Monetary Policy in the Age of Social Media: A Twitter-Based Inflation Analysis*

Benjamin Born, Hrishbh Dalal, Nora Lamersdorf, Sascha Steffen

September 7, 2023

Abstract

We develop a high-frequency inflation index derived from German tweets through sophisticated NLP methods. This index aligns closely with realized inflation and consumer inflation expectations, both nationally and regionally, offering enhanced predictive precision over current benchmarks. Notably, it responds to monetary policy shifts, rising post-easing and falling after unexpected tightenings. The influence is particularly pronounced from tweets by private users during recent periods of elevated inflation. Elevated inflation expectations correlate with reduced consumer spending, as gauged from online transaction data, particularly on discretionary goods. Consequently, this Twitter-centric index offers a valuable real-time tool to assess prevailing inflation sentiments.

Keywords: Inflation expectations, monetary policy, household consumption, text mining, machine learning, Twitter, big data, transformer models, ChatGPT

JEL-Codes: E31, D84, E58, C55

*Born: Frankfurt School of Finance & Management, CEPR, CESifo, and ifo Institute, b.born@fs.de, Dalal: Frankfurt School of Finance & Management, h.dalal@fs.de, Lamersdorf: Frankfurt School of Finance & Management, n.lamersdorf@fs.de, Steffen: Frankfurt School of Finance & Management, s.steffen@fs.de. We thank BMWK and DLR for funding. Survey data on inflation expectations are provided by Deutsche Bundesbank under the project number 2022\0105 and by Gesellschaft für Konsumforschung (GfK). We thank Sarah Arndt (discussant), Felix Chopra (discussant), Sanjiv Das, Philip Schnorpfeil, Michael Weber and participants at the DIW Workshop for Women in Macroeconomics, Finance and Economic History 2023, the CESifo Venice Summer Institute 2023, the 13th ifo Conference on Macroeconomics and Survey Data, the 5th Behavioral Macroeconomics Workshop, the 7th SAFE Household Finance Workshop, and the 2023 EEA-ESEM congress for helpful comments and suggestions.

1 Introduction

Inflation expectations play a pivotal role in shaping the consumption, savings, and investment behaviors of firms and households. They not only serve as a vital predictor for future inflation for central banks and other institutions (see, e.g., ECB, 2021) but can also act as a monetary policy instrument if central banks can effectively influence them (Coibion et al., 2020a). Therefore, understanding these expectations is essential for central banks to assess the impact of their announcements and to tailor their policy strategies accordingly.

Yet, accurately capturing inflation expectations presents challenges, as highlighted by Weber et al. (2022). Factors such as the precise phrasing of survey questions, potential biases introduced by inadvertent priming, and the typically limited sample size can skew results. Moreover, survey data often comes with significant delays. Leveraging new data sources and harnessing the power of recent machine learning advancements could address some of these limitations inherent in traditional survey- or market-based measures.

In this study, we delve into the potential of leveraging Twitter text content to gauge inflation expectations. Given Twitter's status as a leading social media platform, it offers a rich, high-frequency snapshot of diverse individual perspectives. Our goal is to extract insights from these myriad opinions about current or anticipated price levels, juxtapose them with established measures of inflation expectations, and uncover the economic implications of our Twitter-derived metric. Specifically, we investigate if monetary policy retains its influence over household inflation expectations. Furthermore, how does this impact the consumption patterns of households? Our real-time, high-frequency index is ideally positioned to address these queries.

To achieve this, we employ cutting-edge natural language processing (NLP) methodologies to craft an inflation index for Germany, drawing from German-language tweets spanning January 2011 to May 2023. Our initial step involves collating all German tweets containing keywords pertinent to prices and inflation. To ensure the authenticity of our data by filtering out potential bot interference and other extraneous content like advertisements, we deploy unsupervised machine learning tools within the BERTopic framework (Grootendorst, 2022). Notably, we opt for the TwHIN-BERT model by Zhang et al. (2022) for word embeddings, given its aptness for social media contexts. By examining the key tokens within these topics, we pinpoint those linked to inflation and subsequently focus on tweets associated with these inflation-centric topics.

Using our trained topic model, we categorized each tweet in our corpus, selecting 1,294,047 tweets (12.12% of the total) that aligned with one of the 19 identified inflation topics. For clarity, these topics were further grouped into four sub-topics which encompass general inflation, energy prices, inflation in relation to monetary policy, and housing-related inflation. The general inflation topic dominates, accounting for over 60% of inflation-related tweets, with a notable surge in 2022. Discussions on energy prices and monetary policy in the context of inflation have been consistent but saw a marked increase in 2021 and 2022.

In the next step, we classify these tweets as "up", "down", or "neutral", representing increasing, decreasing, or unchanged inflation. For this task, we use a large pre-trained neural network language model which we fine-tune on our specific classification task, by training additional layers on top of the pre-trained model. To generate enough training data for this fine-tuning process, we use the ChatGPT API.[1]

To label tweets, we harness machine learning, specifically a fine-tuned neural network language model. We utilize ChatGPT, specifically the gpt-3.5-turbo-0301 model, to categorize tweets into three classes: increasing inflation/prices, decreasing inflation/prices, and other. This approach eliminates the biases and time constraints of manual labeling. We then fine-tune the TwHIN-BERT model, designed for tweets in multiple languages, using 90% of this data. The model's parameters are adjusted to minimize the loss function during training. The final model classifies our corpus into 377,260 tweets about rising inflation/prices, 87,671 about falling inflation/prices, and 829,116 as other, ensuring only pertinent inflation-related tweets are retained.

Finally, for each day, we aggregate all tweets labeled as up and down to a daily inflation index. In addition to an aggregate German Inflation Index, based on information contained in the Twitter users' profile description, we extract their locations and also build regional inflation indices.

The Twitter-based Inflation Index, when juxtaposed with Germany's CPI Inflation and the Bundesbank Online Panel – Households (BOP-HH) inflation expectations, reveals a close alignment for most periods. Notably, both inflation surges and declines, such as those in early 2017 and 2015,

[1] ChatGPT is a language model developed by OpenAI based on the GPT (Generative Pretrained Transformer) architecture and was trained on massive amounts of text data using unsupervised learning techniques.

are mirrored by the Twitter index. However, during the COVID-19 outbreak in spring 2020, the Twitter index deviated from actual inflation, which sharply dropped, even as our index remained stable. Interestingly, inflation expectations during this period remained elevated, aligning more with our index. This suggests our index might better capture short-term inflation expectations than actual inflation.

Comparative data from the GfK survey on inflation expectations further supports this, showing our index's close alignment with public sentiment, especially during significant events like the ECB's PEPP announcement in March 2020. Correlation analyses between our index, CPI data, and both BOP-HH and GfK inflation expectations reveal strong associations, with values ranging from 0.71 to 0.92.

When assessing the predictive power of our Twitter-based index, regression analyses indicate it offers additional explanatory value for both inflation expectations and actual CPI beyond traditional measures. This suggests that the Twitter-based index could serve as a more immediate indicator for inflation trends, supplementing existing metrics.

Specifically, using three autoregressive models, AR(1), as benchmarks, we compare their performance against models that incorporated a lag of our Twitter-based index. The BOP-HH inflation expectations data, given its shorter duration, began with 12 monthly observations from January 2011 and April 2020, with each subsequent model adding a month.

The results reveal that models incorporating our Twitter-based index consistently outperformed the benchmarks. This is evident from the Root Mean Squared Errors (RMSE) values, where a value below one signifies superior performance by the Twitter-augmented models. The most significant improvement was observed in forecasting the BOP-HH's quantitative measure of inflation expectation. Thus, our Twitter-based index effectively captures near-future consumer inflation expectations.

Our Twitter-based Inflation Index offers a distinct advantage over traditional inflation expectation sources due to its high-frequency, real-time nature. This makes it particularly valuable for tracking shifts in inflation sentiment following events like monetary policy announcements. Our analysis, using intraday changes in 2-year OIS rates around the ECB's press release window taken from Altavilla et al. (2019) and a local projection framework, reveals that the index responds notably to ECB's monetary policy shifts. For instance, a tightening monetary policy surprise sees our index decline after just over a week, suggesting the ECB can influence public inflation expectations within days.

Interestingly, while recent studies suggest households typically don't react significantly to monetary policy communication (see, *e.g.* Coibion et al., 2020b; D'Acunto et al., 2021), our findings differ. We further dissected the index to differentiate between tweets from private individuals and media organizations. The general index's response to policy announcements is predominantly driven by private individuals. Consistent with Weber et al. (2023), this is especially pronounced during high inflation periods, where households seem more attuned to inflation news and monetary policy decisions. In contrast, media organizations' tweets reflect prevailing inflation trends post-ECB announcements, rather than the specifics of the policy surprise.

In essence, during high inflation periods, households appear to be more responsive to monetary policy announcements, aligning their inflation expectations accordingly. Our index, therefore, presents a real-time tool for policymakers to assess their impact on public sentiment.

In the final part of the paper, we explore the nexus between inflation expectations and consumer behavior. Adopting a high-frequency analytical approach, we use data pertaining to online transactions spanning a diverse array of products and retailers within Germany, from January 2021 to May 2023, provided by the FinTech *Grips Intelligence*. In total, our dataset comprises observations from 379 retailers, accounting for in excess of 1 billion transactions. Our empirical findings suggest a pronounced correlation wherein elevated inflation expectations correspond with a contraction in consumer expenditure, a trend particularly salient for goods deemed discretionary in nature. Consequently and because of their economic impact, real-time knowledge of shifts in inflation expectations can offer invaluable insights, not only for policymakers but also for corporations.

Related literature. Our paper contributes to the literature on eliciting inflation expectations. Weber et al. (2022) provide a detailed description of the measurement, the determinants, and the importance of inflation expectations, focusing on surveys conducted with households and firms. Coibion et al. (2020a) analyze the implications of inflation expectations and asses their potential as a policy tool for central banks. A survey of the recent literature on inflation expectations, with a focus on households, can be found in D'Acunto et al. (2022). Andre et al. (2022) collect and investigate the narratives that people have about the macroeconomy and how these shape their inflation expectations, emphasizing the important role of the media in this process. Weber et al. (2023) show that households seem to care about inflation much more in periods when inflation is high. We corroborate their findings by showing that our index reacts to monetary policy announcements when inflation is high.

Measuring monetary policy shocks as high-frequency changes in risk-free interest rates around monetary policy announcements, we rely on a broad recent literature (see, e.g., Kuttner, 2001; Gürkaynak, Sack, and Swanson, 2005; Gertler and Karadi, 2015; Nakamura and Steinsson, 2018; Miranda-Agrippino and Ricco, 2021). Since we focus on German tweets, we use data on ECB monetary policy surprises taken from the data set by Altavilla et al. (2019) that are extended until October 2022.

Furthermore, we contribute to the literature on using new (big) data sources and artificial intelligence to answer economic questions. For example, Twitter and other text data have been used in the macroeconomic literature to, *e.g.*, measure how central bank communication is received by non-experts (Ehrmann and Wabitsch, 2022), to identify monetary policy surprises (Masciandaro et al., 2022), or to predict macroeconomic outcomes, such as economic activity (Aprigliano et al., 2022) or GDP (Barbaglia et al., 2022).

Angelico et al. (2022) explore a similar research question, as they use Italian tweets to measure inflation expectations until the end of 2019, using more traditional NLP methods. In addition to using much more novel machine learning techniques to reduce the extent of manual and, hence, error-prone steps, we explore whether Twitter can help measure consumers' expectations not only during the COVID pandemic but also in times when inflation is at levels that we have not seen for more than 70 years. Furthermore, we split our index into different user classes and show that private individuals react to monetary policy announcements.

Müller et al. (2022) focus on Germany, as well, and build an indicator to measure the media coverage of inflation based on three major newspapers. However, besides having a disproportionate hike in the index in summer 2022, their index only covers the extent to which inflation is covered in the media but does not feature any direction, that is, whether inflation is increasing or decreasing.

The rest of the paper is structured as follows: Sections 2 and 3 present our data sources and the methods to build our Twitter-based inflation index. In section 4, we compare the index with actual inflation data and consumers' inflation expectations and conduct a forecasting exercise. In section

5, we show the effects of monetary policy on our index, while Section 6 looks at the link between inflation expectations and private consumption. Section 7 concludes.

2 Data sources

Scraping and cleaning Twitter data. We download potentially relevant tweets using the Twitter Application Programming Interface (API) for academic research. We select all German-language tweets between the start of Twitter on 21 March 2006 and 31 May 2023 that contain at least one of the following keywords: *price, cost of living, high bill, inflation, expensive, gasoline price, high rent, low rent, energy costs, deflation, disinflation, sale, sell-off, low bill, low cost, cheap.*[2] This gives us a data set of 10,678,524 tweets from more than one million users including their metadata like the biography of the Twitter user, the number of likes, retweets, etc.

Although this set of tweets contains words that are important in the context of inflation, these keywords could also be used in another context. For example, the word price could likely be related to any form of advertisement or another context that could be considered as noise for our purposes. Therefore, we need to reduce the amount of noise in our data set, as we describe below.

Other data sources. To compare our Twitter-based index with survey measures of households' inflation expectations, we use two monthly data sources. The first one is the Bundesbank Online Panel Households (BOP-HH) from Deutsche Bundesbank, which is an online survey conducted regularly since April 2020 once a month with a sample size for each wave between 2,500 and 5,000 individuals.[3] As a measure of inflation expectations, we take the weighted mean of the answers to the question: "What do you think the rate of inflation/deflation will roughly be over the next twelve months?" We trim the data below -12% and above 12% as it is usually done by Deutsche Bundesbank.[4] For the German aggregate, we use data provided in Scientific Use Files. For the regional analyses, we access the data on site to be able to match it with information on the respondents' locations. For comparison, we also use the BOP-HH survey measure of inflation perceptions, which is available on a quarterly basis until the end of 2021 and on a monthly basis

[2]The original list of German keywords can be found in Appendix A.

[3]Data DOI: 10.12757/Bbk.BOPHH.202204.01, see Schmidt et al. (2022)

[4]We also run robustness checks for all our analyses for which we trim the data below -24% and above 24% and the results are very similar (available on request).

from January 2021 on. The exact question is: “What do you think the rate of inflation or deflation in Germany was over the past twelve months?” We trim the data below -12% and above 12%, as well.

As the BOP-HH data only starts in April 2020, we use as a second source for inflation expectations the micro data underlying the GfK Consumer Climate Indicator for Germany. In a monthly survey, GfK asks 2,000 individuals questions about inflation expectations, buying propensities, and personal economic situations. We are interested in the answer to the question: “How will consumer prices evolve during the next twelve months compared to the previous twelve months?" In contrast to the BOP-HH, the respondents do not answer an exact number but give a qualitative answer.[5] As a measure of inflation expectations, we take the share of respondents that answered “Prices will increase more".

Monthly data on realized consumer price inflation (CPI) in Germany is obtained from Destatis, the Federal Statistical Office of Germany.

Data on ECB monetary policy surprises are taken from the data set by Altavilla et al. (2019) that are extended until October 2022.

3 Building the Inflation Index

Building our Twitter-based inflation index consists mainly of three steps, once we have downloaded all the tweets that are potentially related to inflation. First, we clean the remaining tweets from noise to pin down the ones that are actually about inflation, using topic modeling. Then, we classify the tweets that belong to an inflation topic as referring to increasing or decreasing inflation. Finally, we aggregate the tweets of these two classes into a daily inflation index.

3.1 Selecting the relevant tweets

To filter out tweets that are not relevant for our purposes, we mainly rely on topic modeling in order to detect the topics that actually deal with inflation. By cleaning the tweets from topics that are not directly related to inflation, we also drop a large part of Twitter bot activity that is

[5]Specifically, the respondents could answer, “Prices will increase more," “Prices will increase by the same," “Prices will increase less," “Prices will stay the same," or “Prices will decrease."

often related to advertisement.[6] However, before we split the tweets into different topics, we exploit one specific feature of Twitter bots, namely that they often either repeat the exact tweets or post extremely similar ones. For example, we noticed that there are many tweets that contain the same sentence(s), added by a link that varies from tweet to tweet. Therefore, we first clean the tweets, removing hashtags, user mentions, unnecessary white spaces and, importantly, links and then drop duplicates based on the cleaned text. This simple approach leads already to more than 2 million or around 20% of the tweets being removed. With this approach we ensure that we keep the relevant information in these bot activities that could influence people's inflation expectations once but avoid that our index becomes biased by them.

To split the remaining tweets into different topics, we use the topic modeling technique BERTopic (Grootendorst, 2022) that is based on pre-trained transformer-based language models.[7] Besides creating more coherent and better interpretable topics than other topic models such as, e.g., Latent Dirichlet Allocation (LDA), BERTopic has the advantages that it can handle large volumes of text data and is computationally efficient and flexibly adjustable to a specific setting.[8] The pipeline within BERTopic consists of three main steps[9]: First, the text has to be converted to a vector representation—the embedding—for which pre-trained transformer-based models are used. One of the most commonly used models and the basis for many other state-of-the-art models is the language model BERT, which was developed by Google, uses artificial neural networks, and has been trained on very large data sets, including more than 50 languages.[10] By default, BERTopic uses the SBERT model, which is a modification of the BERT architecture that is specifically designed to generate high-quality sentence embeddings. However, because of our specific application context, we use for the embeddings the TwHIN-BERT model by Zhang et al. (2022).[11] As this model is

[6]Twitter bots are software applications that control Twitter accounts via the Twitter API and are able to tweet autonomously. As they can perform many Twitter activities, such as tweeting, re-tweeting, liking, or direct messaging on a very large scale, they can exert great influence on many users and can be used for, *e.g.*, (mis)information campaigns or advertising purposes.

[7]Transformer-based language models are neural network architectures that use attention mechanisms to selectively process different parts of input text, and have achieved state-of-the-art performance on a wide range of NLP tasks.

[8]For example, in contrast to more traditional models like LDA, BERTopic uses graphics processing units (GPU) instead of central processing units (CPU), making it much more time efficient, which is a relevant advantage given our relatively large data set.

[9]For a more detailed description of the single steps involved in BERTopic, see Grootendorst (2022).

[10]BERT stands for Bidirectional Encoder Representations from Transformers.

[11]TwHIN-BERT is based on the BERT architecture and part of the Hugging Face's Transformers library that provides a range of pre-trained transformer-based models for various natural language processing tasks. Hugging Face is a company and open-source community that develops and maintains NLP tools and frameworks (see https://huggingface.co/).

specifically trained on seven billion tweets covering over 100 distinct languages, it should be well suited to analyse our German-language inflation tweets.

In the second step, a dimensionality reduction is performed and the reduced embeddings are clustered.[12] It follows a tokenization step in which the text data for each topic is transformed into numerical representations. Finally, to assign a topic to each cluster, topic representations are extracted, using a class-based term frequency-inverse document frequency (c-TF-IDF) algorithm that makes the topics more easily interpretable for humans. The c-TF-IDF methods weights a term within a topic based on its importance for that cluster. Specifically, the weight $W_{t,c}$ for term t within cluster c is calculated as:

$$W_{t,c} = tf_{t,c} \cdot log(1 + \frac{A}{tf_t}),$$

where $tf_{t,c}$ is the term frequency of term t in cluster c, A is the average number of terms per class, and tf_t is the frequency of term t across all classes. The novelty in this step lies in calculating the importance of a word for a specific topic and not for the entire corpus, as it is often done, such that the single topics are more easily to interpret.

In general, the brevity of tweets could pose a challenge for topic models. Therefore, to train our model, we create documents consisting of five tweets of the same user in chronological order, leading to around 2 million documents. We clean the documents to remove web addresses, user mentions, hashtag symbols, and extra white spaces. In principle, the model is able to handle text without this cleaning process but we still perform this step as it simplifies the interpretation of the resulting topics. For the embeddings, we use ngrams between 1 and 2 words. To decide on the number of topics, we explore different numbers and finally set the number of topics to 150 as these ones seem the best to interpret and to be able to separate different topics from each other. Such a large number of topics further helps disentangling bot activity from the tweets that we are interested in.[13] We manually go through the most important tokens for these topics and select 19 topics that are dealing with inflation.

[12]For the dimensionality reduction a uniform manifold approximation and projection (UMAP) technique is applied (see McInnes et al., 2018). The embeddings are clustered using HDBSCAN, a hierarchical density-based clustering approach developed by McInnes et al. (2017).

[13]For example, there is a bot tweeting about cheap gasoline prices that now is assigned to an own topic and not to a broader topic about energy and gasoline prices.

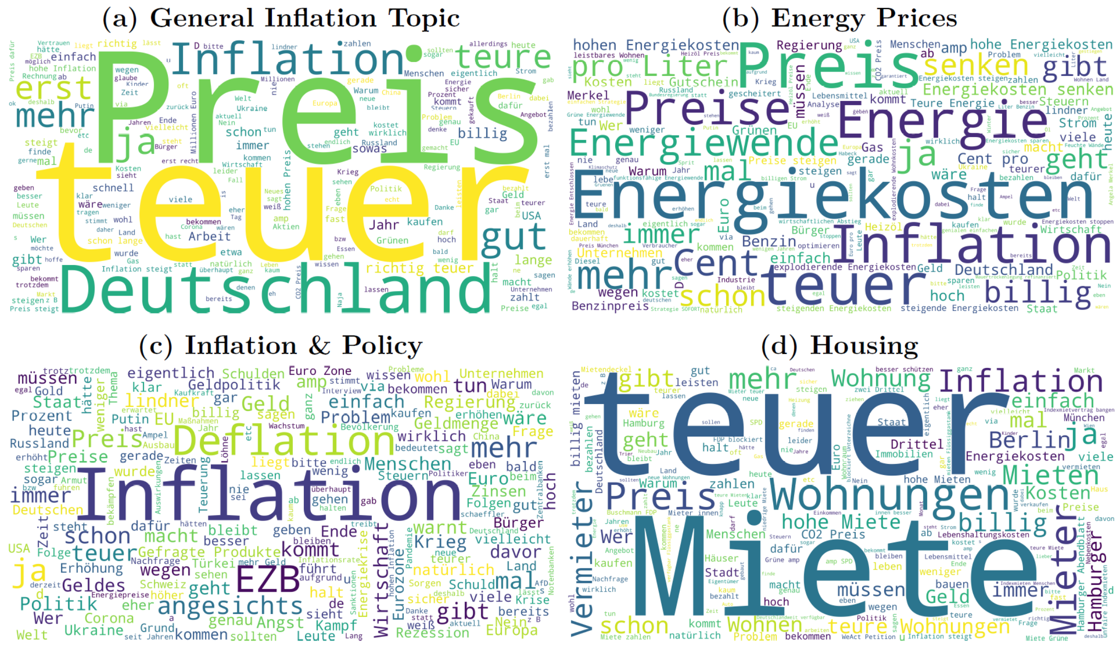


Figure 1
Grouped word clouds based on the tweets within the 19 inflation topics

In the next step, we use our trained topic model and infer the most likely topic for each tweet in our text corpus. Afterwards, we select those tweets that are assigned to one of the 19 inflation topics, which are 1,294,047 tweets, i.e. 12.12% of all the tweets we downloaded. For visualisation, we group these 19 topics into four different sub-topics, the word clouds of which are shown in Figure 1 based on the occurrences of the words in the tweets belonging to these sub-topics. The larger a word appears in a word cloud, the more important is this word for the specific topic.

According to their most important tokens, the inflation topics seem to cover a general inflation topic, a topic regarding energy prices, inflation in the context of (monetary) policy, and inflation in the context of housing. Figure 2 shows the number of tweets per year that are dealing with these four groups of inflation topics over time. The data for 2023 cover only the first five months of the year. The general inflation topic is the largest one, covering more than 60% of all the tweets related to inflation, and rising sharply in 2022. Twitter discussions on energy prices and (monetary) policy in the context of inflation are present in all the years but, naturally, became much larger in 2021 and, especially, in 2022 compare.

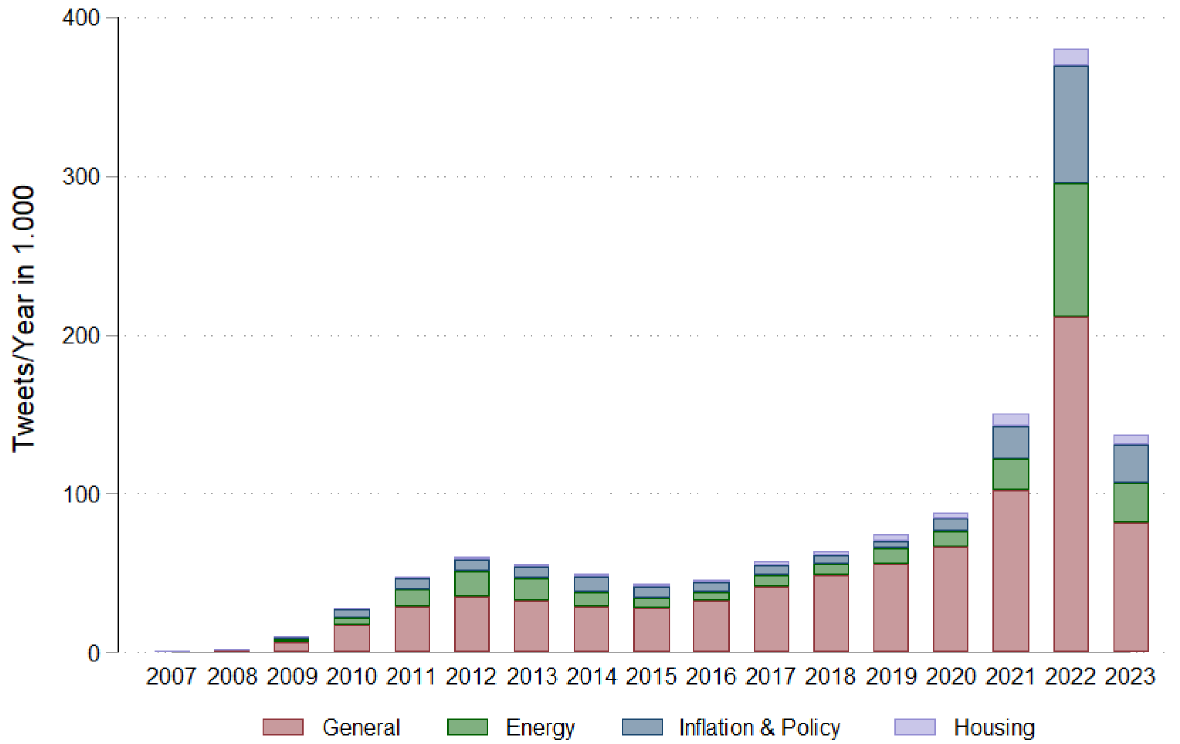


Figure 2
Number of tweets per year within the selected inflation topics over time

3.2 Classifying the tweets

We now need to classify the selected tweets based on whether they describe increasing or decreasing inflation or take a neutral stance on it. In principle, one can adopt different approaches for this task. For example, Angelico et al. (2022), create a dictionary by manually labeling bi- and trigrams (2- or 3-word sequences) as up or down, depending on whether they describe increasing or decreasing inflation. However, since we have more than 10 million different bi- and trigrams in our final text corpus, we could only label a small subset of them manually. Besides, labeling the tweets with this kind of dictionary-based approach—even if potentially added with an algorithm for combinations, special cases, etc.—has several shortcomings. First of all, it is a very subjective approach, as it depends on the researcher to either label the ngrams manually or to define rules on how to label them. In addition, this approach only focuses on specific tokens in a tweet and does not exploit their context. Hence, besides being inefficient, such an approach is likely to be error-prone and can lead to misclassifications of tweets.

Therefore, we rely on machine learning methods to label the tweets. Specifically, we use a pre-trained neural network language model which we fine-tune for our specific task, by training additional layers on top of the pre-trained model. For this additional fine-tuning step, we need a set of training data consisting of tweets and their corresponding labels for one of the three classes: *increasing prices or increasing inflation*, *decreasing prices or decreasing inflation*, and *other*.[14] Instead of labelling these tweets manually, which would be very time consuming and would need to be split among several persons who each could have an idiosyncratic bias, we use ChatGPT.[15] Specifically, we use the *gpt-3.5-turbo-0301* model through the OpenAI API to feed in a suitable set of tweets and obtain their corresponding labels until we reach an amount of 6,500 tweets for each category, respectively. We describe the exact prompt and the selection of the tweets that are labelled in Appendix B.

We use 90% of these training data to fine-tune the pre-trained neural network model TwHIN-BERT by Zhang et al. (2022), which we already used to select the relevant topics and which is explicitly trained on tweets in many languages. Note that for the classification task, we use the original tweets since the model is not only able to handle uncleaned text but can exploit the information contained in text that is often removed like special characters or emojis. During the training process, the model learns from each tweet in the training data set and adjusts its parameters after each step, where a step consists of multiple tweets with the step size decreasing throughout the training process. The parameters are adjusted in order to minimize the loss function, which measures the difference between the predicted output of the model and the actual output for the training examples.[16]

After 250 steps, the accuracy score is calculated, which is defined as the ratio of the number of correctly classified tweets to the total number of tweets in the validation set, and the current version of the model is saved if the accuracy is larger than the one from the previously saved model. We iterate this process three times for the entire training data set—corresponding to three epochs—and

[14] In German these are *steigende Preise oder steigende Inflation*, *sinkende Preise oder sinkende Inflation*, and *andere*. We do not distinguish between changes in prices and inflation as these two concepts are often mixed in people's discussions about inflation.

[15] ChatGPT is an autoregressive language model developed by OpenAI based on the GPT (Generative Pretrained Transformer) architecture and was trained on massive amounts of text data using unsupervised learning techniques. For more information on the language model GPT-3 that is currently used for ChatGPT, see Brown et al. (2020).

[16] We use cross-entropy loss as the loss function, which is defined as $L_{CE} = \sum_{i=1}^{n} t_i log(p_i)$, where t_i is the true label $\in [0, 1]$ and p_i is the predicted probability distribution.

select the best model. We evaluate this final model using the 10% of the training data that we left as a validation set. The corresponding accuracy score is 0.76, which is reached during the second epoch.[17] In addition, we evaluate the final model using a data set of around 2,000 manually labelled tweets, which gives us an accuracy score of 0.64.[18]

Using this final language model, we label all the tweets in our text corpus and obtain 377,260 tweets that deal with increasing inflation or prices, 87,671 tweets dealing with decreasing inflation or prices, and 829,116 tweets that are assigned the class *other*. This classification step serves as another filter to only keep relevant tweets about inflation. If a tweet is not clearly about increasing or decreasing prices or inflation, it will be assigned to the third class *other*.

3.3 Aggregation to index

To aggregate the labeled tweets to an inflation index, we first take the daily sums of all the tweets belonging to the classes increasing and decreasing inflation, respectively. This gives us for each day an up and a down index. To obtain a daily measure of inflation, the *Inflation Index*, we subtract the down from the up index, that is for each day, t, $InflationIndex_t = UpIndex_t - DownIndex_t$. Figure 3 shows the daily Inflation Index for the period 28 April 2007 (when the first inflation tweet occurred) until 31 May 2023. As at the beginning of the entire sample, the volume of inflation tweets is very low and only starts to pick up in 2011, we start our sample for the analyses in the remainder of this paper on 1 January 2011. In the earlier years of the sample, the index is below 0 on a number of days, especially at the end of 2014 and the beginning of 2015, when inflation in Germany was close to and, in January 2015, below 0. However, at the end of our sample, in the summer 2022, the index becomes very large, as inflation rises significantly, before decreasing again toward the end of 2022.

[17] Another commonly used evaluation metric is the precision. We we also keep track of this metric which moves very similarly to the accuracy score during the training process with values in a very similar range.

[18] We use the manually labelled data as the benchmark but the decision whether a tweet belongs to one of the three classes is often even for humans not unambiguous.

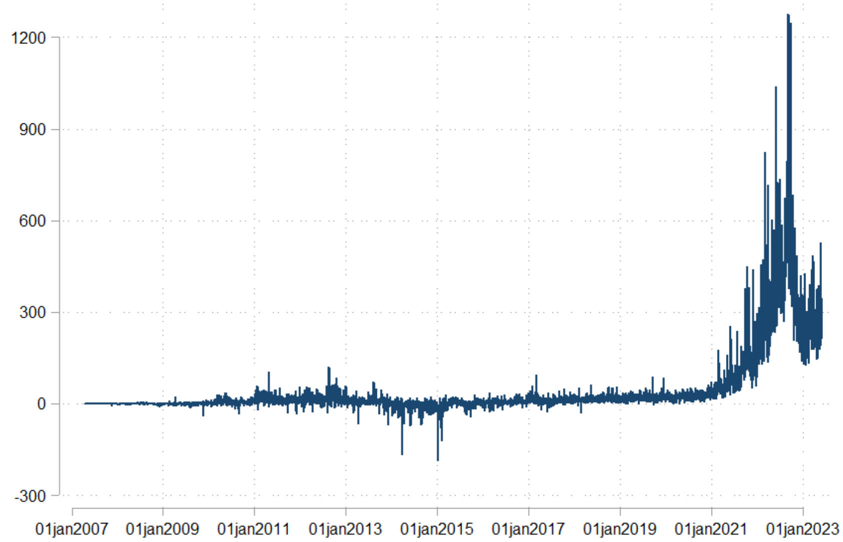


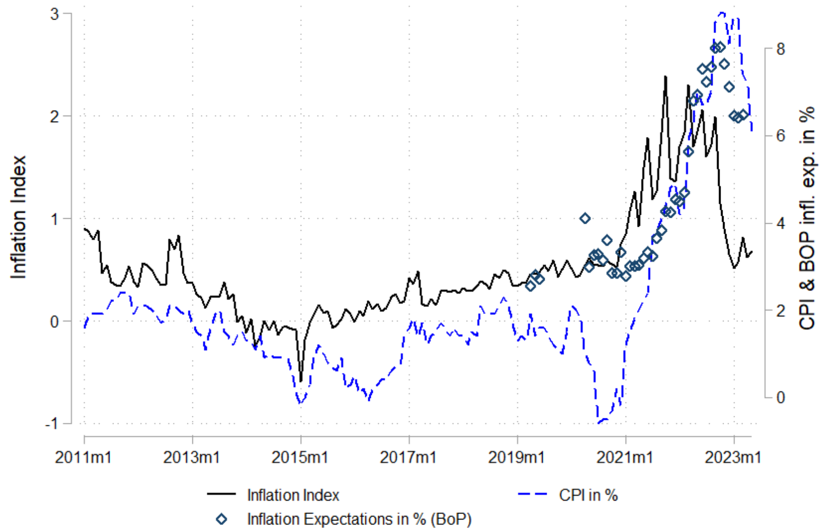
Figure 3
Twitter-based Inflation Index between April 2007 and May 2023

4 Validating the index

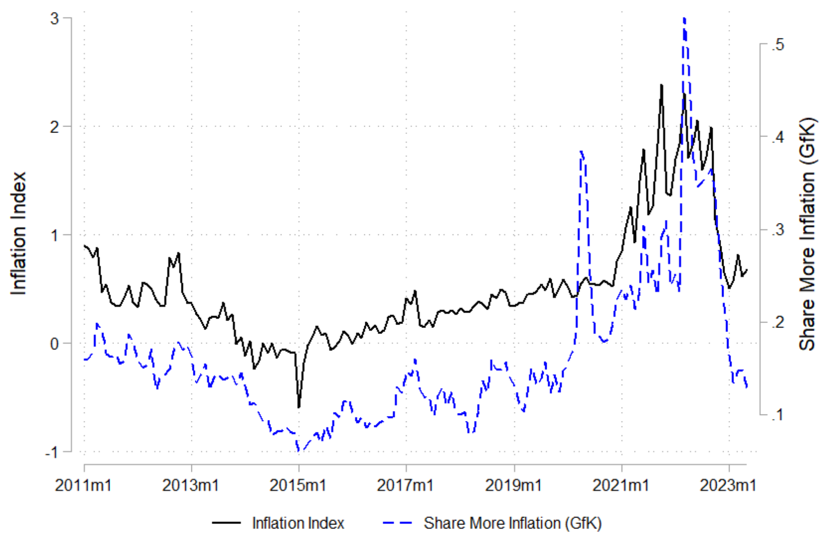
4.1 Comparison with CPI and survey data on inflation expectations

Figure 4a shows the time series of the Twitter-based Inflation Index (left y-axis) together with CPI Inflation in Germany and inflation expectations for the next 12 months from the Bundesbank Online Panel – Households (BOP-HH) (right y-axis) between January 2011 and May 2023 and between April 2019 and March 2023, respectively. Since the latter are reported on a monthly basis, we take monthly averages of our Inflation Index. Furthermore, to facilitate the visual comparison, before taking averages, we standardize the daily series by dividing it by three times the standard deviation. Here, we use the rolling window standard deviation with a window length of 10 years, as the level and the variance of our index are very different at the beginning of the sample compared to the end of the sample.[19] For the majority of our sample period, our index seems to capture the evolution of inflation very closely. Importantly, both increases in inflation, for example at the beginning of 2017, and decreases, for example at the beginning of 2015, in inflation are mimicked by the Twitter-based index. Also the large uptake of inflation from 2021 on and the cooling-off from summer 2022 on are present in our index. However, in spring 2020 (the time of the outbreak of

[19]Using the rolling window standard deviation has the additional advantage that our index depends less on the exact sample period that is used to compute it and can easily be updated.



(a) Inflation Index, CPI and BOP-HH inflation expectations



(b) Inflation Index and GfK inflation expectations

Figure 4
Twitter-based Inflation Index, CPI and inflation expectations

Inflation expectations in panel (a) are from the Bundesbank Online Panel – Households (BOP-HH), obtained from the Research Data and Service Centre (RDSC) of Deutsche Bundesbank, from April 2019 to March 2023, own calculations. Inflation expectations in panel (b) are from the GfK survey and calculated as the share of respondents expecting an increase in prices over the next 12 months.

the COVID-19 pandemic), the Twitter-based index and CPI data seem to depart from each other. Whereas our index stays in the range of the previous years and even increases a bit, actual inflation in Germany falls quite sharply to its lowest value in the entire sample.

However, Figure 4a also shows that, whereas actual inflation falls below 0% in spring 2020, inflation expectations are at 4% in April 2020 and around 3% in the subsequent months. Hence, our index evolves similarly to this measure of inflation expectations.

As the BOP-HH data on quantitative inflation expectations only start on a regular basis in April 2020 (with three pilot waves in 2019), we turn to a different measure from the GfK survey on inflation expectations. In Figure 4b, we plot the share of respondents expecting an increase in prices for the next 12 months as an alternative measure for inflation expectations. The figure shows that throughout the entire period, our Inflation Index matches the evolution of GfK inflation expectations very closely. Similar to the BOP-HH inflation expectations, the share of respondents to the GfK survey expecting an increase in prices for the next 12 month increases sharply in April and May 2020 before falling again in the subsequent months. Whereas our index does not show this large spike, it also does not fall with actual inflation, which seems to be driven to some extent by discussions on Twitter about the ECB's Pandemic Emergency Purchase Programme (PEPP) announcement on 18 March 2020.

For a further comparison of our Twitter-based Inflation Index with actual CPI data and BOP-HH and GfK data on inflation expectations, we provide correlations of the series for different time periods in Table 1. For comparison, we also provide correlations of the index with BOP-HH inflation perceptions. For the entire time period (last row), correlations of our Inflation Index with all four series are very high with values between 0.71 and 0.92. To ensure that these high correlations are not driven by the large spike in 2022, we also report correlations for shorter samples. The correlations with actual CPI data, starting in 2011 are always relatively high, except when we end the sample in 2020 due to the negative correlation in this specific year that we described above. However, as conjectured already before, also when we end the sample in 2020, the correlation between our Twitter-based indices and inflation expectations from both BOP-HH and GfK data are still relatively high. Comparing the correlations of the Inflation Index with survey measures of inflation expectations and perceptions from the BOP-HH, one notices that both are very high, but the former are even higher for most of the time. Therefore, it seems that our index captures especially inflation expectations very closely.[20]

[20]In Appendix C, we describe how we build separate Inflation Indices for different regions in Germany and plot them together with regional qualitative inflation expectations. The graphs show that the index is also highly correlated with inflation expectations on a regional level. Hence, our results for Germany are not driven by any specific region.

Table 1
Correlations of CPI, inflation expectations and perceptions with the Inflation Index

	CPI	$E_t^{GfK}\pi_{t,t+12}$	$E_t^{BOP}\pi_{t,t+12}$	$E_t^{BOP}\pi_{t-12,t}$
until Dec. 2018	0.70	0.69		
until Dec. 2019	0.63	0.66		
until Dec. 2020	0.27	0.66	0.45	
until Dec. 2021	0.67	0.68	0.74	0.77
until Dec. 2022	0.89	0.76	0.94	0.86
until May 2023	0.88	0.71	0.92	0.77

The table reports correlations between our Twitter-based Net Index and actual CPI data and inflation expectations and perceptions for different time periods. CPI data for Germany is from the Federal Statistical Office of Germany (Destatis). Inflation expectations $E_t^{GfK}\pi_{t,t+12}$ are from the GfK survey and calculated as the share of respondents expecting an increase in prices over the next 12 months. Inflation expectations $E_t^{BOP}\pi_{t,t+12}$ and perceptions $E_t^{BOP}\pi_{t-12,t}$ are from the Bundesbank Online Panel – Households (BOP-HH), obtained from the Research Data and Service Centre (RDSC) of Deutsche Bundesbank, from April 2019 to March 2023, own calculations.

4.2 Can the index explain actual CPI data and inflation expectations?

In this section, we explore whether our Twitter-based index can explain inflation expectations and actual CPI data and could, therefore, serve as an alternative and, especially, an earlier indicator. Therefore, we analyze whether the index has additional information to explain measures of inflation expectation and CPI data on top of their lagged values. Specifically, we regress both the BOP-HH and GfK inflation expectations and the actual CPI on their lagged values, one lag of CPI data, and our Twitter-based index. The latter is not lagged but, in contrast to before, we do not take the average of entire months to construct a monthly index, but only the average of the first 16 days, since 90% of the respondents only answer the BOP-HH survey after the 16th of each month. For an adequate comparison, we should therefore include the information on Twitter for the first half of a month, as well. We run four regressions, where the dependent variables are BOP-HH inflation expectations, $E_t^{BOP}\pi_{t,t+12}$, GfK inflation expectations, $E_t^{GfK}\pi_{t,t+12}$, and actual CPI, CPI_t, where for the latter, we add two sets of controls.

Table 2 reports the regression results. All variables have been standardized, such that one standard deviation increase in, *e.g.*, last month's BOP-HH inflation expectation leads to a 0.59 standard deviation increase in this month's BOP-HH inflation expectation. For all regressions, the table shows positive and significant coefficients for our Twitter-based Inflation Index, indicating that it has additional explanatory power for both inflation expectations and actual CPI beyond

Table 2
Can Twitter Inflation Index explain Inflation Expectations and actual CPI?

	$E_t^{BOP}\pi_{t,t+12}$	$E_t^{GfK}\pi_{t,t+12}$	CPI_t	CPI_t
$E_{t-1}^{BOP}\pi_{t-1,t+11}$	0.59***			-0.08
	(0.08)			(0.13)
$E_{t-1}^{GfK}\pi_{t-1,t+11}$		0.77***		0.04
		(0.05)		(0.05)
CPI_{t-1}	0.11**	-0.22***	0.87***	0.92***
	(0.04)	(0.07)	(0.03)	(0.08)
Inflation Index_t	0.20***	0.35***	0.15***	0.14**
	(0.03)	(0.07)	(0.03)	(0.06)
Constant	0.57***	0.50***	0.08***	0.09
	(0.11)	(0.10)	(0.02)	(0.21)
N	37	148	148	39
Adj. R^2	0.97	0.81	0.97	0.97

The dependent variable $E_t^{BOP}\pi_{t,t+12}$ represents monthly inflation expectations from the Bundesbank Online Panel – Households (BOP-HH) for the next 12 months, obtained from the Research Data and Service Centre (RDSC) of Deutsche Bundesbank, from April 2019 to March 2023, own calculations. The dependent variable $E_t^{GfK}\pi_{t,t+12}$ represents monthly inflation expectations from the GfK survey, calculated as the share of respondents expecting an increase in prices over the next 12 months. The dependent variable CPI_t is the monthly CPI for Germany from the Federal Statistical Office of Germany (Destatis). Our Twitter-based Inflation Index_t is calculated as described above, except that for the monthly aggregates, we take the averages of only theqee first 16 days of a given month. The time period is, except for $E_t^{BOP}\pi_{t,t+12}$, January 2011 until May 2023. All variables have been standardized and standard errors are reported in parentheses. *** denote a significance level of p < 0.01, ** p < 0.05, and * p < 0.1.

their lags and, in the case of the former, actual inflation data. Even when we regress actual CPI data on their lagged values and both measures of inflation expectation, our Inflation Index provides additional explanatory power (column four). The inclusion of our Inflation Index in these four regressions increases the Adjusted R^2 by between 0.01 and 0.03. Hence, to a small extent, our Twitter-based index provides additional information for survey-based measures of inflation beyond existing measures like lagged survey variables or actual CPI data. In addition, it can also explain actual CPI data beyond past CPI data and traditional measures of inflation expectation.

4.3 Forecasting exercise

We now explore whether our Twitter-based inflation index helps in forecasting inflation expectations and actual inflation. As benchmark models for both measures of inflation expectations and actual CPI data, we estimate three autoregressive models, AR(1), for which we choose the lag length

1 according to the BIC criterion. Since the BOP-HH inflation expectations data only covers a relatively short time period, we use as a first in-sample 12 monthly observations, starting in January 2011 and April 2020, respectively. We then incrementally add one additional month to the in-sample. In addition to these benchmark models, we estimate three corresponding competing models, for which we add to the aforementioned AR(1) models one lag of the Twitter-based Inflation Index, respectively. For all models, we forecast one to three months ahead.

Table 3
Forecasting inflation expectations and actual CPI

	$h = 1$	$h = 2$	$h = 3$
AR(1): $E_t^{BOP}\pi_{t,t+12}$, RMSE	0.619	1.061	1.541
AR(1): $E_t^{GfK}\pi_{t,t+12}$, RMSE	0.040	0.056	0.065
AR(1): $\pi_{t-12,t}$, RMSE	0.414	0.622	0.802
incl. Inflation Index : $E_t^{BOP}\pi_{t,t+12}$, relative RMSE	0.89	0.713	0.779
incl. Inflation Index : $E_t^{GfK}\pi_{t,t+12}$, relative RMSE	1.092	0.965	0.95
incl. Inflation Index : $\pi_{t-12,t}$, relative RMSE	0.994	0.896	0.896

The table reports the RMSE of the benchmark AR(1) models to forecast BOP-HH (first row), GfK inflation expectations (second row), and actual CPI data (third row) and the ratios of the RMSE of the respective model including our Inflation Index relative to the benchmark model (rows four to six) for horizon h from 1 to 3 months ahead. For the latter, values below 1 indicate that the competing model performs better than the benchmark one. A recursive estimation scheme is applied with a first in-sample of 12 observations. Inflation expectations, $E_t^{BOP}\pi_{t,t+12}$, are from the Bundesbank Online Panel – Households (BOP-HH), obtained from Research Data and Service Centre (RDSC) of Deutsche Bundesbank, from April 2020 to March 2023, own calculations. The time period is, except for $E_t^{BOP}\pi_{t,t+12}$, January 2011 until May 2023.

Table 3 shows in the first three rows the Root Mean Squared Errors (RMSE) for the three benchmark AR(1) models without our Twitter-based Inflation Index. In rows four to six, we report the RMSE of the competing models relative to the ones from the benchmark. Hence, a value below one indicates that the competing model outperforms the benchmark one. The table shows that for almost all three variables and all forecasting horizons, including our Twitter-based index improves the forecasting performance, as the relative RMSE are always below one. It improves the forecasting performance the most for the BOP-HH's quantitative measure of inflation expectation. Therefore, we conclude that our Twitter-based inflation index is well-suited to capture consumer expectations about inflation in the near future.[21]

[21] It could also be the case that it improves the forecast of actual CPI data, but we are aware that for a proper forecasting exercise of actual inflation, one would have to include many other variables, as well.

5 Can monetary policy influence inflation expectations?

5.1 Inflation expectations and monetary policy surprises

A significant benefit of our Inflation Index over traditional inflation expectation measures, such as surveys, is its high-frequency characteristic. This index offers a real-time insight into perceptions of inflation and shifts resulting from specific events. Notably, monetary policy announcements might influence these expectations. Given that guiding inflation expectations is a pivotal objective of monetary policy, assessing its efficacy at a high frequency might be particularly important for central banks.

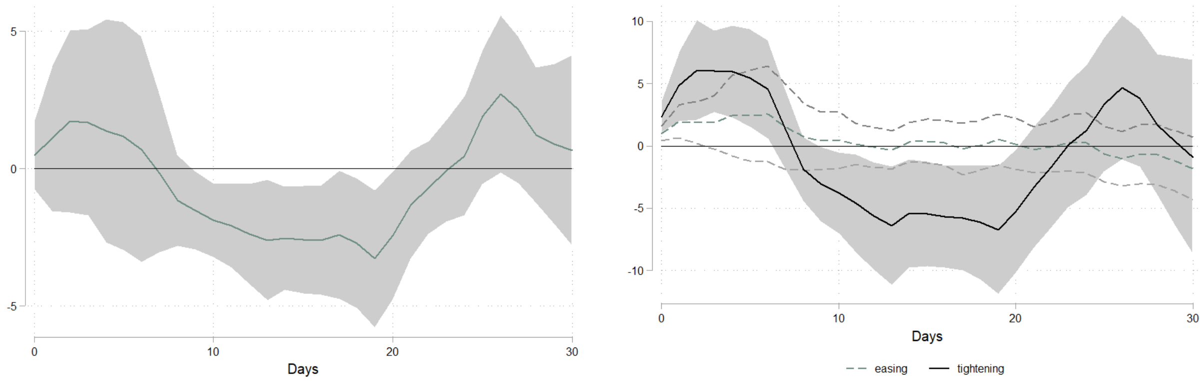
To analyze the effects of ECB monetary policy announcements on inflation expectations, we run local projections following Òscar Jordà (2005) at a daily frequency for each horizon h:

$$y_{t+h} - y_{t-1} = \alpha_h + \beta_h \epsilon_t + \gamma_h X_t + u_{t+h} \ , \tag{1}$$

where y_t is the seven-day moving average of our Inflation Index (including y_{t-6} to y_t), α_h is a horizon-specific constant, ϵ_t are intraday changes in 2-year OIS rates around the ECB's press release window taken from Altavilla et al. (2019), and X_t is a vector of control variables including up to 5 lags of the Inflation Index and last month's CPI. Standard errors are computed using the Newey and West covariance estimator, which allows for autocorrelation and heteroscedasticity. The maximum lag length is set to $h+1$, such that it increases with the horizon h. β_h then gives the response of the Inflation Index to monetary policy surprises at a horizon h.

Figure 5a illustrates that following a tightening monetary policy surprise, our Inflation Index experiences a decline just over a week later. A monetary policy surprise of 10 basis points results in a 30-point drop in our Inflation Index, translating to 30 more tweets anticipating decreasing inflation than increasing inflation. This suggests that the ECB might effectively sway public inflation perceptions within a matter of days. Up to this point, we have postulated that both easing and tightening surprises exert symmetrical impacts. To account for potential asymmetries, we will proceed with local projections, distinguishing between easing and tightening surprises:

$$y_{t+h} - y_{t-1} = \alpha_h + \beta^h_{tight} \mathbb{1}_{\{\epsilon_t > 0\}} \epsilon_t + \beta^h_{easing} \mathbb{1}_{\{\epsilon_t < 0\}} \epsilon_t + \gamma_h X_t + u_{t+h}, \tag{2}$$



(a) All surprises **(b) Easing and tightening surprises separately**

Figure 5
Impulse responses of the Twitter-based Inflation Index to monetary policy surprises

The figures show impulse responses of the seven-day moving average of the Inflation Index to 1 basis point intraday changes in 2-year OIS rates around the ECB's press release window taken from Altavilla et al. (2019). In panel (a), we show the β_h from Equation (1). In panel (b), we show $-\beta^h_{easing}$ and β^h_{tight} from Equation (2). The time period is January 2011 until December 2022. We use Newey-West standard errors and show confidence bands at the 90% level.

where β^h_{tight} gives the response of the Inflation Index to tightening monetary policy surprises at a horizon h and β^h_{easing} gives the response of the Inflation Index to easing ones. To facilitate the interpretation, we show $-\beta^h_{easing}$ in the corresponding graph. If $-\beta^h_{easing}$ is positive, this implies that our Inflation Index is increasing after a monetary policy easing surprise h days after the announcement.

Figure 5b reveals distinct impacts of the two types of shocks on our Inflation Index. While the index exhibits minimal response to easing surprises, it demonstrates a more pronounced reaction to tightening ones. Initially, following a monetary policy tightening, the Inflation Index rises for several days, only to experience a significant drop after a week. A plausible interpretation is that in the immediate aftermath of a tightening monetary policy surprise, discussions on Twitter revolve around the rationale behind the central bank's decisions, such as excessively high inflation. However, as days progress, the sentiment may shift towards confidence in the ECB's capacity to curb inflation, leading to diminished inflation expectations.

5.2 How do households respond to monetary policy during a surge in inflation?

The responsiveness of our Inflation Index to monetary policy announcements might be unexpected, especially considering recent findings suggesting that households typically exhibit limited reactions to monetary policy communications (refer, for instance, to Coibion et al., 2020b; D'Acunto et al., 2021). This prompts the inquiry: which tweets are reflected in our index, and who are the primary contributors influencing the Inflation Index's reaction to monetary policy announcements? Specifically, are we capturing the reactions of households or primarily media coverage on monetary policy decisions?

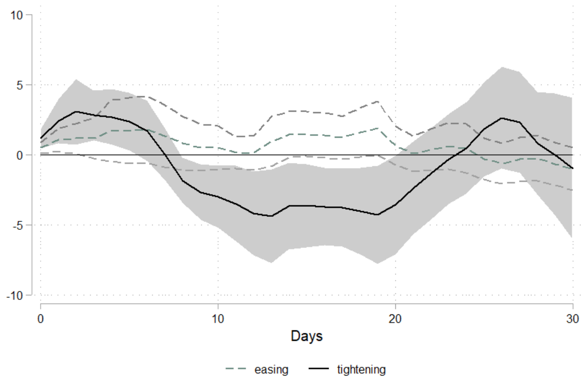
To address these questions, we create sub-indices comprising tweets on inflation from both private individuals and media entities. To categorize the accounts tweeting about inflation into groups like private individuals, media organizations, and other potential categories, we employ the pre-trained TwHIN-BERT model, refining it for this specific purpose. For the fine-tuning process's training data, we predominantly utilize the OpenAI API, supplementing it with manually curated entries for media companies.

Upon classification, we identify 64,458 accounts as private users and 4,664 as media organizations.[22] Recognizing that not all media accounts are economically oriented and thus may not be pivotal in Twitter discussions on monetary policy, we refine this list based on specific criteria: only accounts that have published a minimum of 50 tweets on inflation and boast a follower count exceeding 10,000 are retained, resulting in 86 pertinent accounts. A comprehensive overview of the classification task, the precise API prompt, and a showcase of the most significant media accounts in our dataset can be found in Appendix D.

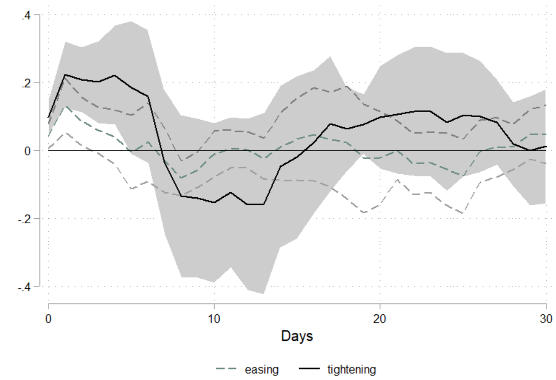
Figure 6 shows the impulse responses of the Inflation Indices for both private individuals (left panel) and media organizations (right panel), delineated by easing and tightening surprises. The response pattern for private individuals closely mirrors that of the overarching index depicted in Figure 5b. Following tightening shocks, the index rises for several days and then markedly declines after a week. In contrast, the index for media organizations exhibits a brief surge post-tightening shocks without any subsequent statistically significant fluctuation. This suggests that the previously observed response of the general Inflation Index to monetary policy announcements is predominantly influenced by private individuals.

The apparent responsiveness of private individuals to monetary policy announcements seems

[22]For the classification task, we delineated six user categories: *private individuals*, *individual journalist*, *influencer*, *media organization*, *business organization*, and *other organization*. Notably, the classes of private individuals and media organizations, which are our primary focus, appear to be more accurately classified than the rest, leading us to concentrate on these two.



(a) Private individuals



(b) Media organisations

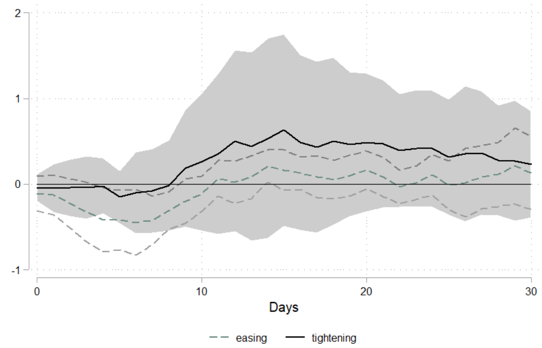
Figure 6
Impulse responses of the Twitter-based Inflation Indices of private individuals and media organizations to monetary policy surprises

The figures show impulse responses of the seven-day moving averages of the Inflation Indices of private individuals and media organizations to 1 basis point intraday changes in 2-year OIS rates around the ECB's press release window taken from Altavilla et al. (2019). We show $-\beta^h_{easing}$ and β^h_{tight} separately from Equation (2). The time period is January 2011 until December 2022. We use Newey-West standard errors and show confidence bands at the 90% level.

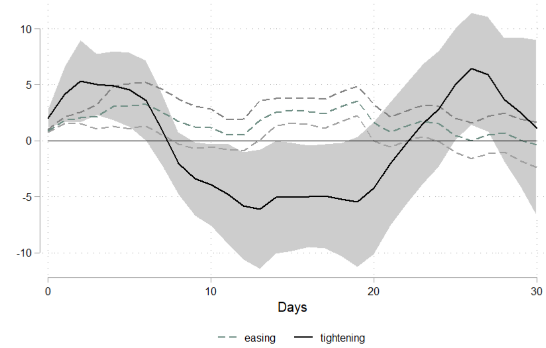
at odds with the literature previously referenced. Yet, Weber et al. (2023) offers recent findings suggesting that during periods of high inflation, households are notably more informed about inflation dynamics. Given that individuals tend to be more concerned about inflation during these times, it's logical they would be more attuned to inflation-related news, including monetary policy decisions. Consequently, we bifurcate the response of the Inflation Index for private individuals into two distinct phases: 2011 to 2020, representing a low inflation period, and 2021 to 2022, indicative of a high inflation era.

The resulting impulse responses are depicted in Figure 7. The top two charts represent the households' index responses, while the bottom two correspond to media organizations. For households, the left panel 7a reveals that during the 2011-2020 period of low inflation, the Inflation Index exhibits minimal reaction to both easing and tightening surprises, with a slight dip observed 5-7 days post an easing shock. Conversely, the right panel 7b during the high inflation period of 2021-2022, demonstrates a significant index response to both types of shocks. In the week following a monetary policy announcement, the index rises for both easing and tightening events.[23] The subsequent patterns align with anticipated policy outcomes: the index ascends post easing surprises

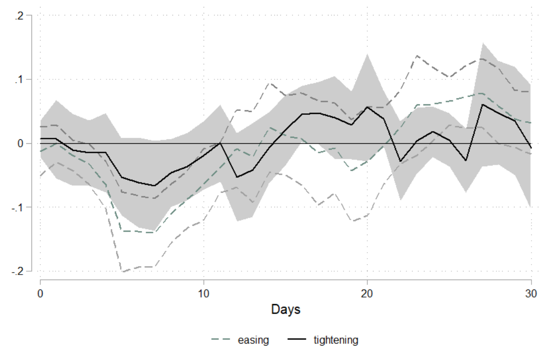
[23]This trend might partially be attributed to our choice of plotting the seven-day moving average of our index, given the potential variability in Twitter activity across different days. Depending on the specific announcement day and concurrent events, reaction times might differ, justifying a smoothed series over several days.



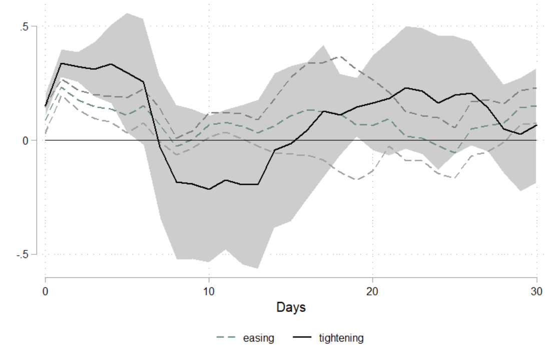
(a) Private: 2011-2020



(b) Private: 2021-2022



(c) Media: 2011-2020



(d) Media: 2021-2022

Figure 7

Impulse responses of the Twitter-based Inflation Index of private individuals and media organizations to monetary policy surprises

The figures show impulse responses of the seven-day moving averages of the Inflation Indices of private individuals and media organizations to 1 basis point intraday changes in 2-year OIS rates around the ECB's press release window taken from Altavilla et al. (2019) for two different time periods: 2011 until 2020 (panels 7a and 7c) and 2021 until 2022 (panel 7b and 7d). We show $-\beta^h_{easing}$ and β^h_{tight} separately from Equation (2). The overall time period is January 2011 until December 2022. We use Newey-West standard errors and show confidence bands at the 90% level.

and descends following tightening ones, suggesting that monetary policy announcements effectively influence household inflation expectations. In contrast, media organizations exhibit pronounced responses only within the initial week post a monetary policy announcement. During the low inflation phase, their index diminishes after easing shocks. In the high inflation period, it escalates post both easing and tightening events. This behavior implies that media entities tend to tweet about low inflation/deflation during low inflation times and about heightened inflation during high inflation periods post ECB's proclamations, irrespective of the magnitude of the monetary policy surprise.

In summary, our findings indicate that households have recently become more attuned to and influenced by monetary policy announcements. Particularly during periods of elevated inflation that impact daily living, monetary policy appears to effectively shape their inflation expectations within a matter of days. Our index can act as a real-time barometer for monetary policymakers to assess their reach and influence on households.

6 Inflation expectations and consumer behavior

Having documented the capability to gauge inflation expectations with high-frequency precision and their responsiveness to monetary policy shifts — particularly during heightened inflationary periods — our final inquiry is its tangible impact on the real economy, specifically focusing on the interplay between inflation expectations and household consumption dynamics.

Theoretically, multiple channels might explain the link between inflation expectations and household consumption. One potential channel, the substitution channel, might prompt an uptick in consumption in the face of rising inflation expectations. Conversely, concerns about the potential erosion in real income might compel individuals to curtail their expenditure when they anticipate higher inflation. In this segment, we scrutinize the correlation between consumption patterns across diverse product segments and our inflation index. This analysis leverages data on e-commerce transactions spanning a vast spectrum of products and retailers in Germany from January 2021 to May 2023. This dataset is obtained from the FinTech *Grips Intelligence*, which monitors e-commerce sales across 5,000 product categories, encompassing 30,000 retailers and brands globally.

We focus on data from firms that consistently report their sales to prevent sample bias due to firms' entry or exit. Specifically, we only include firms operating in Germany that have reported for at least 26 of the 29 months in our dataset, resulting in 379 firms accounting for over 1 billion product purchases. These products are segmented into 21 primary categories, each with various subcategories. Figure 8 illustrates these 21 categories and their respective proportions of total purchases. Notably, "Home & Garden" dominates, with "Apparel & Accessories"—primarily clothing—trailing as the second-largest category.

To shed light on the short-term relation between inflation expectations and consumption behavior,

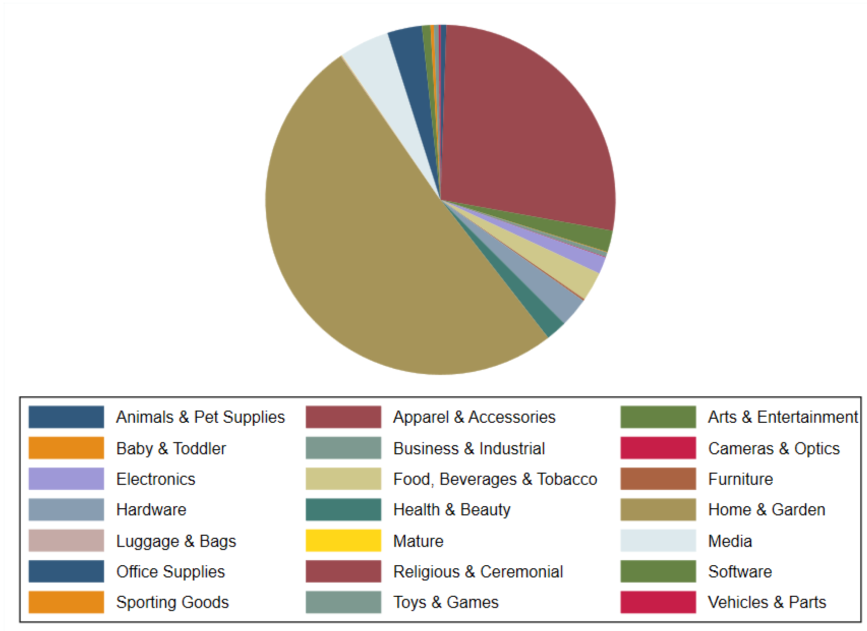


Figure 8
Shares of the number of purchased products across different product categories

we regress the number of purchased products in a category j on a constant and our Inflation Index, as described by equation (3):

$$\text{No. of purchased products}_j = \alpha + \beta_j \text{Inflation Index} + \epsilon_j \tag{3}$$

For both the number of goods purchased and our index, we utilize the seven-day moving averages, encompassing the current day and the preceding six days, and then standardize these values. The correlation findings can be found in Appendix E, specifically in Table 8. As consumption varies significantly across the 21 product categories, we sort them into three different groups according to their level of discretion.[24]

The coefficients β_j from equation (3) for the three different levels of discretion are shown in Table 4. The number of purchased good that are most discretionary are negatively correlated with *Inflation Index*. Specifically, when the index increases by one standard deviation, consumption of products like clothes, cameras or toys decreases by 0.34 standard deviations. Consumption in the

[24]See the notes to Table 4. Differentiating between durable and non-durable goods, as done by Coibion et al. (2022), would be insightful. However, given our data's focus on online purchases, few items pertain to immediate consumption, such as grocery shopping. Predominantly, the products lean towards intermediate or long-term use, such as clothing, tableware, or cameras. Even within the "Food, Beverages & Tobacco" category, the items are not daily essentials but more specialized goods such as liquors, wines, and teas.

Table 4
Correlations of our index and item quantity purchased for different levels of discretion - standardized (7-day moving averages)

	Discretion level 1	Discretion level 2	Discretion level 3
Inflation Index	-0.34***	-0.10***	0.01
	(0.03)	(0.03)	(0.03)
Constant	2.98***	0.34***	0.70***
	(0.06)	(0.06)	(0.05)
N	881	881	881
Adj. R^2	0.11	0.01	0.00

The table reports the coefficients β_j from equation (3) for the different levels of discretion. We group the 21 product categories into three different levels of discretion as follows Level 1: Cameras & Optics, Electronics, Apparel & Accessories, Arts & Entertainment, Health & Beauty, Mature, Media, Sporting Goods, Toys & Games; level 2: Food, Beverages & Tobacco, Home & Garden, Furniture, Luggage & Bags, Software, Vehicles & Parts, Baby & Toddler, Business & Industrial, Hardware, Office Supplies; level 3: Animals & Pet Supplies, Religious & Ceremonial. For each level of discretion, we sum up the total amount of purchased products, take the seven-day moving average including the current day and standardize the series. The time period is 01 January 2021 until 31 May 2023 and the sample includes 379 firms.

next level of discretion is negatively correlated with our index, as well, but the effect is muted. In contrast, consumption of less discretionary goods like pet supply, which is the vast majority in this group, is uncorrelated with our index.

Overall, this section indicates that rising inflation expectations correlate with reduced consumption, particularly for more discretionary goods. Given that shifts in inflation expectations can swiftly impact the real economy, real-time awareness of these changes is crucial for both policymakers and businesses.

7 Conclusion

Inflation expectations are crucial in determining people's consumption and investment behavior and, therefore, the evolution of the economy. Yet, their measurement is usually costly, time-consuming, and only available with a considerable time lag. In the age of big data and artificial intelligence, the natural question arises whether we can exploit new methods and data to provide an alternative assessment of consumers' expectations.

In this paper, we use natural language processing techniques to develop an inflation index based on German-language tweets. We apply state-of-the-art machine learning methods to clean tweets about prices and inflation from potential noise and to specify whether a tweet discusses increasing or decreasing inflation. Based on the difference between increasing and decreasing inflation tweets, we build an inflation index that exhibits high correlations with actual inflation data and, especially, inflation perceptions and expectations measures from consumer expectations surveys.

In exercises analyzing the informativeness and forecasting power of our Inflation Index, we find that it adds additional explanatory power beyond existing measures and helps in forecasting future values, especially for quantitative inflation expectations. Therefore, it seems that our Twitter-based index could, indeed, serve as a valuable real-time indicator for households' short-term expectations of inflation.

In addition to an aggregate German Inflation Index, we also build regional indices, based on Twitter users' information on their locations. We find that our index matches the evolution of inflation expectations on a regional level, as well.

To exploit the advantage that our Inflation Index is available on a high frequency, we analyze the effects of monetary policy announcements on our index within the next 30 days. We find that, after a monetary policy tightening, our index decreases after around one week. This response is driven by tightening shocks and indicates that the ECB is indeed able to lower inflation expectations after a tightening surprise.

After classifying Twitter users into different user categories, we find that it is private individuals who respond to monetary policy announcements. Especially in the past two years, when inflation was high, households seem to react significantly to both easing and tightening surprises, in line with recent literature showing that households care more about inflation, and, hence, should follow any news affecting it, when inflation is relatively high. These findings further underline the potential usefulness of a real-time indicator of inflation expectations for monetary policymakers to gauge their effects on households' expectations.

References

Altavilla, C., L. Brugnolini, R. S. Gürkaynak, R. Motto, and G. Ragusa (2019). Measuring Euro Area Monetary Policy. *Journal of Monetary Economics 108*, 162–179.

Andre, P., I. Haaland, C. Roth, and J. Wohlfart (2022). Narratives about the Macroeconomy. *mimeo*.

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

Aprigliano, V., S. Emiliozzi, G. Guaitoli, A. Luciani, J. Marcucci, and L. Monteforte (2022). The Power of Text-based Indicators in Forecasting Italian Economic Activity. *International Journal of Forecasting*.

Barbaglia, L., S. Consolia, and S. Manzan (2022). Forecasting GDP in Europe with Textual Data. *Mimeo*.

Brown, T. B., B. Mann, N. Ryder, M. Subbiah, J. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, S. Agarwal, A. Herbert-Voss, G. Krueger, T. Henighan, R. Child, A. Ramesh, D. M. Ziegler, J. Wu, C. Winter, C. Hesse, M. Chen, E. Sigler, M. Litwin, S. Gray, B. Chess, J. Clark, C. Berner, S. McCandlish, A. Radford, I. Sutskever, and D. Amodei (2020). Language models are few-shot learners. *arXiv preprint: arXiv:2005.14165*.

Coibion, O., Y. Gorodnichenko, S. Kumar, and M. Pedemonte (2020a). Inflation Expectations as a Policy Tool? *Journal of International Economics 124*, 103297. NBER International Seminar on Macroeconomics 2019.

Coibion, O., Y. Gorodnichenko, S. Kumar, and M. Pedemonte (2020b). Inflation Expectations as a Policy Tool? *Journal of International Economics 124*, 103297.

Coibion, O., Y. Gorodnichenko, and M. Weber (2022). Monetary policy communications and their effects on household inflation expectations. *Journal of Political Economy 130*(6), 1537–1584.

D'Acunto, F., D. Hoang, and M. Weber (2021, 07). Managing Households' Expectations with Unconventional Policies. *The Review of Financial Studies 35*(4), 1597–1642.

D'Acunto, F., U. Malmendier, and M. Weber (2022). What Do the Data Tell Us About Inflation Expectations? *NBER Working Paper 29825*.

ECB (2021). Inflation Expectations and their Role in Eurosystem Forecasting. *ECB Occasional Paper Series 264*.

Ehrmann, M. and A. Wabitsch (2022). Central Bank Communication with Non-Experts – A Road to Nowhere? *Journal of Monetary Economics 127*, 69–85.

Gertler, M. and P. Karadi (2015). Monetary policy surprises, credit costs, and economic activity. *American Economic Journal: Macroeconomics 7*(1), 44–76.

Goldfayn-Frank, O. and J. Wohlfart (2020). Expectation formation in a new environment: Evidence from the german reunification. *Journal of Monetary Economics 115*, 301–320.

Grootendorst, M. (2022). BERTopic: Neural Topic Modeling with a Class-Based TF-IDF Procedure. *arXiv preprint 2203.05794*.

Gürkaynak, R., B. Sack, and E. Swanson (2005). Do actions speak louder than words? the response of asset prices to monetary policy actions and statements. *International Journal of Central Banking 1*(1), 55–93.

He, P., J. Gao, and W. Chen (2021). DeBERTaV3: Improving DeBERTa using ELECTRA-Style Pre-Training with Gradient-Disentangled Embedding Sharing. *arXiv 2111.09543*.

Kuttner, K. (2001). Monetary policy surprises and interest rates: Evidence from the fed funds futures market. *Journal of Monetary Economics 47*(3), 523 – 544.

Laurer, M., W. van Atteveldt, A. S. Casas, and K. Welbers (2022). Less Annotating, More Classifying – Addressing the Data Scarcity Issue of Supervised Machine Learning with Deep Transfer Learning and BERT - NLI. *mimeo*.

Masciandaro, D., D. Romelli, and G. Rubera (2022). Monetary Policy and Financial Markets: Evidence from Twitter Traffic. *Bocconi Working Paper Series 160*.

McInnes, L., J. Healy, and S. Astels (2017). hdbscan: Hierarchical Density Based Clustering. *Journal of Open Source Software 2*, 205.

McInnes, L., J. Healy, and J. Melville (2018). UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *arXiv preprint 1802.03426*.

Miranda-Agrippino, S. and G. Ricco (2021). The Transmission of Monetary Policy Shocks. *American Economic Journal: Macroeconomics*.

Müller, H., T. Schmidt, J. Rieger, L. M. Hufnagel, and N. Hornig (2022). A German Inflation Narrative - How the Media frame Price Dynamics: Results from a RollingLDA Analysis. *DoCMA Working Paper 9*.

Nakamura, E. and J. Steinsson (2018). High-Frequency Identification of Monetary Non-Neutrality: The Information Effect. *The Quarterly Journal of Economics 133*(2), 1283–1330,.

Schmidt, T., D. Boddin, S. Helmschrott, O. Goldfayn-Frank, M. Cato, and et. al. (2022). Bundesbank-Online-Panel-Households (BOP-HH). Version: 01. *Deutsche Bundesbank Dataset*.

Weber, M., B. Candia, T. Ropele, R. Lluberas, S. Frache, B. Meyer, S. Kumar, Y. Gorodnichenko, D. Georgarakos, O. Coibion, G. Kenny, and J. Ponce (2023). Tell Me Something I Don't Already Know: Learning in Low and High-Inflation Settings. *mimeo*.

Weber, M., F. D'Acunto, Y. Gorodnichenko, and O. Coibion (2022, August). The Subjective Inflation Expectations of Households and Firms: Measurement, Determinants, and Implications. *Journal of Economic Perspectives 36*(3), 157–84.

Zhang, X., Y. Malkov, O. Florez, S. Park, B. McWilliams, J. Han, and A. El-Kishky (2022). Twhin-bert: A socially-enriched pre-trained language model for multilingual tweet representations. *arXiv preprint arXiv:2209.07562*.

Òscar Jordà (2005). Estimation and Inference of Impulse Responses by Local Projections. *American Economic Review 95*(1), 161–182.

A Downloaded Tweets—Details

We use the following German keywords to select the tweets that we download for our initial data set: preis OR Lebenshaltungskosten OR (hohe rechnung) OR inflation OR teure OR teuer OR benzinpreis OR (hohe miete) OR (niedrige miete) OR energiekosten OR deflation OR disinflation OR schlussverkauf OR abverkauf OR (niedrige rechnung) OR (niedrige kosten) OR billig.

B Details on generating the training data set for the tweet classification

To generate training data, we use the the *gpt-3.5-turbo-0301* model through the OpenAI API to label a given set of tweet. The exact prompt is the following:

Classify this German tweet into one of the three categories: 'steigende Preise oder steigende Inflation oder hohe Inflation', 'sinkende Preise oder sinkende Inflation oder niedrige Inflation' or 'andere'. Confidence of prediction (COP): low, medium, high. The tweet is: '[tweet]' Output expected in the form of - Label: xxx, Explanation: xxx, COP: xxx.

Consider, for example, the following tweet:

#inflation steigt weiter in die Höhe und dann verkündet #Habeck diese Woche auch noch Alarmstufe Gas. Wir müssen das bezahlen was die Politiker vergeigt haben. Sie mussten unbedingt Sanktionen gegen Russland verhängen. Quelle des Screenshots: https:[...]

This tweet leads to the following response: Label: *steigende preise oder steigende inflation*, Explanation: *The tweet mentions inflation rising and the need to pay for mistakes made by politicians*, COP: *medium*.

We set the system role to *You are an economist* and transform the responses in a parsable manner. We add the requirement of giving a reason for the choice of the class for a specific tweet to better judge the ability of ChatGPT to classify the inflation tweets. We manually browse randomly through a subset of the labelled tweets and can confirm that ChatGPT is able to understand the main messages of the tweets—at least comparable to a human, as sometimes even for humans, it is not easy to decide on the correct category of a tweet.

We aim at generating a balanced training data set with equal amounts of tweets for each class.

However, since in reality the three classes are heavily unbalanced with the low inflation class being much smaller than the others, this is not straightforward to achieve.

Therefore, we first perform a zero-shot classification method, for which we do not have to provide any annotated training data, but only the three classification labels. Specifically, we use the machine learning algorithm "mDeBERTa-v3-base-mnli-xnli" by Laurer et al. (2022), which aims at understanding many different languages and performing any kind of classification task. This algorithm is a fine-tuned version of the "DeBERTaV3-base" transformer by He et al. (2021) from Microsoft, which is an improved version of the original BERT language model. The idea of this algorithm is that it is pre-trained so well that it does not need any or only very few additional annotated data to learn how to perform a given task.[25] Since the model was fine-tuned using different languages, we can directly apply it to our German-language tweets. Applying this zero-shot classification algorithm to our set of tweets, returns 447,924 tweets that deal with increasing inflation, 73,259 tweets dealing with decreasing inflation, and 234,516 tweets that do not indicate the direction of inflation.

To evaluate the quality of this approach, we calculate the accuracy score based on the manually labelled set of tweets, which is 0.46—hence, significantly lower than the one for our final fine-tuned model.

In addition to the respective label for each tweet, the algorithm also returns a score which indicates how certain the decision for a specific task is, *i.e.* how large the prediction probability is. We use this score to sort the tweets within each category in descending order based on the model's certainty that a specific tweets belongs to a certain class. Besides having equally distributed classes in the training data set, it is also important to include both easy and hard examples, where the former can be expected to have a high zero-shot prediction probability and the latter a low one. Therefore, after first choosing randomly 2,000 tweets from each class, we add another 2,000 tweets per class from both easy and hard examples in a 1:3 ratio, resulting in 12,000 tweets. We feed more difficult classification problems to ChatGPT to not waste API calls on those tweets that are

[25]The original BERT model was trained on 16 gigabytes of books and Wikipedia texts, added by 145 gigabytes data on news articles, links on Reddit, and story-like texts for the "DeBERTaV3-base" transformer by Microsoft. In addition, for the "mDeBERTa-v3-base-mnli-xnli" model, the previous model's pre-training is further fine-tuned by using more than a million classification examples from different Natural Language Inference (NLI) data sets—specifically, the English "MNLI" data set and the "XNLI" data set, containing 15 different languages, including German. For further details, see Laurer et al. (2022).

very easy to classify, anyway. Since this approach still returns much more tweets belonging to the increasing or other class, we add more tweets to the ChatGPT prompt in a different ratio. We now add 8,000 tweets from the increasing and other class based on the zero-shot algorithm's classification and 16,000 from the decreasing class. Afterwards, we have 6,500 tweets belonging to the decreasing class and take as many tweets from the the other two classes [which ones? randomly?], as well to have a balanced final training data set.

C Regional Inflation Indices

We also explore whether there are regional differences in our Twitter-based Inflation Index and, if so, whether these are mirrored in survey-based inflation expectations. For the latter, we use the information on the state of the respondent in the GfK data and calculate weighted averages for each state in Germany. To obtain geographical information on Twitter users, we exploit information from the users' descriptions. More than 61% of all the users in our sample provide some information about their location. As this information is not standardized (users can write anything they want), we use an API called *Nominatim* to obtain the exact name of a given location, as long as this location exists. In Figure 9, we show a map with the locations of the tweets that deal with inflation. As some users in our sample write in their user description that they are based in Germany, without further indicating the region, we dropped them from this regional sample as they would be mapped to the center of Germany belonging to Thuringia. The size of the bubbles in the figure indicates the number of tweets from a specific location.

To match the locations to our inflation tweets and build regional Inflation Indices, we have to make two assumptions: First, we assume that the vast majority of the users indicate their true location. Second, as for each user, there is only one (the latest) location assigned to all her/his tweets during the entire period, we assume that the share of people who changed their place of living during this period is not too high. We aggregate our state-level indices into four regional indices by taking the sum across the respective states of up and down tweets before generating the final indices. For inflation expectations, we take the sum of all the respondents expecting higher prices that live in the respective region.

Figure 10 shows both our Inflation Index and the GfK measure for inflation expectations, which

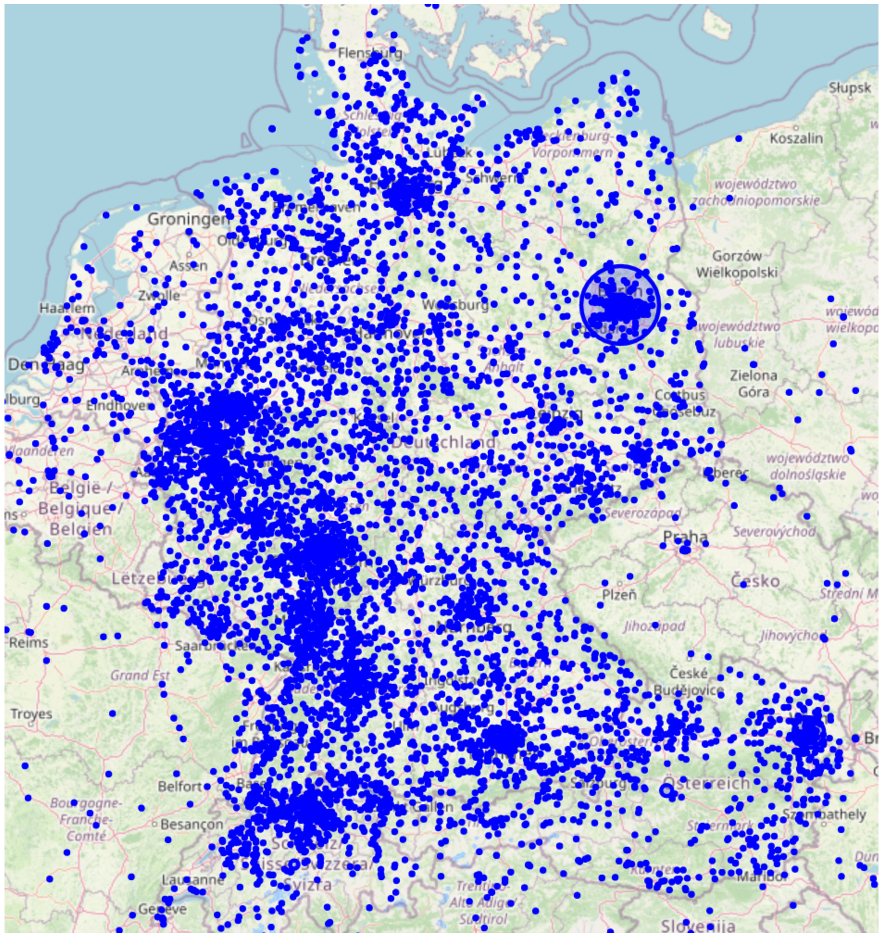


Figure 9
Locations of inflation tweets

is the share of respondents expecting an increase in prices for the next 12 months, for four regions: east (Mecklenburg-Western Pomerania, Brandenburg, Berlin, Saxony, Saxony-Anhalt, Thuringia), north (Hamburg, Schleswig-Holstein, Bremen, Lower Saxony), south (Hesse, Baden-Wuerttemberg, Bavaria), and west (North Rhine-Westphalia, Rhineland-Palatinate, Saarland).

The figure shows that the share of respondents expecting higher prices is higher in the east of Germany than in the other regions. This is in line with findings of, *e.g.* Goldfayn-Frank and Wohlfart (2020). However, the general evolution of inflation expectations over time is relatively similar across the regions. Importantly, one can see that our index matches inflation expectations closely on a regional level, as well.

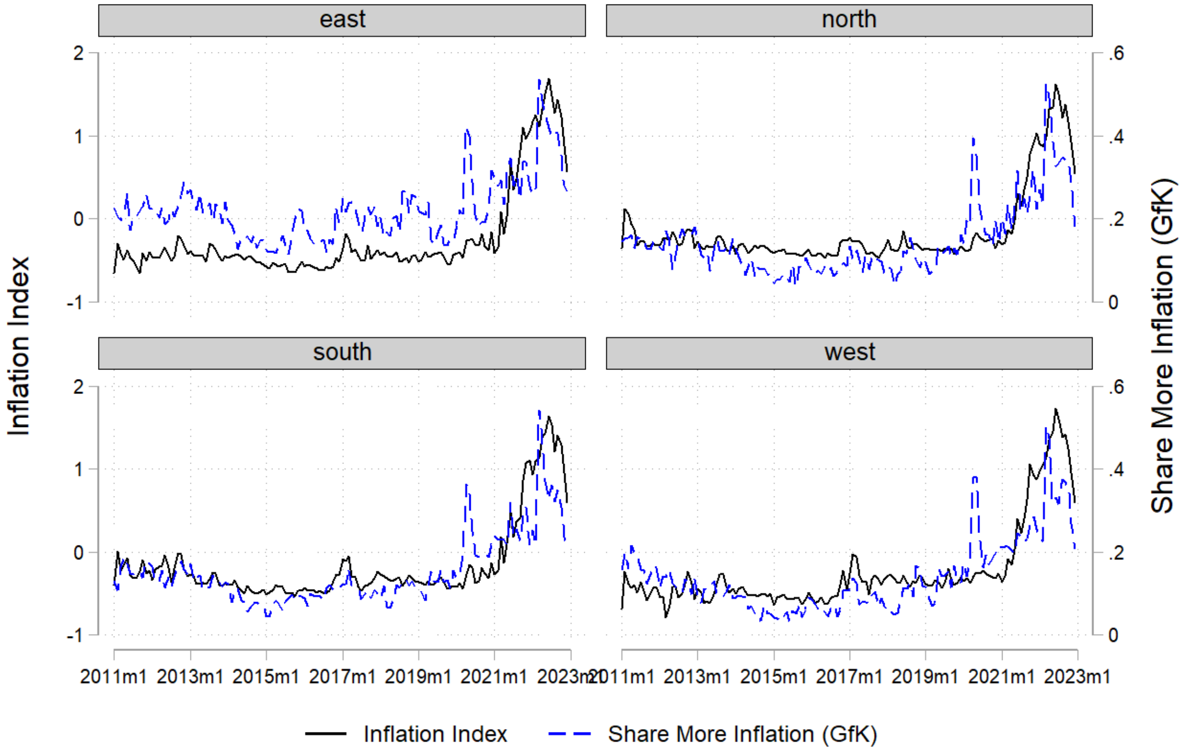


Figure 10
Inflation Index and GfK inflation expectations by regions

D Classification of different users

To classify the 230,000 user accounts that tweet about inflation into private individuals, media organisations and other possible user categories, we use a pre-trained machine learning model (specifically, the *twhin-bert-large* model) which we fine-tune for this specific task. We provide six different labels, *private individuals*, *individual journalist*, *influencer*, *media organization*, *business organisation*, and *other organisation*. To generate training data for this fine-tuning process, we again use the OpenAI API to label almost 45,000 users, which we use to fine-tune the *twhin-bert-large* model.[26] After inferencing on all the users in our sample, it became clear that the model performs differently well across the different classes, as manual checks but also the precision scores in Table 5

[26]We need this large amount of labelled user accounts to obtain enough training data for all the different categories as we need an equal amount of training data for each class.

show.[27] The table shows that the model performs best in the categories *private individuals* and *media organization*, which are also the most important ones.

We obtain 64,458 private user accounts and 4,664 belonging to media organisations. Since not all of these media accounts are focusing on economic context and are, therefore, likely to lead the discussions on Twitter about monetary policy decisions, we select a subset of these media accounts according to the following criteria: we only keep the accounts that post at least 50 inflation tweets and have at least 10,000 followers, leading to 86 user accounts, the most important of which are shown in Table 6.

Table 5
Precision scores of the different user categories

User class	Precision
private individuals	0.89
individual journalist	0.61
influencer	0.62
media organization	0.84
business organisation	0.67
other organisation	0.74

Table 6
The ten most important media organisations

Username	No. of inflation tweets	No. of followers
hb_finanzen	499	56,023
faznet	458	798,004
handelsblatt	414	377,159
FAZ_Finance	379	44,364
boersenzeitung	352	30,785
RND_de	351	62,692
zeitonline	292	2,573,925
tagesschau	280	3,928,233
derspiegel	270	3,111,783
SPIEGEL_alles	233	117,122

E Correlations for individual product categories

[27] We rely on the precision score here since it is the most suitable measure for our propose. It is especially useful in situations where the emphasis is on minimizing false positives, ensuring that when the model predicts a positive outcome, it is highly likely to be correct—*i.e.*, when the model predicts a user to be a household the probability that this user is indeed a household is very high.

	Animals & Pet Supplies	Apparel & Accessories	Arts & Entertainment	Baby & Toddler	Business & Industrial	Cameras & Optics	Electronics	Food, Beverages&Tobacco	Furniture	Hardware
net_count_ma6sd	0.01	-0.34***	0.26***	-0.18***	-0.15***	-0.18***	-0.47***	-0.19***	-0.18***	-0.38***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Constant	0.92***	2.63***	1.86***	1.95***	1.40***	0.89***	2.12***	2.65***	1.21***	2.75***
	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.05)	(0.06)	(0.06)	(0.06)
N	881	881	881	881	881	881	881	881	881	881
r2_a	-0.00	0.11	0.07	0.03	0.02	0.03	0.22	0.03	0.03	0.14

Table 7
Correlations of our index and the item quantity purchased in different product categories - standardized (6-day moving averages)- 1

	Health & Beauty	Home & Garden	Luggage & Bags	Mature	Media	Office Supplies	Religious & Ceremonial	Software	Sporting Goods	Toys & Games	Vehicles & Parts
net_count_ma6sd	-0.29***	-0.10***	-0.13***	0.06*	-0.12***	0.35***	0.01	0.06*	0.11***	-0.13***	-0.12***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.08)	(0.03)	(0.03)	(0.03)	(0.03)
Constant	2.84***	0.31***	1.98***	0.31***	2.88***	2.86***	0.68***	1.15***	1.00***	1.31***	2.04***
	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.14)	(0.06)	(0.06)	(0.06)	(0.06)
N	881	881	881	793	881	881	194	881	881	881	881
r2_a	0.08	0.01	0.02	0.00	0.01	0.12	-0.01	0.00	0.01	0.02	0.01

Table 8
Correlations of our index and the item quantity purchased in different product categories - standardized (6-day moving averages) - 2