PLACEHOLDER Title

Jasper Bär*

Abstract

PLACEHOLDER Abstract

Keywords: PLACEHOLDER Keyword 1, Keyword 2, Keyword 3, Keyword 4, Keyword 5

*PLACEHOLDER
  Email address: ja.baer@stat-econ.uni-kiel.de (Jasper Bär)
  URL: PLACEHOLDER Github url

1. Introduction

Central bank communication has become an important tool of modern monetary policy (Blinder et al., 2017). Central bankers use communication is a important tool for guiding inflation expectations and ensuring trust. In recent years, most of the communication was aimed at financial experts. However, central banks are increasingly reaching out to the general public (Blinder et al., 2022). The media plays a crucial rule in this mechanism by making the information of the central bank communication available to the general public, which makes it crucial to understand how central bank communication is reported in the media for understanding the effects of central bank communication on the general public.

Most of the literature focuses on a direct link between central bank communication and financial variables. (To close to Picault et al. (2022).) Similar to Picault et al. (2022), I adapt a framework by Hayo & Neuenkirch (2015) where media acts a channel between the central bank communication and the perception of monetary policy by financial markets. Following this line of thought I assume that the media coverage of central bank communication acts as a channel between the central bank and the households. I assume that households consume the news about the central bank communication rather than the original central bank communication. This view is in line with Gardt et al. (2022) who show that according to the ECB Knowledge and Attitudes Survey, May 2021 households mostly hear about the ECB from television and newspapers (online and printed) and rarely by using direct sources like the ECB website. (To close to Blinder et al. (2022).)

Since Carroll (2003) created a model for inflation expectations that considers the amount of media coverage several studies have analyzed the role of media coverage for the inflation expectation forming process Ehrmann et al. (2017), (To close to Larsen et al. (2021), more citations?) Newer studies, extend the analysis to also incorporate the tone (Lamla & Lein, 2014) and the topics (Larsen et al., 2021) of media coverage.

Using the data-driven lexicon approach by Picault & Renault (2017), I find that news and ECB communication is strongly correlated with the current inflation in Germany with a few noticeable exceptions. Furthermore, the idiosyncratic error from a linear regression of media coverage on ECB communication is strongly correlated with the inflation expectation gap of German households.

I investigate the influence of media coverage in German newspapers of ECB press conferences on household's inflation expectation accuracy. Firstly, I investigate if the information given in regard to inflation in media coverage deviates from the inflation information in the original press conferences by the ECB which would introduce a bias and could potentially influence how the general public perceives monetary policy decisions. Secondly, I investigate how media coverage and central bank communication can be used to explain the household's inflation expectation errors.

This paper, is to my knowledge, the first paper which explicitly investigates the link between ECB communication and news coverage in regard to household inflation expectations. Furthermore, it contributes to the literature by demonstrating how the news about inflation deviates from the corresponding ECB communication, allowing to better understand how the ECB communication is conveyed to the general public and explains the channel through which the general public reacts to ECB communication.

(NOTE 1: I could maybe ask Tri and take his method instead. It is still supervised, but I could slightly modify it to produce a lexicon. It would most likely be a improvement over Picault et. al (2017) due to the word embeddings) (NOTE 2: Before I can describe the model which I use for the second part, I need to certain what exactly I want to regress and how to do it.)

2. Data

My dataset consists of two different parts. The textual data from the ECB and the Media and the quantitative data regarding inflation and inflation ex-

pectations. In section 2.1 I describe the textual data. In section 3 I describe how I transformed the textual data into quantitative data and in section 2.2 I describe the remaining quantitative data.

2.1. Textual Data

I collected two separate datasets for the textual analysis, one for the ECB communication and one for the media news. The ECB's main channel for communication to the media are their press conferences (CITE?). The press conference takes place after the ECB's Governing Council took their monetary policy decision every six weeks. I collected ECB press conference from 2000 until 2022 with a webscrapper from the official ECB website [1]. I discarded the Q&A section and all sections which carry no significance, like greetings and acknowledgments, leaving only the introductory statements (cite other who done it similarly).

(Note 1: News Before I can do the description of the news dataset, I need to know which data I am allowed to use. The next part describes how I clean the news dataset

Note 2: News dataset is in German, ECB dataset in English

To filter out any data that is not relevant for inflation, I only included sentences which contain the word "Inflation" (inflation) and synonyms of inflation like "Preisteigerung" (price increase). Furthermore, I filtered out all purely financial news, like reporting of stock movements or business news (better word?).

Note 3: Provide list of synonyms somewhere ?

Note 4: Not sure if this is the best way to select inflation related sentences. Maybe some kind of topic modeling would be better suited.

To reduce the dimensionality of the data I apply several pre-processing steps to the two datasets which are commonly used in the literature: Lowercasing, removing punctuation, removing stopwords (e.g., and, but, or) which carry no information and removing numbers.

NOTE 5: Currently the news dataset includes all inflation related sentences

---

[1]The press conferences can be found under https://www.ecb.europa.eu/press/pressconf/html/index.en.html

2.2. Quantitative Data

The ECB surveys each quarter professional forecasters regarding their annual HICP expectations and provides the results to the general public. The survey ranges from the first quarter of 1999 up until the last quarter of 2022. I use the one-year-ahead annual HCPI point forecast averaged over all responds as my professional inflation forecast.

The consumer inflation expectation is based on the results from the monthly European Commission's Business and Consumer Survey. The survey collects information about households inflation expectations by asking them if they expect that the prices increase more rapidly, increase at the same rate, remain unchanged or decrease in the next 12 months. To transform the qualitative survey answers into quantitative inflation expectations, I use an rolling-window regression based approach by Lahiri & Zhao (2015).

3. Text Classification

3.1. Method for Quantifying the Central Bank Communication and News Coverage

In the following section I will lay out how I quantified the ECB communication and news. I do this by separating my dataset into individual sentences and then classify each sentence according to predefined categories. Two main approaches are mostly used for text classification, the lexicon-based approach, and the machine learning approach. For the lexicon approach a text is classified by taking a list of words, the lexicon, where each word belongs to one of the desired categories, such as "good" for positive sentiment or "bad" for negative sentiment. The text is classified by counting the occurrence of each of these word

in the text. The lexicon approach can be modified by adding linguistic rules like negations (CITE?). The machine learning text classification approach classifies a text based on supervised or semi-supervised models to a given category. In recent years, deep-learning models have been widely used for text classification. (Continue this part with some examples) The ECB press conferences are classified into three categories where each category consists of three classes. The first two categories are taken from Picault et. al (2017). The first category describes the monetary stance expressed as monetary hawkish, monetary neutral or monetary dovish. The second category describes the economic outlook either as positive, neutral, or negative. I add a third class which describes the inflation outlook either as increasing, steady or decreasing. This third class allows for a direct comparison of the inflation expectations communicated in the ECB press conference to the inflation expectations expressed by the news media. The News articles are classified into two categories where each category consists of three classes. The first category describes the economic effect of the inflation expressed in the news which can be either positive, neutral, or negative. The second category describes the inflation direction which can be either increasing, neutral or decreasing.

3.2. Training Dataset

In the next section I will use supervised models to classify my data into several categories. I trained the models with two trainings dataset which I created, one for the news and one for the ECB press conference. I manually classified 3000 randomly drawn sentences from the ECB press conferences according to the three categories where each sentence was labeled with one of the three classes for each category. Similarly, 3000 sentence were randomly drawn from the news corpus and labeled based on the two categories. The sentences were then used to classify the remaining sentences from my dataset with the method described in the next two sections.

Is it obvious enough why I used two different training datasets? I should further describe my annotation scheme in the appendix and give some examples.

The simple implementation and transparency of Lexicon-based sentiment classification has made it widely used in the literature Shapiro (2020), (CITE), (CITE), (CITE), (CITE). Several lexicons exist for sentiment classification (CITE), (CITE), (CITE). However, these lexicons are not optimized for ECB communication or economic news. (FIND EXAMPLE). Picault et. al (2017) solve this problem by manually classifying each sentence in the ECB press conference based on their monetary stance and economic outlook. The classified sentences are then used to build a lexicon from the word used in the press conferences to classify the monetary stance and economic outlook of a press conference. I implement their approach with my own training dataset. The resulting lexicon is then used to classify each sentence in the ECB press conference for each of the three categories with the formula: (FORMULA IN LATEX, same as in Picault et. al (2017) or any other similar papers. It's just positive words – negative words divided by all words,) with the class score s, the press conference i, the category c, the class j and the number of word occurrences n. A final score for each category in each press conference is calculated by taking the averaging the category score over all sentences from the corresponding press conference. I slightly deviate from Picault et. al (2017) by first calculating the category score for each sentence instead of directly calculating score for the full press conference. My approach however is in line with most of the literature (Example from book or paper?). Similiar to the press conferences, I calculate the category score for the news. First I create a lexicon for each of the two categories with the method from Picault et. al (2017). The category score for each news sentence in my dataset is calculated with (EQ). A monthly category score is calculated by averaging the sentence score for each month. I further deviate from Picault et. al (2017) or similar papers like Marozzi (2021) (Still WP, and he at least uses valence shifters) by applying several linguistic rules to take grammatical(?) relations into account. (I added negation handling and I will add some more rules. Not sure if I should describe that in the main text or in the appendix. The rules are using POS tags to identify negative and positive phrases, e.g., a negative adjective in front of a positive noun results in a negative phrase instead

7

of a neutral phrase.

4. Descriptive Analysis

4.1. News Indexes

(FIG) depicts the INI together with the quarterly GERMAN CPI growth rate from 2000 to 2018. The index strongly correlates with the inflation (correlation value) with the noticeable exception of the inflation decrease during the Great Recession which was not fully depicted by the INI. (FIG) depicts quarterly GERMAN CPI growth from 200 to 2018 together with the number of sentences which I identified as inflation related. Before 2014 an increase in inflation related news is typically accompanied by a high inflation phase. However, from 2014 to 2017 the amount of inflation related news is relatively high during the low inflation phase. (Is there more to say? Should I search for reasons why the INI behaves strangely during the Great Recession? Any ideas?) (Note: Maybe I could also mention the number of inflation related news which is highest for high inflation periods especially in 2007 and 2008.)

4.2. ECB Indexes

(FIG) depicts the ECB inflation index in green next to household and professional inflation expectations. (Note: The German household inflation expectations are normalized) (Repeat same step for ECB index – Outlook vs Eurozone industrial production, Monetary – vs interest rate, Inflation – Eurozone inflation) (NOTE: The results are quite similar to Picault et. al (2017) and Marozzi (2021), which is reassuring, but I only add 4 more years to the dataset. Therefore, my approach adds little new information regarding the ECB indices.)

4.3. Difference between ECB and News Indices (Placeholder title)

(Is an extra subsection for the comparison of ECB and News Indices smart?)

(Rule based forecast for high differences in German household expt) The inflation expectation error of household is in the literature typically defined as

the difference between household inflation expectations and professional inflation expectations (CITE) (CITE). To measure the error, I use the EU household inflation expectation described in and the professional inflation forecasts for the Eurozone described in (SECTION 2). My interest lies in the deviation between the household inflation expectations from the optimal forecasts. Therefore, I follow Lamla (2014) and calculate the household forecast errors as the difference between the household inflation expectations and professional inflation forecast for the Eurozone in the coming 12 months. I deviate from Lamla (2014) by taking the non-absolute error. (CITE) and (CITE) show that households react differently to news regarding high and low inflation. (Lamla (2014) uses the absolute error. However, wouldn't it make more sense to take the non-absolute error in case of asymmetric behavior? Some papers indicate asymmetric behavior when it comes to inflation in news. Even if the behavior is symmetric, it could still be interesting to communicate that) (How to continue from here? I probaly should define difference metric between the ECB inflation index and the news inflation index. Lamla (2014) simply do some OLS regression (Expecation gap on media indices. However, he also created a theorie which ties his regression together and lets him explain his results. I could try to modify his theory or come up with something on my own. Would that make sense???)

5. Model

6. Estimation

7. Conclusions

References

Blinder, A., Ehrmann, M., De Haan, J., & Jansen, D.-J. (2017). Necessity as the mother of invention: Monetary policy after the crisis. *Economic Policy*, *32.92*, 707–755.

Blinder, A., Ehrmann, M., De Haan, J., & Jansen, D.-J. (2022). Central bank communication with the general public: Promise or false hope? *Journal of Economic Literature (forthcoming)*, .

Carroll, C. D. (2003). Macroeconomic expectations of households and professional forecasters. *The Quarterly Journal of Economics*, *118.1*, 269–298.

Ehrmann, M., Pfajfar, D., & Santoro, E. (2017). Consumers' attitudes and their inflation expectations. *International Journal of Central Banking*, *13.1*, 225–259.

Gardt, M., Siria, A., Mee, S., & Glöckler, G. (2022). Ecb communication with the wider public. *Economic Bulletin Articles*, *8*.

Hayo, B., & Neuenkirch, E. (2015). Self-monitoring or reliance on media reporting: How do financial market participants process central bank news? *Journal of Banking & Finance*, *59*, 27–37.

Lahiri, K., & Zhao, Y. (2015). Quantifying survey expectations: A critical review and generalization of the carlson–parkin method. *International Journal of Forecasting*, *31.1*, 51–62.

Lamla, M. J., & Lein, S. M. (2014). The role of media for consumers' inflation expectation formation. *Journal of Economic Behavior & Organization*, *106*, 62–77.

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

Picault, M., Pinter, J., & Renault, T. (2022). Media sentiment on monetary policy: Determinants and relevance for inflation expectations. *Journal of International Money and Finance*, *124*.

Picault, M., & Renault, T. (2017). Words are not all created equal: A new measure of ecb communication. *Journal of International Money and Finance*, *79*, 136–156.

10

Appendix A. Quantification of Inflation Expectations