2407.00890*. 3

Cascaldi-Garcia, D., Ferreira, T. R., Giannone, D., and Modugno, M. (2024). Back to the present: Learning about the euro area through a now-casting model. *International Journal of Forecasting*, 40(2):661–686. 20

Coibion, O. and Gorodnichenko, Y. (2025). Expectations matter: the new causal macroeconomics of surveys and experiments. In *Expectations Matter*. Princeton University Press. 20

Consoli, S., Barbaglia, L., and Manzan, S. (2022). Fine-grained, aspect-based sentiment analysis on economic and financial lexicon. *Knowledge-Based Systems*, 247:108781. 7, 8

Daas, P. J. H. and Puts, M. J. H. (2014). Social media sentiment and consumer confidence. Statistics Paper Series 5, Statistics Netherlands (CBS), The Hague, Netherlands. 20

Del Monaco, A., Longo, L., Marcucci, J., and Tafani, I. (2025). Reddit's pulse on us inflation: Forecasting with large language models. Working paper, Bank of Italy. 3

Ehrmann, M. and Wabitsch, A. (2022). Central bank communication with non-experts–a road to nowhere? *Journal of Monetary Economics*, 127:69–85. 2

European Commission (2025). *The European media industry outlook – September 2025*. Publications Office of the European Union. Prepared by the Directorate-General for Communications Networks, Content and Technology. 2

European Parliament (2025). Social media survey 2025. Accessed: 2026-05-06. 2

Faria-e Castro, M. and Leibovici, F. (2024). Artificial intelligence and inflation forecasts. Technical report. 3

Ghysels, E., Kvedaras, V., and Zemlys, V. (2016). Mixed frequency data sampling regression models: The r package midasr. *Journal of statistical software*, 72:1–35. 13

Giacomini, R. and Rossi, B. (2010). Forecast comparisons in unstable environments. *Journal of Applied Econometrics*, 25(4):595–620. 37

Gorodnichenko, Y., Pham, T., and Talavera, O. (2024). Central bank communication on social media: What, to whom, and how? *Journal of Econometrics*, page 105869. 3, 20

Granziera, E., Larsen, V. H., Meggiorini, G., and Melosi, L. (2025). Speaking of inflation: the influence of fed speeches on expectations. 7, 8, 30, 31

Gueta, A., Feder, A., Gekhman, Z., Goldstein, A., and Reichart, R. (2025). Can llms learn macroeconomic narratives from social media? In *Findings of the Association for Computational Linguistics: NAACL 2025*, pages 57–78. 4

Hansen, A. L., Horton, J. J., Kazinnik, S., Puzzello, D., and Zarifhonarvar, A. (2024). Simulating the survey of professional forecasters. *Available at SSRN*. 4

Heikkinen, J. and Heimonen, K. (2025). Media tone: The role of news and social media on heterogeneous inflation expectations. *Journal of Forecasting*, 44(3):881–921. 3

Liu, Q. and Son, H. (2024). Methods for aggregating investor sentiment from social media. *Humanities and Social Sciences Communications*, 11(1):1–22. 43

Long, S., Lucey, B., Xie, Y., and Yarovaya, L. (2023). "i just like the stock": The role of reddit sentiment in the gamestop share rally. *Financial Review*, 58(1):19–37. 6

Loughran, T. and McDonald, B. (2011). When is a liability not a liability? textual analysis, dictionaries, and 10-ks. *The Journal of finance*, 66(1):35–65. 6

Macaulay, A. and Song, W. (2023). Narrative-driven fluctuations in sentiment: Evidence linking traditional and social media. Technical report, Bank of Canada Ottawa. 3

Online appendix

A Details on social media signal construction

A.1 Dictionary-based signal

In Table A.4 we provide the words that reflect positive or negative sentiment/signal towards economic concepts. The dictionary is taken from Granziera et al. (2025).

Positive Sentiment (+1)	**Negative Sentiment (-1)**
Boost	Below
Climb	Collapse
Elevate	Damp
Escalate	Deteriorate
Expand	Decline
Foster	Diminish
High	Down
Increase	Drop
Height	Ease
Intensify	Fall
Jump	Low
Persist	Modest
Pressure	Moderated
Moderate	Muted
Rise	Plummet
Risk	Reduction
Remain	Restrain
Rising	Retreat
Rose	Set Back
Risen	Slow
Soar	Soft
Solid	Subdued
Spike	Weak
Sustain	
Strong	
Strength	
Surge	
Upward	
Up	
Upside Risk	

Table A.4: Sentiment Classification of Words

A.2 Signal construction with social interaction

To further illustrate the steps involved in the construction of Reddit signals with social interaction, Figure A.11 highlights two points. First, it provides an example of how the LLM classifies a submission differently from a dictionary approach (in this example NEUTRAL (LLM) vs. DOWN (dictionary)). Since we specifically instruct the LLM to predict the *future* direction of the economic concept, it is arguably correct in classifying as NEUTRAL the submission's statement, which is simply a description of the current state of the unemployment rate. The dictionary approach of Granziera et al. (2025), on the other hand, attributes a downward direction to the submission, due to the presence of the keyword "low". Second, the graphic shows a situation where the sentiment conveyed in the unweighted first-level comments following the submission overturns the score assigned to it. The commenters view the record low of the unemployment rate as a turning point in the business cycle and predict that it will rebound in the future, leading to an overall upward expectation. For a threshold of 0.1, this is sufficiently strong to re-classify the label of the submission from NEUTRAL to UP.

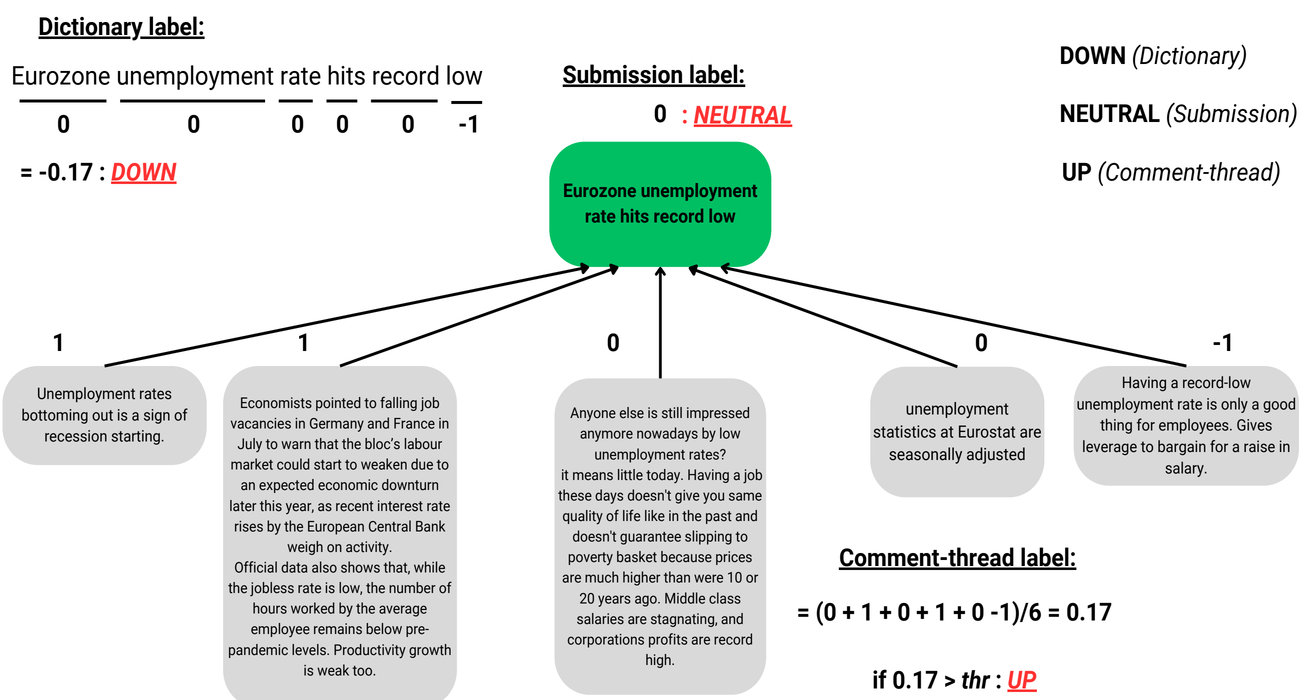


Figure A.11: Illustration of a submission-comment network. The initial submission is labeled as DOWN (-1) by the dictionary classifier, and NEUTRAL (0) by the LLM. The submission's signal is adjusted to UP using comment-based signals and a threshold $\tau = 0.1$.

Next, we check how often and when such re-classifications based on social interaction occur. We observe that in the case of inflation (HICP), re-classification occurs in 11% of the cases. For unemployment (the overall rate) the share increases to 15%. For both variables, social interaction therefore leads to hundreds of re-classifications. Since nothing else changes between a MIDAS model that uses only submissions and the one that accounts for social interaction, it is this re-classification which reduces nowcasting errors.

Re-classification of Reddit signals for inflation

Submissions	Comments	Number of re-classifications
-1	0	154
-1	1	8
0	-1	43
0	1	196
1	-1	2
1	0	124

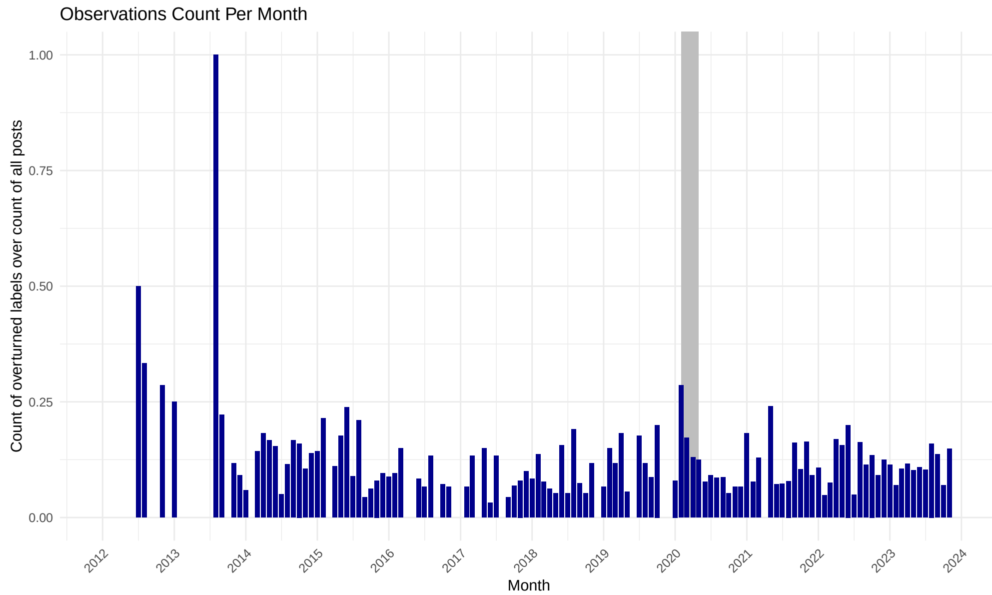
Table A.5: Re-classifications of inflation signals attached to a post based (1) only on the original submission or (2) factoring in the sentiment attached to the resulting comments. There are in total 527 re-classifications out of 4825 (10.92%). Comments uses threshold 0.3, no upvote/downvote scoring and only the first level of comments as this was the winning configuration.

Re-classification of Reddit signals for unemployment

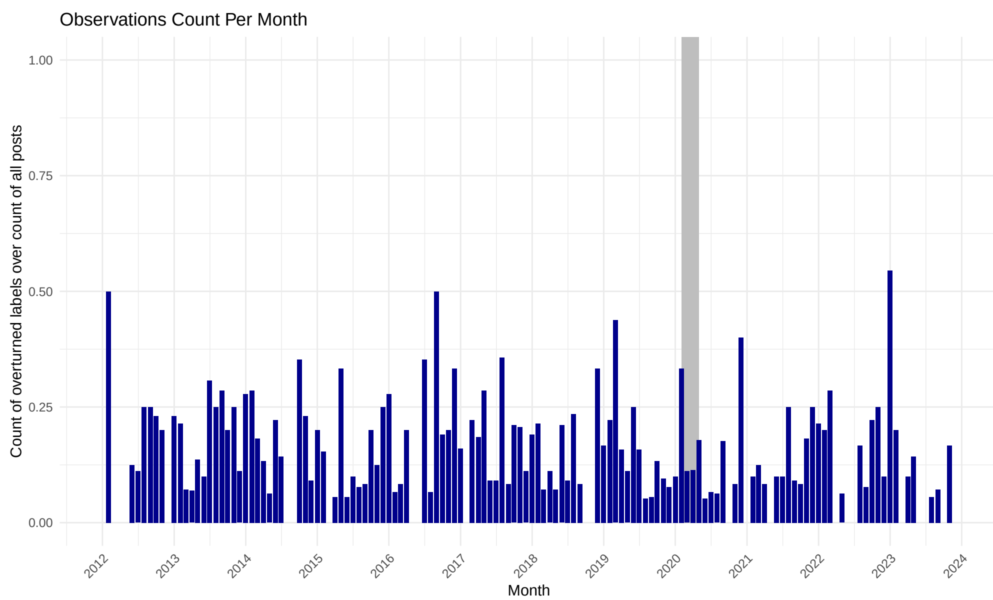
Submissions	Comments	Number of re-classifications
-1	0	111
-1	1	2
0	-1	58
0	1	73
1	-1	0
1	0	47

Table A.6: Re-classifications of unemployment signals attached to a post based (1) only on the original submission or (2) factoring in the sentiment attached to the resulting comments. There are in total 291 re-classifications out of 1934 (15.05%). Comments uses threshold 0.3, no upvote/downvote scoring and a dictionary filter for comments as this was the winning configuration.

Tables A.5 and A.6 show the direction of change that comes about through the incorporation of comments into the signal for the winning specifications. In the case of inflation, the total signal is revised upwards whenever the re-classification is from -1 to 0, -1 to 1 and from 0 to 1. This happens 366 times (counting -1 to 1 twice) while only 171 downgrades occur, a ratio of 2.1. Therefore, the overall sentiment for the direction of inflation in the sample we have considered was generally revised upwards. For unemployment, there are 188 upward revisions and 105 downward revisions, a ratio of 1.8. The results of the horse race suggest that this sharpens the signal, as the nowcast accuracy improves. Since we have observed the clearest nowcasting gains in unusual time periods, we also check if these re-classifications are concentrated in and around these times. Figure A.12 shows that this is not the case, which leads us to conclude that this type of regularization is not episodic but constitutes a systematic feature of social media signals that researchers can use to improve their information set.



(a) HICP inflation



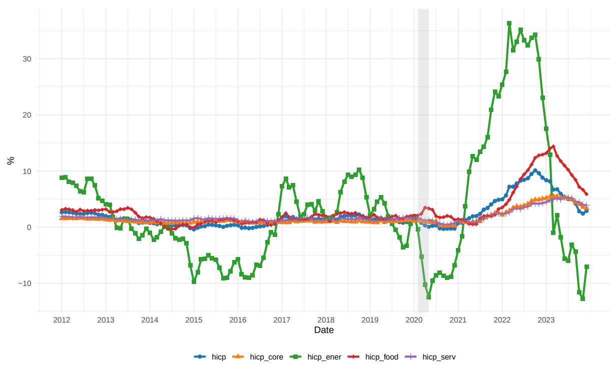
(b) Unemployment rate

Figure A.12: Frequency of re-classifications as a fraction of total submissions over time. Grey shaded area is the COVID-19 recession.

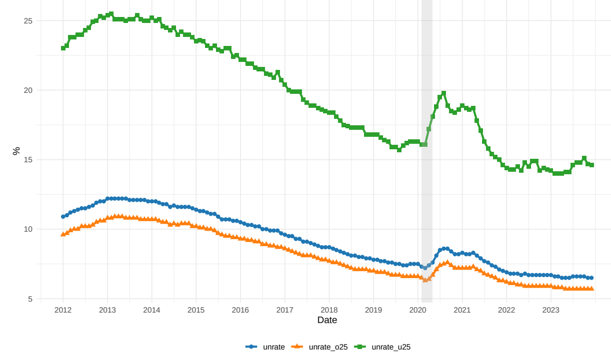
B Details on nowcasting horse-race

B.1 Horse race target variables

The target variables for the horse-race are depicted in Figure B.13. Clearly energy price inflation was more volatile than the other series in the sample we examine. Similarly, the youth unemployment rate was at a substantially higher level than the aggregate or over 25 years unemployment rates.



(a) Inflation data

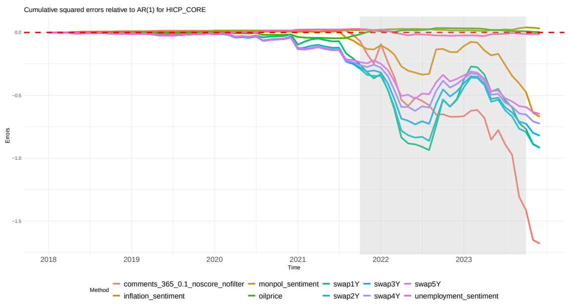


(b) Unemployment data

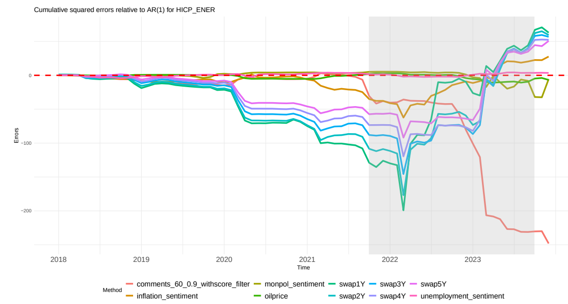
Figure B.13: Time series of observed HICP inflation and unemployment data from the EA-MD data base of Barigozzi and Lissona (2024). Grey shaded area is the COVID-19 recession.

B.2 More daily horse-race results

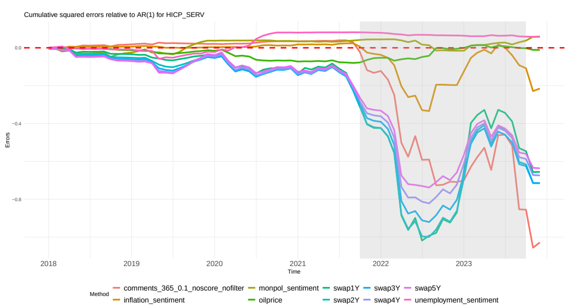
Figure B.14 presents the cumulated squared errors in differences to the AR(1) benchmark for the components of inflation and the unemployment rate. For all target variables, the Reddit signal involving social interaction produces the best result at the end of the sample. For inflation, the swap series are useful nowcasting predictors on the onset of the high inflation period, but less so at the end. For the unemployment rate components, the only real competitor to our Reddit series is the unemployment sentiment of Barbaglia et al. (2024).



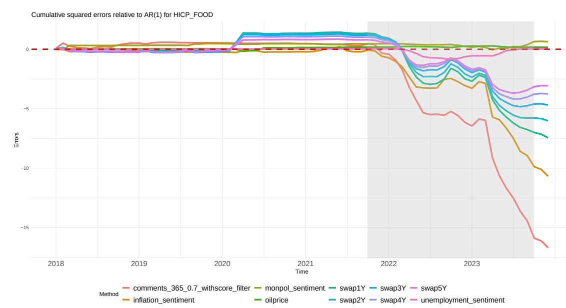
(a) Core inflation



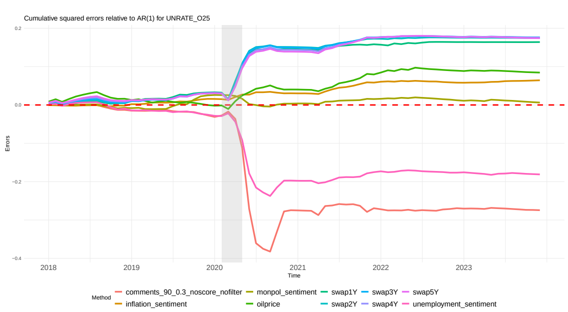
(b) Energy price inflation



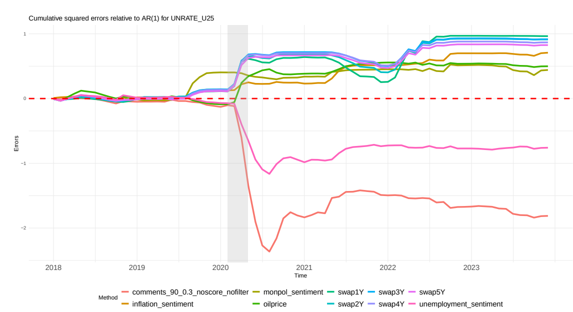
(c) Service price inflation



(d) Food price inflation



(e) Unemployment >25

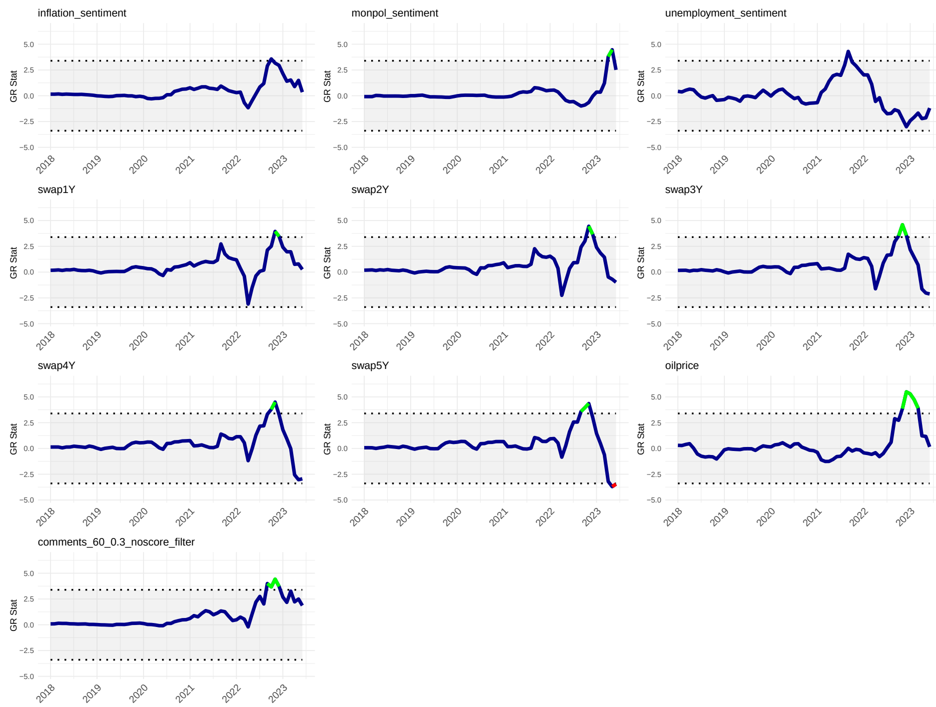


(f) Unemployment <25

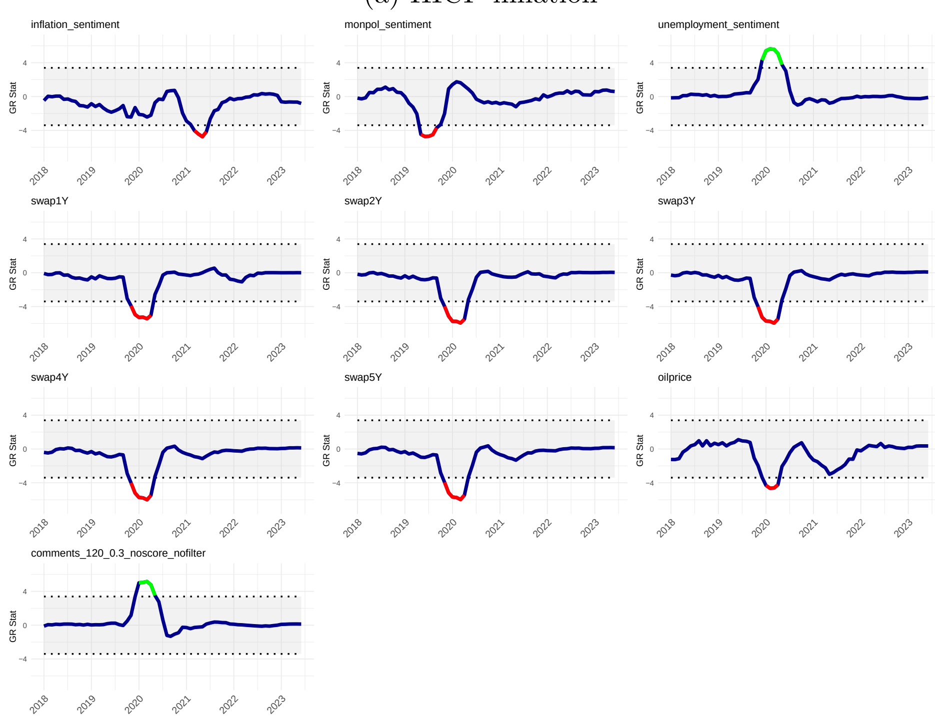
Figure B.14: Cumulative squared errors of MIDAS models minus cumulative squared errors of AR(1) over time. Values below 0 mean better performance than the AR(1). Shaded areas are periods when HICP inflation was above 4% and the COVID-19 recession months respectively.

B.3 Nowcasting gains stability tests

Given that the performance gains appear to be driven by unusual times, we check the models' performance using the instability test of Giacomini and Rossi (2010), where we choose the moving-window in the evaluation sample to be of size 10% ($\mu = 0.1$). The test provides evidence on the time periods in which the the MIDAS specification achieves significant gains (or sustains losses) compared to the AR(1) benchmark. Figure B.15 shows that, indeed, the Reddit signal for inflation is significantly better than the benchmark during the hyperinflation period around the end of 2022. In the case of the unemployment rate, the indicator provides improvements over the benchmark in the COVID-19 period.



(a) HICP inflation

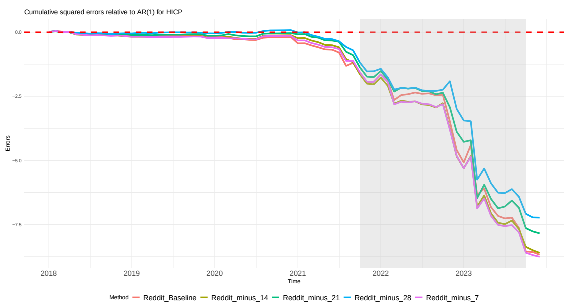


(b) Unemployment rate

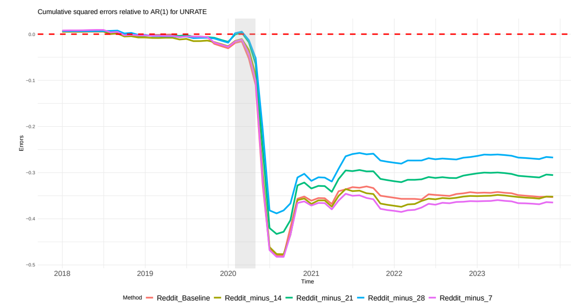
Figure B.15: Giacomini-Rossi test statistics and critical values over time. If the statistic is within the bounds there is no evidence of significantly different performance. If the statistic exceeds the upper bound (green chunks) the model outperforms the AR(1), otherwise (red chunks) it underperforms.

B.4 The role of daily information flow

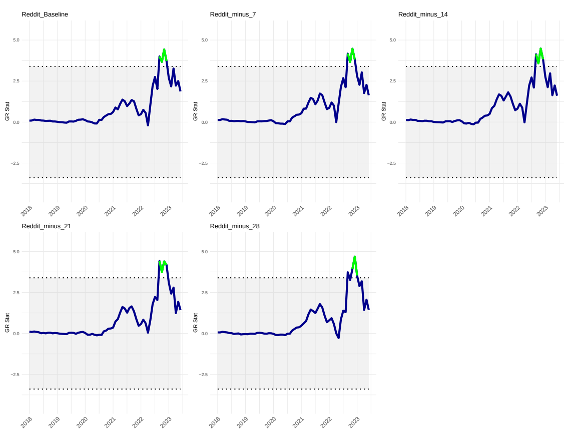
To better judge the real-time value of the daily Reddit series, we conduct an additional check. We relax the assumption that the daily information is used once all of it has been revealed at the end of the target period. Instead, we assume that the information cutoff occurs 7, 14, 21, and 28 days before the end of the period that is to be nowcasted. This resembles the nowcasting setup of Bańbura et al. (2013), who emulate the flow of information through the months of the quarter to be nowcasted. In principle, more information should improve the nowcasts, but a practitioner may not have the full month available at the point when the nowcast is to be released.



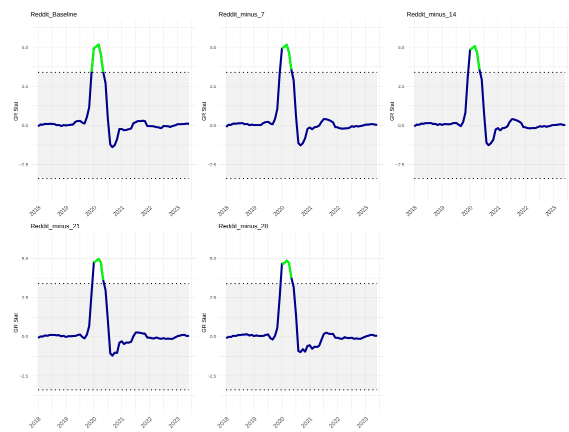
(a) Cumulative Errors HICP inflation



(b) Cumulative Errors Unemployment rate



(c) Giacomini-Rossi HICP inflation



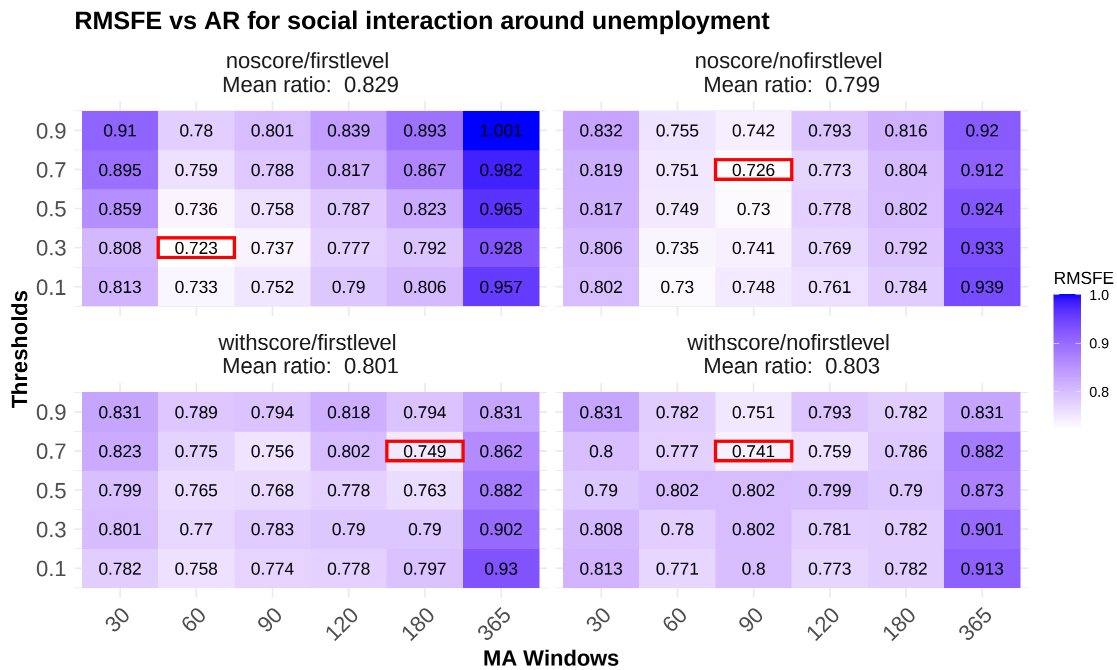
(d) Giacomini-Rossi Unemployment rate

Figure B.16: Cumulative squared errors of MIDAS with information cutoffs 0,7,14,21, and 28 days before the end of the nowcast period. Values are calculated subtracting cumulative squared errors of AR(1) over time. Values below 0 mean better performance than the AR(1). Shaded areas are periods when HICP inflation was above 4% and the COVID-19 recession months respectively. Giacomini-Rossi test statistics and critical values over time. If the statistic is within the bounds there is no evidence of significantly different performance. If the statistic exceeds the upper bound (green chunks) the model outperforms the AR(1), otherwise (red chunks) it underperforms.

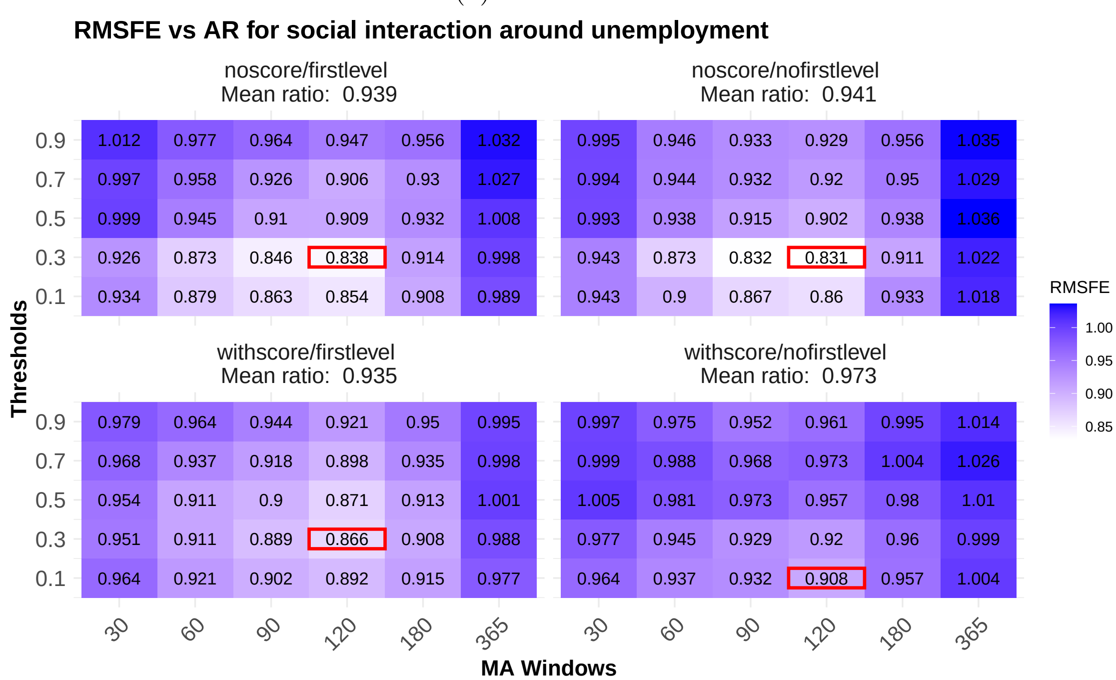
Figure B.16 reports the result of the MIDAS models for the best performing Reddit specifications (in terms of RMSFE) for HICP inflation and the unemployment rate with varying information cutoffs. For both variables, there is some benefit to considering additional information, but this is clear up to around 14 days before the cutoff only. Thereafter performances are essentially equal. The stability tests emphasize that the information contained in the signal is significantly useful in all cases, highlighting the real time value of the Reddit data.

B.5 The role of thresholds and pre-filtering for comments

While the choice of smoothing window is clearly important in all plots there is at least one specification involving comments that beats the optimal submissions-only model, for the unemployment rate this is more obvious than for the inflation data. To better understand what components – if any – of the comments signals (smoothing, thresholds, scoring, or pre-filtering) contribute systematically to these performance gains, we break the RMSFE scores against the AR(1) down by these four dimensions.



(a) HICP inflation



(b) Unemployment rate

Figure B.17: Relative RMSFE scores against an AR(1) for MIDAS specifications estimated with different variations of Reddit comments signals. Brighter colors imply better scores, values below 1 imply that the Reddit signal improves upon the AR(1).

The results of this are displayed in Figure B.17. For HICP inflation we see that there is no clear pattern that would indicate significant gains stemming from one of the four dimensions that influence the construction of the Reddit signals. Using the scoring rule based on upvotes and downvotes seems to produce slightly better results than the specification which does not incorporate this extra layer of information even though the overall best performing model does not apply it. The pre-filtering using the list of keywords does not appear to be relevant and there is no clear guidance on the value of the threshold which should optimally be employed. The results are clearer in the case of the unemployment rate. The pre-filtering improves performance, especially in combination with the upvote/downvote scoring, at least on average. Again, the standout performance is achieved by a specification which uses neither of these. A low threshold together with a smoothing window of 120 is optimal.

C Robustness checks

We conduct robustness checks on the two key contributions of this paper. The results are reported in Figure C.19.

C.1 Construction of the Reddit comments signal

We have claimed that it is important to account for the full comment structure below a Reddit submission to refine the social signal. To do this, we have opted for a voting system that incorporates comments in an unweighted or weighted fashion (based on upvotes), in a separate labeling step following the initial labeling of the original submission. In a robustness check, we give the entire text consisting of the submission together with all comments to the LLM and ask it to assign a sentiment score based on the full text, i.e. the implicit weighting of submissions and responses is done with the LLM. The advantage of this is that the researcher does not have to choose a specific aggregation rule, as there is no consensus in the literature on how this should be done (Liu and Son, 2024). The disadvantage is that the implicit voting scheme happens inside a blackbox. We use the following prompt:

```
You are a forecaster and want to predict the direction of future
{concept} in Europe from social network conversations (post and related
comments). The following document will report sentences potentially
referring to Europe: {df['title'][i]}. You have to print a signal
that can be UP (if the document is signalling {concept} going up in
the short/long-term run), DOWN (if the sentence is signalling {concept}
going down in the short/long-term run) or NEUTRAL (if the sentence is
neutral or does not signal a particular direction on {concept}). Print
only the results of the signal, do not summarize the sentence nor give
any reason on your choice. Even if more sentences or paragraphs are
provided to you, you only have to print one signal that can be UP, DOWN
or NEUTRAL.
```

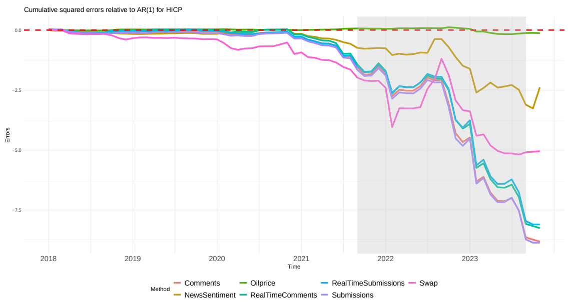
Figure C.18: LLM prompt used with LLaMa-3-70b-instruct. *concept* is either unemployment rate or inflation rate, `df['title'][i]` contains the submission as well as the attached comments.

C.2 Selection of the best performing nowcasting information set

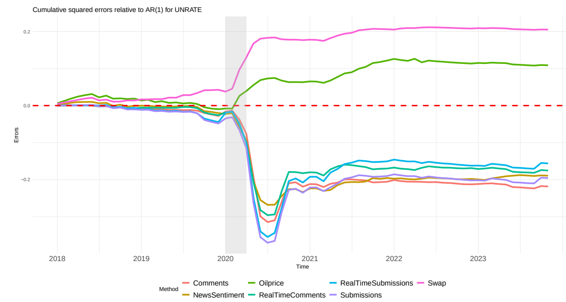
We have evaluated different Reddit signals in an out-of-sample exercise and concluded that using comments is uniformly preferable to other high-frequency indicators. Based on the RMSE results, the clear recommendations for a nowcaster starting to use social media monitoring today are reported in Table 3.2. As we have shown in Figure 3.6, the choice of the MA filter is quite important for this. Therefore, we consider another way of evaluating performance of the indicators by emulating a researcher having access to the Reddit signals in real time and selecting the optimal MA smoothing window at each point in the forecast evaluation sample, based on continuously updating the RMSE statistic available to her (this strategy is called "real time" below).

C.3 Robustness checks results

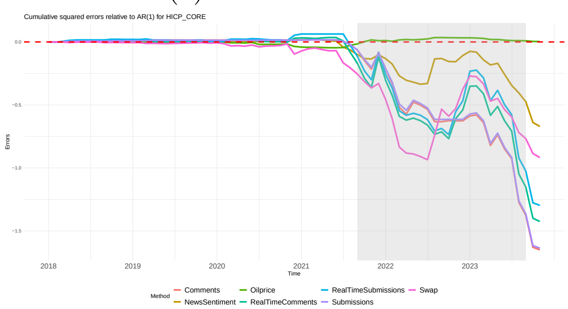
Results for the cumulated squared errors relative to those of the AR(1) benchmark are reported in Figure C.19. The endpoint of each line is a measure of the sum of squared errors for the out-of-sample period. To simplify the presentation, we report for each outcome variable the statistics for the overall best signal based only on submissions, the overall best signal based on comments and submissions fed jointly to the LLM, the overall best newspaper sentiment indicator, the best inflation swap, the oil price, as well as the submissions and comments signals, constructed by a real time nowcaster based on available RMSE information.



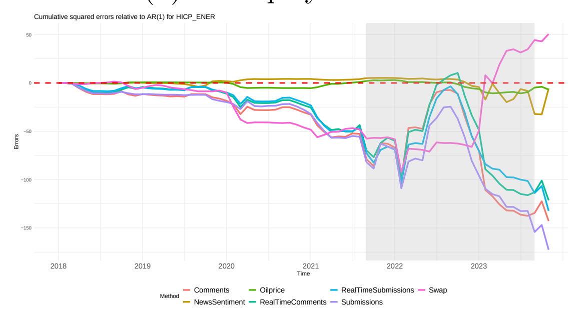
(a) HICP inflation



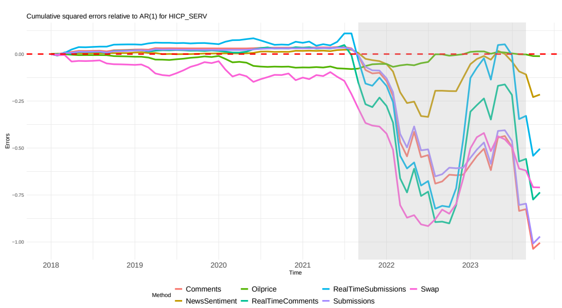
(b) Unemployment rate



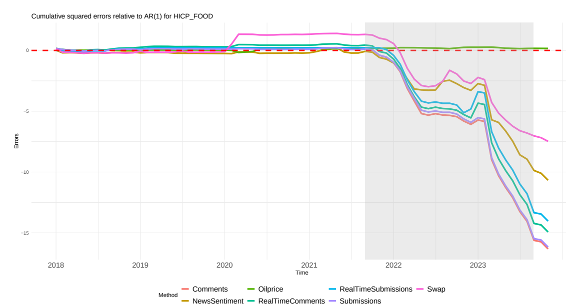
(c) Core inflation



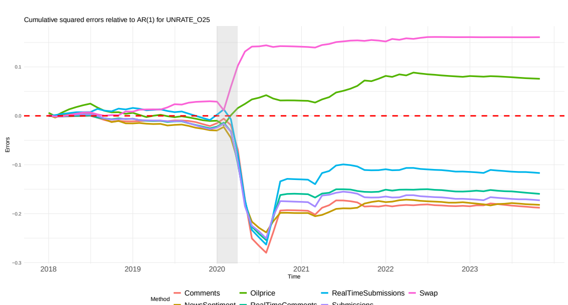
(d) Energy price inflation



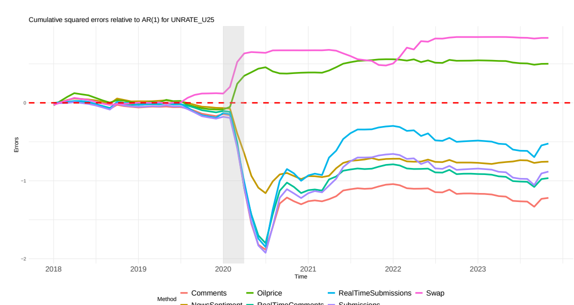
(e) Service price inflation



(f) Food price inflation



(g) Unemployment >25



(h) Unemployment <25

Figure C.19: Cumulative squared errors of MIDAS models minus cumulative squared errors of AR(1) over time. Values below 0 mean better performance than the AR(1). Shaded areas are periods when HICP inflation was above 4% and the COVID-19 recession months respectively.

The checks confirm the results of the main body. First, they suggests that Reddit indicators provide the best overall results for all outcome variables, although in the cases of Energy Price Inflation and the overall HICP index, submissions outperform comments incorporated directly into the LLM call. Second, a nowcaster selecting the optimal MA filter available in real time should generally expect to make a slightly larger error than one that sticks to a single MA window (provided the correct one is chosen) for future nowcasts. Third, differences between comments and submissions are very small with this kind of comment incorporation. Comparing these robustness results to the ones of the main body using a voting scheme to incorporate comments, we find that the out-of-sample error for the LLM based comments incorporation is essentially the same as for the voting based system, if anything slightly higher. This comparison is depicted in Table C.7. It appears that the way comments are incorporated is less important than ensuring that they are actually included in the signal.

Robustness Results Table

Target	*Submissions*	*Submissions RT*	*Comments Voting*	*Comments Voting RT*	*Comments LLM*	*Comments LLM RT*
HICP	0.37	0.38	0.36	0.37	0.36	0.37
HICP Core	0.24	0.25	0.24	0.25	0.24	0.25
HICP Energy	2.98	3.17	2.96	3.04	3.2	3.26
HICP Food	0.47	0.48	0.44	0.46	0.45	0.47
HICP Services	0.24	0.24	0.23	0.24	0.23	0.24
Unemployment	0.11	0.11	0.10	0.11	0.11	0.12
Unemployment > 25	0.10	0.11	0.09	0.09	0.10	0.10
Unemployment < 25	0.35	0.35	0.34	0.34	0.35	0.36

Table C.7: RMSE for different Reddit signals and outcomes estimated using MIDAS. RT stands for "real time."

D EA-MD variable descriptions

Table D.8 shows the description from the predictors from the EA-MD dataset (Barigozzi and Lissona, 2024).

ID	*Series Name*
(1)	Exchange rates (contemporaneous)
REER42	Real Exchange Rate (42 main industrial countries)
ERUS	Real Exchange Rate (42 main industrial countries)
(2)	Interest rates (contemporaneous)
IRT3M	3-Month Interest Rate
IRT6M	6-Month Interest Rate
LTIRT	Long-Term Interest Rate (EMU Criterion)
(3)	Industrial Production and Turnover (2-months lag)
IPMN_EA	Industrial Production: Manufacturing
IPCAG_EA	Industrial Production: Capital Goods
IPCOG_EA	Industrial Production: Consumer Goods
IPDCOG_EA	Industrial Production: Durable Consumer Goods
IPNDCOG_EA	Industrial Production: Non-Durable Consumer Goods
IPING_EA	Industrial Production: Intermediate Goods
IPNRG_EA	Industrial Production: Energy
TRNMN_EA	Turnover Index: Manufacturing
TRNCAG_EA	Turnover Index: Capital Goods
TRNCOG_EA	Turnover Index: Consumer Goods
TRNDCOG_EA	Turnover Index: Durable Consumer Goods
TRNNDCOG_EA	Turnover Index: Non-Durable Consumer Goods
TRNING_EA	Turnover Index: Intermediate Goods
TRNNRG_EA	Turnover Index: Energy
(4)	Prices (contemporaneous)
PPICAG_EA	Producer Price Index: Capital Goods
PPICOG_EA	Producer Price Index: Consumer Goods
PPINDCOG_EA	Producer Price Index: Non-Durable Consumer Goods
PPIDCOG_EA	Producer Price Index: Durable Consumer Goods
PPIING_EA	Producer Price Index: Intermediate Goods
PPINRG_EA	Producer Price Index: Energy
HICPOV_EA	HICP: Overall Index
HICPNEF_EA	HICP: no Energy & Food
HICPG_EA	HICP: Goods
HICPIN_EA	HICP: Industrial Goods
HICPSV_EA	HICP: Services
HICPNG_EA	HICP: Energy
DFGDP_EA	GDP Deflator
HPRC_EA	Residential Property Prices
CURR_EACC	Currency in Circulation
(4)	Unemployment (1-month lag)
UNETOT	Unemployment: Total
UNEO25	Unemployment: Over 25 years
UNEU25	Unemployment: Under 25 years
(5)	Others (contemporaneous)
SHIX_EA	Share Prices
CAREG_EA	Passenger Car Registrations

Table D.8: Euro area variables used in the analysis, from EA-MD (Barigozzi and Lissona, 2024).

E Reddit composition

In Table E.9 we present statistics on the composition of Reddit submissions related to unemployment and inflation. Each submission includes several attributes that allow us to summarize its content. Among these, the submission flair is one of the most informative, as it indicates the type of content submitted. In the table, we report the three most recurrent flair types: Missing flair, News flair, and Data flair. The first category (Missing flair) is common because assigning a flair is not compulsory. The News flair indicates that the submission is based on a news source, while the Data flair generally refers to posts including quantitative information, often retrieved from public datasets. These two categories are closely related, as both reflect newspaper-style information shared within the community. However, given the large share of submissions with Missing flair, it is likely that many of these also correspond to news posts, since users can post news without assigning a flair. To clarify this, we compute the percentage of submissions containing selftext, which identifies posts that include original written content as opposed to direct links to external news sources. Only 16.9% of inflation-related submissions and 14.2% of unemployment-related submissions contain selftext, meaning that the majority of submissions are not original commentary but rather external content. Taken together, these results suggest that a signal based solely on submissions primarily captures news selected by the Reddit community rather than users' personal views or qualitative opinions.

	Inflation	Unemployment
Missing flair	50.1%	69.1%
News flair	28.1%	10.8%
Data flair	7.2%	8.3%
Selftext	16.9%	14.2%

Table E.9: Percentage of submissions.

F Real-time dashboard

The inflation-nowcast-socialplatform dashboard consists in our effort of collecting Reddit data in real-time and processing them with the Llama model. In the dashboard, we provide a real-time forecasts (1-step-ahead) for inflation and unemployment rate in the Euro Area[7].

[7]If the dashboard does not load, please click the button that appears: "Yes, get this app back up!".

G BCS questions

Consumer survey question on unemployment:

> *How do you expect the number of people unemployed in this country to change over the next 12 months? The number will ...*

Possible answers are increase sharply, increase slightly, remain teh same, fall slightly, fall sharply and Don't Know.

Industry survey on employment:

> *How do you expect your firm's total employment to change over the next 3 months? It will. . .*

Industry survey on selling prices:

> *How do you expect your selling prices to change over the next 3 months? They will. . .*

For the industry survey, possible responses are increase (+), remain unchanged (=), decrease (-) and Refuses/Not applicable.