Analyzing the Impact of Media Inflation Narratives

by Tammo Thobe*
Data Analysis Lab
Minor Digital Sciences

January 22, 2024

Abstract

This research delves into the intricate relationship between media narratives, inflation, and economic expectations. In the wake of a sudden surge in inflation, this study explores the impact of news coverage as well as sentiment in the context of inflation on inflation expectations and economic outcomes. Utilizing data from the New York Times, the project analyzes news articles related to inflation from January 2010 to December 2023.

*tammo.thobe@student.uibk.ac.at

Contents

1. Introduction

Following a prolonged period of relatively low inflation, the world has recently seen an unprecedented surge in inflation, bringing the issue to the forefront of macroeconomic debates. Although there is an abundance of research regarding the topic of inflation, many aspects remain unclear and warrant further investigation.

A cornerstone of macroeconomic theory depends on the finding that macroeconomic variables are influenced by expectations.[1] This finding was revolutionized by the rational expectations hypothesis, first formulated by Muth in 1961. In general it states that individuals expectations on average correspond to the predictions of economic theory. [2]

However, critique of the rational expectations theory has been that it remains unclear how individuals obtain their knowledge to form expectations and that the theory does not capture peculiar features of empirical macroeconomics. Carrol adresses this by analyzing an alternative model of expectations formation, in which agents obtain their macroeconomic knowledge from the news, which in turn obtain their views from professionals who are assumed to be somewhat rational.[1]

There is another possible model for expectations, which is relatively new to economics. "Narrative Economics" consider the possibility that economic stories and their pervasiveness can help explaining economic outcomes.[3]

This projects aims to analyze the impact of narratives in the media regarding inflation, how these narratives are formed and their impact on inflation expectations. Four different research questions are stated to guide this analysis. Firstly, when does the media report about inflation, is higher inflation correlated with higher news reporting and vice versa? Secondly, is greater news coverage associated with more rational household inflation expectations? Thirdly, the paper asks whether media narratives affect inflation and inflation expectations, whether they are correlated and if narratives can help predict inflation and expectations. Finally the question is stated, how does the macroeconomy, especially consumer spending, react to inflation news shocks?

The remainder of this paper is organized as follows: Section 2 describes the data as well as the process of data preprocessing and finally the methodology used for the data analysis. Section 3 lays out the results of the analysis, while section 4 contains the discussion in regards to the topic along with a discussion with regards to ethical and legal implications of the project. Finally section 5 concludes.

2. Methodology

The following chapter introduces the work flow in obtaining and preprocessing the data for the subsequent analysis. Furthermore the methodology for the analysis and its fit for answering the above defined research questions are described.

2.1. Data description

To analyze media narratives, a data foundation had to be layed out. As described above there exists an economic model in which agents derive their macroeconomic knowledge from newspapers, but are assumed to only absorb the content probabilistically.[1] The possibility remains that agents obtain their views from online sources, such as Twitter (X) or from public discourse, however due to availability as well as time constraints and based on the model by Carrol, this project focuses on news related narratives in the context of inflation.

To obtain data, news articles were scraped from the web, using the New York Times API. The original purpose was to create a data set containing a variety of newspaper sources to be able to have a large sample with a wide variety of narratives. However it turned out that obtaining news articles without costs and more importantly in a small amount of time was not possible. Therefore the project focuses on articles obtained from the New York Times.

The first step was to acquire an API key from the New York Times website. Then a crawler was build, that uses the API key, a specific time frame and the term "inflation" to gather URLs corresponding to the described input. After that a function was created, that used the URLs to extract the articles body, meaning the text itself. This process was repeated for articles between the first of January 2010 and the 12th of December 2023, resulting in a dataset of 15808 news articles containing the keyword "inflation".

2.2. Data preparation and processing

An overview of the project workflow is given in Figure 1. Once the data set was crawled from the New York Times webpage, the data had to be prepared and processed for the analysis. The focus of this project was on Inflation media narratives in the context of the United States. The first step in preprocessing then had to be to classify the articles based on the geographical focus of each article and then subsetting the data to only keep articles with a focus on the United States.

For this purpose the semi-supervised geographical news classifier "Newsmap" was used, which constructs a classifier from a corpus through a manually gathered dictionary. This classification method makes manually classifying thousands of news stories to train a model unnecessary and is therefore more efficient than comparable methods.[4] First a corpus was created, which was then tokenized and stopwords as well as punctuation were removed. From these tokens two document feature matrices were created, one general and one that records the class membership of the documents using the English seed geographic dictionary. These were than used to train the newsmap model and subsequently predict the most associated geographical locations. Furthermore, as an additional criterion articles were classified to the US, when they contained the term "Wall Street". This process resulted in a subset of the data of 8731 articles that geographically focused on the United States.

When crawling the news articles, retrieval was already conditioned on including the term inflation. Nevertheless, inflation can be incorporated in a variety of contexts and to ensure that it is used in an economic context the data had to be classified by topic.

To be able to classify the articles a Latent Dirichlet Allocation (LDA) was employed, which is a Bayesian model where each document is depicted as a mixture of finite topics, which in turn are modeled as an infinite mixture over an underlying set probabilities associated with the topics. Similar to the newsmap classifier above, LDA models do not require manual classification and are time and energy efficient.[5] As described above a corpus was created and the elements were then tokenized. Additionally to the process above, only the top 20 percent of the most frequent features that appear in less than 10 percent of all documents were kept in the document feature matrix, to focus on common but distinguishing features.

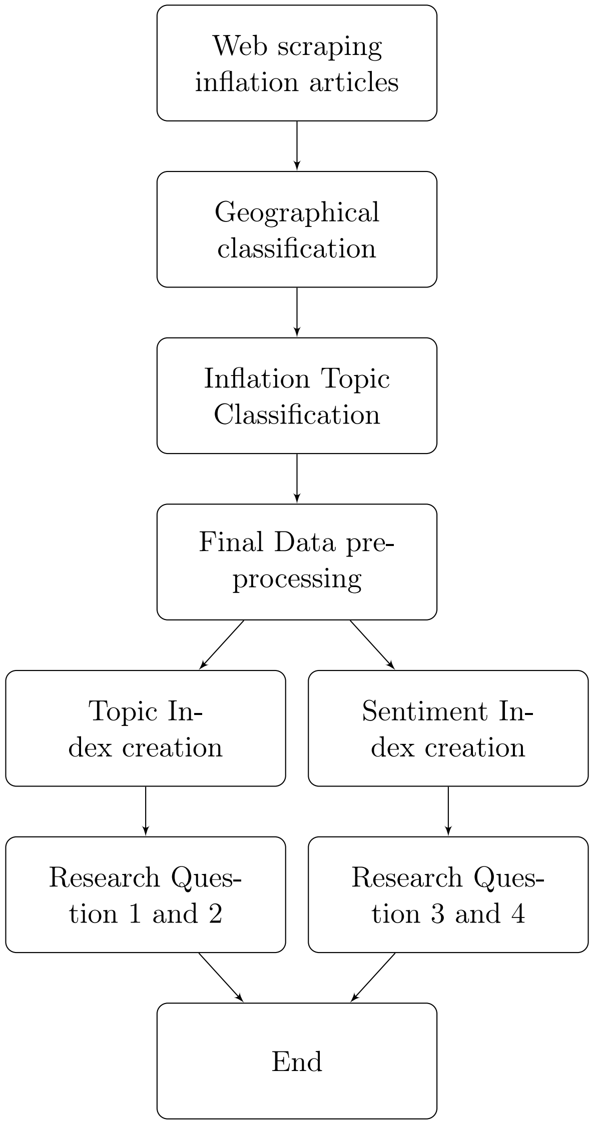

Figure 1: Overview Project Workflow

Then the LDA was applied to the document feature matrix. The parameter defining the number of topics to be retrieved was set to ten, which was established through trial and error. The resulting topics were then defined as "general politics", "home prices, consumer prices", "student finances, student loans and tuition's", "medical prices, drug prices and premiums", "politics related to inflation", "professional inflation, forecasts, monetary policy", "private finances and income", "other inflation related" and "energy".

Topic two to nine seemed to fit inflation in the context of economics, while topic one and ten contained articles that did not fit the analysis criteria. Subsetting the data set for topics two to nine and removing duplicates resulted in a final data set of 7179 articles.

2.3. Data analysis

The foundation of the project and the answers to the different research questions lie in a pair of indeces. The first indicates the degree of inflation reporting in a given month, called the "Inflation Topic Index" and the second depicts the sentiment in the news reagrding inflation, called the "Inflation Sentiment Index".
Creating the Topic Inflation Index, the articles were first counted on a monthly basis. The specific value of the Index was then created by running a simple min-max function over the specific monthly values.

$$Index_t = \frac{monthlycount_t - min(monthlycount)}{max(monthlycount) - min(monthlycount)} \qquad (1)$$

The values for the Topic Inflation therefore range between zero and one, where zero indicates the minimum value of monthly articles related to inflation and one the respective maximum.
The creation of the Inflation Sentiment Index follows the results of Shapiro et al. (2022). They create a lexicon for economic news articles based on a corpus of 238685 economic news articles. The lexicon assigns a sentiment class $c \in$ {negative, neutral, positive} to each sentence based on the Vader model which also includes the Loughran-McDonald (LM) and Hu-Liu (HL) lexica. Based on this they create a word-by-class matrix which counts the co-occurence of words and sentiment classes. Finally they calculate the "point wise mutual information" (PMI) between a word w and a sentiment class c defined as:

$$PMI(w,c) = log(\frac{p(w,c)}{p(w)p(c)}) \qquad (2)$$

where $p(w$ is a words share of total words, $p(c$ is the sentiment class share of total sentences and $p(w,c)$ is the probability that a specific word and sentiment class co-occur. The PMI enables to generate the degree to which a specific word and sentiment are related. The final sentiment score S for a specific word is then defined as:

$$S(w) = PMI(w, positive) - PMI(w, negative) \qquad (3)$$

and the score is normalized to a range between -1 and 1 for each word.[6]
The authors find that combining the LM and HL lexica and then adding scores for all other words from the "news PMI lexicon" described above (Where the sentiment score for each article is an average of the scores of the respective words), yields the best results in predicting articles classified by humans.[6]
The same approach is used in this project to derive sentiment scores for each article in our dataset. Machine Learning techniques would also have been possible to derive the sentiment scores, however these require large labeled training datasets which are time and ressource consuming. The derived scores are aggregated to a monthly basis to create the Inflation Sentiment Index, described above.
Both indeces are then used to answer the stated research questions. The resulting indeces are plotted in Figure 2 and Figure 4 respectively. Additionally as a lot of the analysis pertains to questions of correlation and causation, OLS regression methods are used, as they are equipped

to answer these questions. For the forecasting exercise ARIMA models were implemented and to analyze the impact of sentiment on macroeconomic aggregates the local projection framework established by Jordà (2005)[7] is implemented.
These methods will be described in more detail in the following chapter.

3. Results

This chapter shows the results and methods of the analysis and attempts to answer the defined research questions.

3.1. News coverage and inflation

The first research question stated: When does the media report about inflation, is higher inflation correlated with higher news reporting and vice versa? For the purpose of answering this question, the Inflation Topic Index was created, capturing the degree of inflation news coverage in a specific month. Figure 1 plots the Index next to the consumer price index for the United States, containing all items and capturing the growth in percent from the same period in the subsequent year. The figure already shows that inflation news coverage and actual inflation are highly

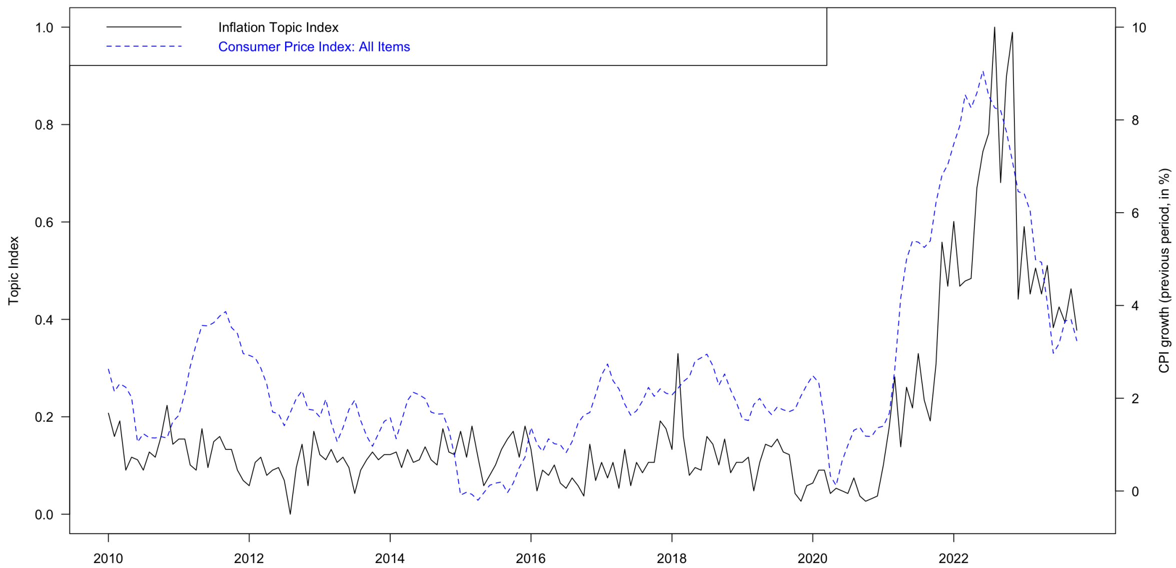

Figure 2: Inflation Topic Index and CPI growth

correlated. The pearson correlation coefficient for both measures equals 0.8289, which is a strong positive correlation, implying that if inflation increases, news reporting increases as well.
Furthermore we can estimate a simple OLS regression, containing the Topic Index as the dependent variable and CPI inflation as an explanatory variable. The results of which are shown in Table 1. The coefficient α_1 is positive and statistically significant for the full sample, which means that increasing CPI by one unit, increases the Topic Index by 0.0745. When inflation rises, the news increases the degree of reporting about inflation as well.

The equation was estimated for two additional subsamples, the first corresponding to observations where inflation was higher than the average inflation in the sample and the second where inflation was lower than on average. The coefficient for higher than average inflation is larger than for the full sample and also statistically significant, while the coefficient for the lower than average sample is very small and statistically insignificant. This indicates that, when inflation is very high, the news reports even stronger about the topic, while when inflation is low the impact on reporting is close to zero or insignificant.

Table 1: Inflation and Inflation News Reporting

Equation Estimated: $TopicIndex_t = \alpha_0 + \alpha_1 * CPI_t$			
Sample	α_0	α_1	R^2
all obs.	-0.0064 (0.8375)	0.0745 (6.679e-08)***	0.687
$CPI_t > mean(CPI)$	-0.1266 (0.2445)	0.0973 (7.57e-06)***	0.65
$CPI_t < mean(CPI)$	0.1026 (1.91e-07)***	0.0037 0.7273	0.0026

The answer to the research question is thus, higher inflation is correlated with higher news coverage. When inflation is above average, the increase in reporting is even stronger, while when inflation is below average, the impact on reporting is negligible.

3.2. News coverage and inflation expectations

The second research question asked whether greater news coverage was related to more rational household inflation expectations, as would be the case in the rational expectations theory.

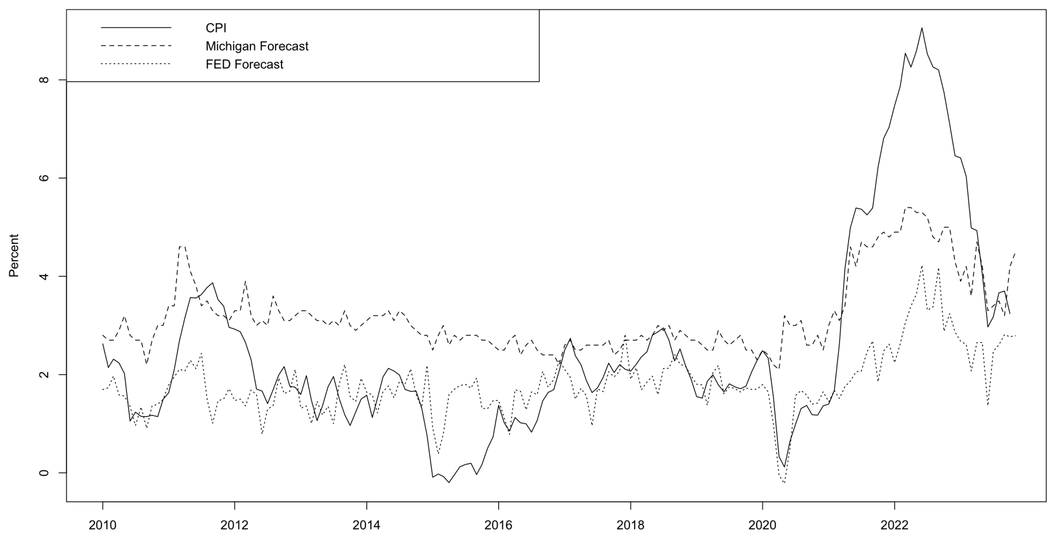

Figure 3: CPI Inflation with Household and Market Inflation Expectations

For this purpose household inflation expectations were derived from the Michigan Consumer Survey, which captures households inflation expectations for the next 12 months. Secondly the Federal Reserve Banks inflation forecast for the next year were derived, corresponding to professional forecasts or market expectations. Figure 3 depicts both measures combined with CPI growth in percent.
Similar to Carroll (2003), more rational household expectations are defined as household expectations closer to professional forecasts. This is captured, by creating a variable as the squared difference between the Michigan Consumer Survey and the FED forecasts $GAPMS_t = (MCS_t - FED_t)^2$. This variable is in turn used in a simple OLS regression as the dependent variable, with the Topic Index as the explanatory variable.[1]
The results are layed out in Table 2.

Table 2: Greater News Coverage and Rational Household Inflation Expectations

Equation Estimated: $GAPMS_t = \alpha_0 + \alpha_1 * TopicIndex_t$			
Sample	α_0	α_1	R^2
2010m1-2023m11	1.91 (2.792e-07)***	1.98 (0.1482)	0.0369
2010m1-2021m12	1.36 (0.01581)*	6.76 (0.09985).	0.0675
2010m1-2020m12	2.46 (0.00084)***	-4.58 (0.35299)	0.0153
2010m1-2019m12	2.13 (5.137e-05)***	-2.32 (0.5123)	0.00492
2010m1-2018m12	2.43 (2.061e-05)***	-3.77 (0.3082)	0.01245
2020m1-2021m12	2.49 (0.07456).	8.0017 (0.07234).	0.1662

The result for the full sample shows a positive coefficient for a_1 which does not correspond to the expected result. It would mean that households expectations get less "rational" when inflation reporting increases, as the gap between expectations and professional forecasts increases.
However, looking at different samples changes the picture. The samples below 2020m12 show the expected negative relationship between $GAPMS$ and $TopicIndex$, when the news increases its reporting about inflation, households expectations become more rational. For some reason the period between 2020 and 2022 is characterized by a strong positive relationship, which biases the whole sample result. This should be investigated in a further analysis.
In general the results hint at a negative relationship, which would mean that households inflation expectations get more rational, meaning they converge to professional forecasts, as the degree of inflation reporting increases. These results should be taken with a grain of salt and should be investigated further, with a larger sample size as for example none of the coefficients are statistically significant.

3.3. Inflation sentiment and expectations

The third research question stated: Do media narratives affect inflation and inflation expectations, are they correlated and if so can narratives help predict inflation and expectations? To answer this question, the Inflation Sentiment Index becomes necessary, which is plotted in Figure 4 below.

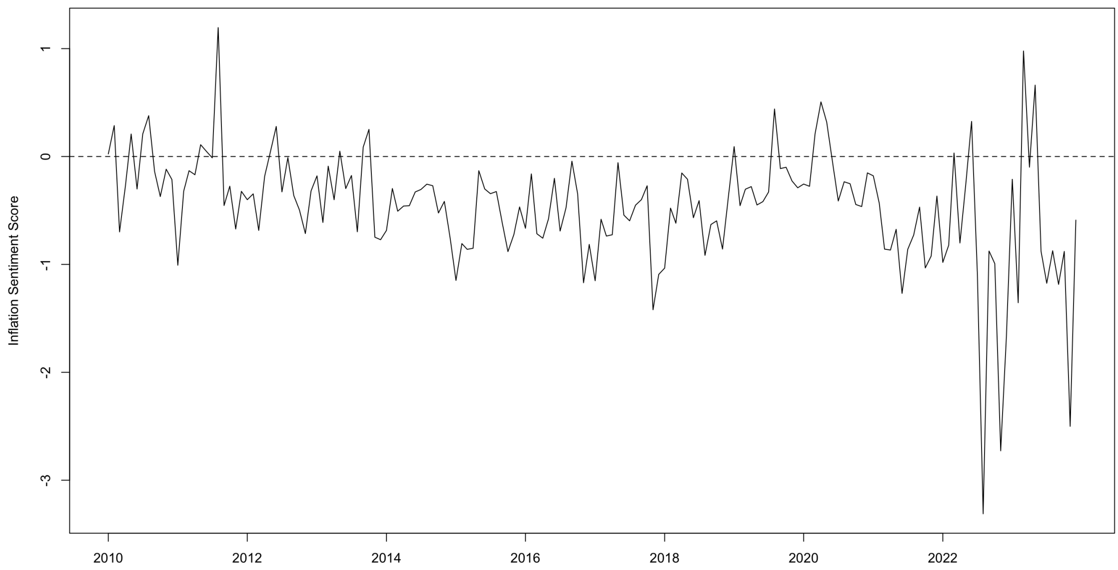

Figure 4: Inflation Sentiment Index

The figure shows that the sentiment remained relatively stable and slightly negative on average in the period of relatively low inflation. When inflation increased we can observe a strong decline in sentiment and also a lot of volatility in inflation news sentiment.
As a first approximation we can observe the correlation between the Sentiment Index and Inflation, as well as the two measures of inflation expectations, described beforehand. Inflation and the Index are negatively correlated (PCC: -0.2906, OLS: $\alpha_1 = -1.14$), the Michigan Consumer Survey (PCC: -0.2524, OLS: $\alpha_1 = -0.36$) as well as the FED Forecast (PCC: -0.3576, OLS: $\alpha_1 = -0.43$) and the Index are also negatively correlated.
This indicates that an increase in inflation sentiment, which is interpreted as an expected decrease in inflation in the news, is associated with a decrease in actual inflation as well as household and market inflation expectations.
What remains to be analyzed is whether the indeces can help predict inflation and inflation expectations and therefore, maybe already capture inflation expectations. To this end, three different ARIMA models with their corresponding p, q, d are identified and estimated and then in a further exercise used to forecast CPI inflation, the Michigan Consumer Survey and the FED forecast respectively for $h = 3$. Furthermore the same models are extended to include the Inflation Sentiment and Inflation Topic Index respectively and then also forecasted for $h = 3$. Table 3 depicts the results for this exercise with the corresponding mean squared prediction error. The first three rows correspond to the forecasting exercise with the simple models, while

rows four to six correspond to the models, including one lag of the Inflation Sentiment Index and Inflation Topic Index respectively. The MSPE for the Inflation Topic Index is depicted in brackets.

Table 3: Inflation Expectation and CPI Forecasting

Model	MSPE
ARIMA(1, 1, 0): CPI	0.1036
ARIMA(1, 1, 0): MCS	0.5412
ARIMA(1, 1, 1): FED	0.0602
incl. Inflation Index: CPI	0.307 (0.1175)
incl. Inflation Index: MCS	0.5351 (0.5447)
incl. Inflation Index: FED	0.0022 (0.0683)

The results show that the Inflation Sentiment Index cannot improve the forecast for CPI, however it increases the forecast accuracy for the Michigan Consumer Survey slightly and for the Federal Reserve Bank quite substantially. The Topic Inflation Index cannot improve the forecast of any measure, however it outperforms the Inflation Sentiment Index in predicting inflation itself.
These results can be interpreted as the Inflation Sentiment Index contains information about future inflation expectations, as it is possible to improve forecasts of measures of expectations. To what extent and the exact reasoning behind these findings need to be analyzed in more detail.

3.4. Inflation sentiment and macroeconomic aggregates

Finally we turn to the last research question: How does the macroeconomy, especially consumer spending, react to inflation sentiment news shocks? A common tool in empirical macroeconomics to estimate the response of aggregates to specific shocks is the local projection method established by Jordà (2005). We can generate so called impulse response funtions (IRFs) by running a sequence of forecasting regressions defined as:

$$x_{t+s} = \alpha^s + \sum_{i=1}^{p} B_i^{s+1} x_{t-i} + u_{t+s}^s \quad s = 0, 1, ..., h, \qquad (4)$$

where $x_t = [cc_t \quad y_t \quad \pi_t \quad i_t]'$ is a vector of model variables, corresponding to real personal consumption expenditures, industrial production (i.e. output), consumer prices (in logs) and the federal funds rate (i.e. interest rate) respectively. This equation is estimated s steps ahead for $h = 20$ horizons and up to $p = 4$ lags of the model variables in our case.
The IRFs are then constructed by:

$$\hat{IR}(t, s, d_i) = \hat{B}_1^s d_i \quad s = 0, 1, ..., h, \qquad (5)$$

where $B_1^0 = I$ and d_i is a $n * 1$ column vector, which first estimates a linear VAR and then

applies a Cholesky decomposition to the variance-covariance matrix, so basically identifies the structural shock from the reduced form shock, as is done in so called VAR models.[7]
A detailed description of how the shocks are identified goes beyond the scope of this project, however the resulting IRFs from the method described above are plotted below in Figure 5.

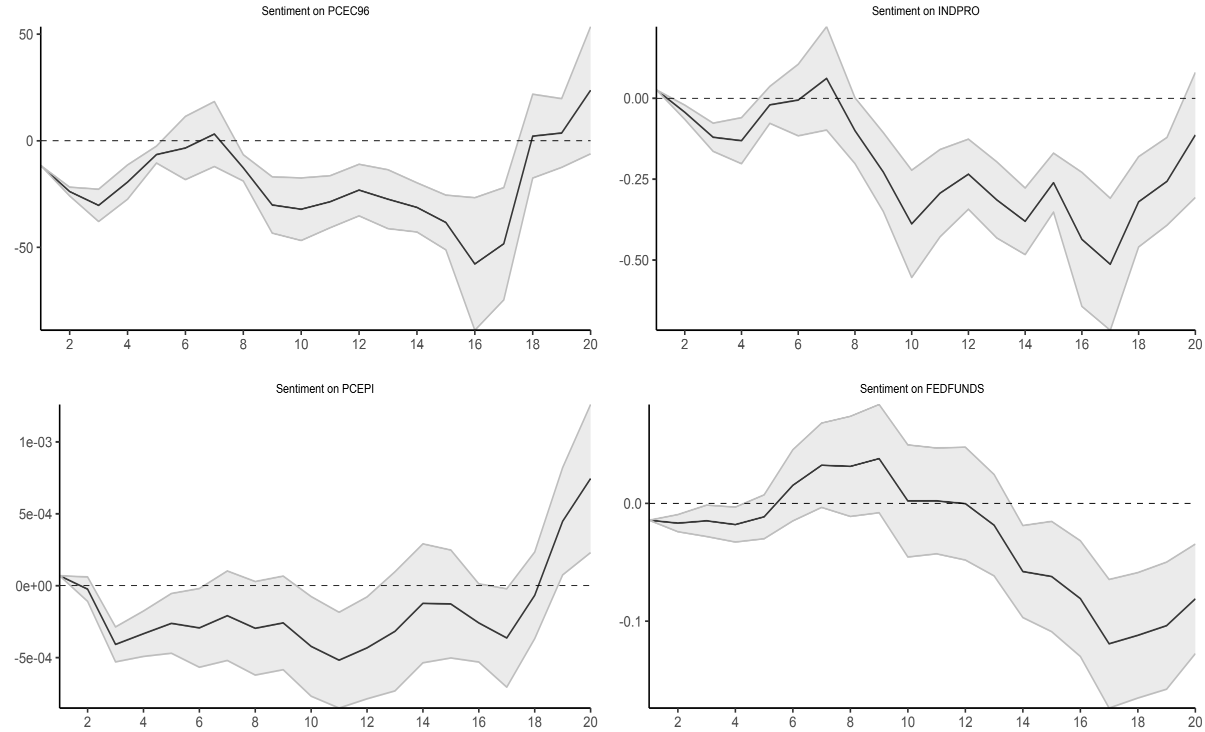

Figure 5: Impulse responses from one standard deviation inflation news sentiment shock on economic activity

A positive inflation news sentiment shock decreases inflation and inflation growth stays negative for about 18 months. Then inflation increases again. This seems sensible, as a positive shock to sentiment is interpreted as a decrease in inflation or expectations. The response of the federal funds rate to a sentiment shock is characterized by a slight increase up until month ten, after which the federal funds rate decreases. This is also roughly in line with the expected response (decreasing interest rate in response to lower inflation), as monetary policy is often quite sluggish in response to changes in inflation, although ten months is still quite slow.
The responses of consumption as well as industrial production are quite spurious. The expected response would have been, that a positive sentiment shock increases consumption as well as output. However consumption decreases, rising shortly up until month seven, then falling again and only increasing from month 16 onwards. Industrial production turns slightly negative, then positive around month seven, to then decline again and remaining negative for the remaining periods, with a slight increase towards the end.
These findings do not show a clear picture and different model specifications as well as variable transformations are necessary to investigate the responses of the macroeconomy to inflation news sentiment shocks further.

4. Discussion

The analysis finds that higher inflation is correlated with higher news coverage, with the effect being even more pronounced when inflation is above average and no significant effect when inflation is below average.
Furthermore the findings hint at more rational household inflation expectations when the news coverage increases. However, the results are not that clear, since there is a period between 2020 and 2022 which skews the results. It is not clear what exactly caused this period of more "irrational" expectations with an increase in news coverage. It could for example be induced by higher uncertainty which was apparent at the time or the incredible speed in which inflation increased during that period of time. It would be very interesting to investigate whether household expectations and their rationality depend on whether inflation is increasing or decreasing. Furthermore a larger sample should be considered.
Additionally the project finds that the relationship between inflation news sentiment and inflation expectations and inflation itself is negative, which would imply that an increase in sentiment would decrease inflation and expectations. This is in line with the expected relationship, but needs to be investigated further, including lags of the Inflation Sentiment Index. The Sentiment Index also aids in forecasting inflation expectations, especially professional forecasts, indicating that the Index might already capture inflation expectations. However, this result needs to be tested for robustness, considering different measures and methods as well.
Finally the response of macroeconomic aggregates to a news inflation sentiment shock are spurious. While the responses for inflation and interest rates seem sensible, the responses of consumption and output are unclear. Different model specifications and variable choices and transformations should be analyzed. It would additionally be interesting to use the Topic Index to derive the shocks and corresponding responses.
In general the dataset needs to be increased, incorporating a wider variety of publishers as well as a longer time horizon. The used dataset is sufficient as a first glimpse and to start investigating the relationships, however due to the smaller size there were some significance issues additionally a larger sample is more equipped to analyze the raised questions in more detail.

The impact of the project extends beyond the pure analysis described beforehand and also contains legal and ethical implications.
When first organizing the project it was essential to look at legal implications of crawling the web. Reading and following the API terms, usage policies and the contents of the provided robots.txt file are necessary in order to remain inside the legal framework of web scraping. Otherwise one could be in conflict with data protection laws such as the General Data Protection Regulation (GDPR), intellectual property rights, copyrights or trademarks.
One notion that is clearly identifiable from the project is that the impact of news extents beyond individuals thoughts and can have real world implications. It is essential for journalists to acknowledge this fact to then be able to adjust their actions correspondingly.
In times of increasing information at ever faster speeds and the impact of the internet, the issue of fake news is becoming more and more apparent. Wrong information can spread as a narrative through a whole economy, changing expectations and in turn possibly economic outcomes.

The power of narratives also in an economic context can not be understated and it is essential that society acknowledges the ethical implications this has for individuals as well as organizations who handle the distribution and communication of knowledge.

5. Conclusion

In conclusion, this study provides valuable insights into the dynamic connections between media narratives, inflation, expectations and economic aggregates. The findings underscore the impact of news coverage on shaping public perceptions, influencing expectations, and potentially impacting economic behaviors. Despite some complexities and nuances in the relationships identified, the project can be seen as a basis for further research on the subject. The ethical considerations highlighted emphasize the responsibility of media professionals in spreading accurate and contextually relevant information. The project also underpins the importance of narratives, especially in the context of economic analyses, especially as this is a relatively new theory in regards to the economic toolkit.

A. Appendix

GitLab Repository

The following chapter contains the names of the scripts in the gitlab repository (https://git.uibk.ac.at/csaz9602/news-inflation-index) corresponding to this project, with a reference to the specific parts of the report.

- Data gathering and preprocessing:
  - Web Scraping NYT per year.R (obtaining data from https://www.nytimes.com, M1/M2)
  - Consolidation NYT.R (consolidating yearly files into one data set, M1)
  - Geographical Classification of NYT articles.R (classifying articles corresponding to the US, M1/M3)
  - Topic Classification.R (additional classification to the inflation topic, M1/M3)
  - Final Data Preparation.R (final adjustments, M1/M2)
  - Sentiment Analysis Preparation.R (preparation for the sentiment analysis, M1/M2)
- Data analysis:
  - Topic Index.R (Creating the Inflation Topic Index, M1/M3)
  - NS_functions.py (Functions necessary for the sentiment score file, M1)
  - main_NS_score.py (Calculating the sentiment score for each article, M1/M3)
  - Inflation Sentiment Index.R (Creating the Inflation Sentiment Index, M1/M3)
- Results:
  - RQ1.R (Answering Research Question 1, M3)
  - RQ2.R (Answering Research Question 2, M3)
  - RQ3.R (Answering Research Question 3, M3)
  - RQ4.R (Answering Research Question 4, M3)

References

[1] C. D. Carroll. "Macroeconomic Expectations of Households and Professional Forecasters". In: *The Quarterly Journal of Economics* 118 (1 Feb. 2003), pp. 269–298. URL: https://www.jstor.org/stable/25053904.

[2] J. F. Muth. "Rational Expectations and the Theory of Price Movements". In: *Econometrica* 29 (3 July 1961), pp. 315–335. URL: https://www.jstor.org/stable/1909635.

[3] R. J. Shiller. "Narrative Economics". In: *American Economic Review* 107 (4 Apr. 2017), pp. 967–1004. URL: https://www.jstor.org/stable/44251584.

[4] K. Watanabe. "Newsmap: semi-supervised approach to geographical news classification". In: *Digital Journalism* (Mar. 2017). URL: http://eprints.lse.ac.uk/69525/.

[5] D. M. Blei, A. Y. Ng, and M. I. Jordan. "Latent Dirichlet Allocation". In: *Journal of Machine Learning Research* 3 (Jan. 2003), pp. 993–1022. URL: https://www.jmlr.org/papers/volume3/blei03a/blei03a.pdf.

[6] A. H. Shapiro, M. Sudhof, and D. J. Wilson. "Measuring news sentiment". In: *Journal of Econometrics* 228 (2022), pp. 221–243. URL: https://doi.org/10.1016/j.econom.2020.07.053.

[7] O. Jordà. "Estimation and Inference of Impulse Responses by Local Projections". In: *American Economic Review* 95 (1 Mar. 2005), pp. 161–182. URL: https://www.jstor.org/stable/4132675.