Inflation Expectations, Sovereign Bond Yields and Media Sentiment on the ECB in Four European Countries

Wiem Ghazouani[a] and Matthieu Picault[a]

[a]Université d'Orléans, Laboratoire d'Economie d'Orléans (LEO)

February 22, 2024

Abstract

[Preliminary Version - Do not Quote]

With 20 countries and 24 languages, the Euro Area presents a unique communication challenge to the European Central Bank especially when countries' inflation diverge. Using a unique dataset of more than 200 000 press articles from 28 journals in 4 countries between 2020 and 2023, we study how the media coverage of the ECB monetary policy varies across the Euro Area by studying both the volume of articles published and the content of the articles. To analyze the sentiment conveyed by the different articles in a multi-language corpus, we rely on both lexicon approach and deep learning Natural Language Processing models to improve the comparability of our quantitative measures. Using the newly released data from the Consumer Expectation Survey, we discuss how the information content of press articles on monetary policy can influence both households inflation expectations and the dynamic and sovereign bonds yields. Using a panel dataset with country specific inflation expectations from both professional and households, macroeconomic control variables and fixed effects, we show that the tonality of the media can significantly influence households inflation expectations between 2020 and 2023.

Keywords: Central Bank, Inflation expectations, Media, Natural Language Processing.

JEL classification: XXX.

1 Introduction

Note to the QFFE organizers: the paper is under construction. At this stage, we have gathered a large database of 223 693 articles, we have processed the data and got preliminary indicators of sentiment and first empirical estimates. We are currently extending the paper as such: computing additional measures of sentiment, comparing the measures, focusing on the articles dealing with the press conference, and developing the effect of sovereign yields.

While the importance of central bank communication has already been widely demonstrated, its ability to reach directly the population relies on different medium. They can include direct communication from members of the institution on TV, radio, newspapers, or through social networks or internet platforms. They also include indirect communications where medias will present, discuss and/or interpret official communication such as Monetary Policy Statements or Press conferences. For most central banks, both the official communications and the medias are using the same language. But the Euro Area introduced, since its creation, another communication challenge for central banker : a population using 24 different languages. Therefore an official communication given in English, as the European Central Bank (ECB) press conferences, will be discussed by the medias in different languages with possible asymmetries across countries and languages. The importance of the media as the relay of central banks official communication has been discussed, among others, by Berger et al. (2011). They show that the reporting from the media varies depending on the economic context and the inflation level. At the same time, the importance and role of expectations and particularly inflation expectations of both firms and households in economic decisions became central for the conduct of monetary policies (Coibion et al. (2020)). Central bank can influence such expectations through its direct communication Lamla and Vinogradov (2019) but also through the media as media coverage is a good predictor of both inflation and inflation expectations in the US (Larsen et al. (2021)). Picault et al. (2022) also show that this reporting from the media is relevant to explain market inflation expectation measures.

As media exert a significant influence on inflation expectations, our empirical inquiry delves into the interplay among inflation expectations, media sentiment, and medium to long-term sovereign bond yields. While the magnitude of the relationship between inflation expectations and bond yields has been intensively discussed, Duffee (2018) corroborates the presence of a positive relation. Using a unique corpus of more than 200 000 press articles from 28 medias in 4 core European countries, our main objectives is to analyze the tonality of the medias when covering informations related to the ECB. Multiple approaches based on Natural Language Processing (NLP), Dictionaries, Machine Learning (ML) and Deep Learning (DL) exist to ensure English and non-English linguistic comparison, statistics, classification and modeling. Sentiment analysis is a computational process that categorizes given data dealing with specific topics into categories. The frameworks divided into three categories: (i) lexicon-based, (ii) corpus-based using machine learning and (iii) hybrid approaches, for both English, other languages and translated languages. Lexicon-based sentiment analysis considers a dictionary to ensure text classification for unannotated data. Tetlock (2007) and Tetlock et al. (2008) uses the Harvard University's General Inquirer (IV-4) developed by Stone et al. (1962) to measure the pessimism degree in media news journals. Their analysis show that pessimist media announcements carrying negative information about enterprises' profits lead to decline the pressure on market prices. Loughran and McDonald (2011) pointed that IV-4 dictionary considers many neural words in economic and finance as negative and constructed a context-specific dictionary dictionary (LM). Loughran and McDonald (2015) demonstrate that sentiments explaining post 10-k filing stock return uncertainty with LM dictionary provide better performance of classification compared to IV-4 dictionary. Garcia (2013) constructed financial news sentiments from the New York Time articles by using the LM dictionary to predict stock returns during the 20th century. Based on both IV-4 and LM dictionaries, Heston and Sinha (2017) also predicted compagnies stock returns from news articles. Sentiment analysis with corpus-based using supervised ML classifiers consists of considering a labelled text data, generally large textual corpus rated by individuals (e.g., tweets and movie reviewers). The annotated set is randomly separated into training set that is used to run the models in order to learn sentiments, and the testing and validation sets are used to evaluate models performances by highlighting how well a ML classifier captured the text polarity. Such method also focuses on feature engineering. It is the process of transforming textual corpus into numerical vectors with NLP feature transformation such as: Bag of Word (BOW), TF-IDF, Word embedding (Wod2Vec) used as input in ML models. It also focuses on model selection based on specific metrics (Accuracy, F1 score, Recall, Precision, etc.) to evaluate the predictive power of classification. For some years research has been directed towards the use of algorithms with transformers such as the Bidirectional Encoder Representations from Transformers (BERT, Devlin et al., 2018), Robustly Optimized BERT Pretraining Approach (RoBERTa, Liu et al., 2020) or Distilling BERT (DistilBERT, Sanh et al., 2019). Synchronously, transformers architectures of those pre-trained models have improved to handle multilingual tasks. Google research has introduced a multilingual version of BERT (M-BERT) . Rather than being trained on monolingual English data and English vocabulary, M-BERT is trained on raw Wikipedia texts expressed in 104 languages with multilingual vocabulary. Wu and Dredze (2019) pointed out that M-BERT seems to produce well cross-linguistic representations for multiple downstream tasks even when its training does not need any supervision or cross-language alignments or explicit objectives. Pires et al. (2019) also underlined how M-BERT is cross-lingually successful. Both papers focused on how the model functions, and revealed that the first layer-wise links cross linguistic performance to the quantity of shared words, and the second layer-wise considers the model ability to transfer between languages based on their word order similarity. The other most common transformers-based models for multilingual tasks are (i) Cross-Lingual Language model (XLM, Conneau and Lample, 2019) and which correspond to an extended version of BERT and aims to ensure cross-lingual classification and machine translation, (ii) XLM-RoBERTa (XLM-R, Conneau et al., 2020), a multilingual version of RoBERTa that is pre-trained with Masked language modeling (MLM) objectives on unlabeled text in 100 languages and Multilingual BART (mBART) introduced by Liu et al. (2020). Our contribution focuses on three strands of the literature. First, we provide a comparison of textual methods to extract sentiment from a multi-language textual dataset to guide futur research in economics, especially inside the Euro-Area. By comparing sentiment measures obtained from lexicon approach and LLM applied to either text in the national language or translated to english, we show that output are highly sensitive to the method selected. Also, evolution of automatic translation algorithm might reduce the reproducibility of research as models are updated without notice. We highlight the divergences of media reporting inside the Euro-Area, supporting the importance, for the ECB, of country-specific forms of communications. Second, we contribute to the existing literature on the relation between media sentiment and inflation expectations. We show that accross countries, media reporting of the ECB actions and policies has a significant influence on households' inflation expectations. An improving sentiment is associated with lower inflation expectations. This finding is robust to the inclusion of macroeconomic variables, monetary policy decisions and professional inflation forecasts. Last, we show that the media sentiment also affects euro-area countries sovereign bond yields with maturities of 5 to 10-year.

Our paper is structured as follow. The second section describes the textual methodology which mainly includes the data, textual techniques and empirical models. The third section discusses the relation between media sentiment and inflation expectations. The fourth section focuses on the effect of media sentiment on bond yields, and the fifth section concludes.

2 Database and measure of sentiment

To understand and analyze the media coverage of the European Central Bank, we manually collected press articles from European countries using Factiva, Europress or webscrapping. Given data availability, we obtained articles from four core European countries (France, Germany, Italy, Spain) covering 28 newspapers. To identify articles related to the euro area, we extracted articles with at least one mention of the European Central Bank or members of its governing council. Overall, the media database includes more than 200 000 articles between 2020 and 2023. Table 1 details the total number of articles from each source.

To measure the sentiment conveyed in the different languages, we classify our articles using both existing dictionaries and pre-trained deep learning NLP models . First, we rely on the automatic translation[1] of existing dictionaries of classified words (the LM dictionary) into the language of the article (Remus et al. (2010)). Then we measure the sentiment of a given article i with :

$$Sent_i^{LM\ trans} = \frac{\sum(Positive\ words_i) - \sum(Negative\ Words_i)}{Number\ of\ Words_i} \tag{1}$$

where $Positive\ words_i$ ($Negative\ Words_i$) is the occurrence of a word classified as positive (negative) in a dictionary. Finally, at a monthly frequency, we compute an aggregated media sentiment indicator equal to the average sentiment of all articles published during the period. The Figure 1 represents the measure of sentiment obtained using a translation of the LM dictionary after the announcement of the ECB Quantitative Easing Program in January 2015.

Following a methodology close to Lin et al. (2022), we also translated all the articles into English and applied the LM dictionary. This approach might provide a more reliable measure of sentiment as the translation is performed with the general context at a sentence level. Using Equation 1, we label this measure of sentiment $Sent_i^{LM}$.

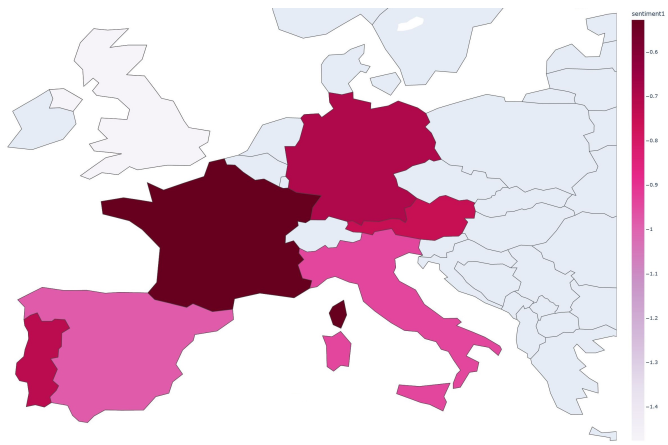
As an alternative measure of sentiment, we used the FinBERT pre-trained classifier from Huang et al. (2022) to obtain the sentiment polarity of a text translated into English. As FinBERT provides, at a sentence level, the probabilities of being labeled as positive, negative or neutral, we compute for each text i the sentiment score with:

$$Sent_i^{Finbert} = \sum_{i=1}^{s} p_s^{Positive} - \sum_{i=1}^{s} p_s^{Negative} \tag{2}$$

[1]All translation are preformed through the Google Translate API.

where s is a sentence of the text, $p_s^{Positive}$ is the Finbert probability of a sentence s being positive and $p_s^{Negative}$, its probability of being negative so that $0 \leq p_s^{Positive}, p_s^{Negative} \leq 1$. The overall sentiment of the text is then the sum of all sentences probabilities.

Figure 1: Monthly Sentiment $Sent_i^{LM\ trans}$ after the announcement of the ECB QE February 2015



Notes: Aggregated media sentiment from 6 European countries (Austria, France, Germany, Italy, Portugal, Spain) and international newspapers positioned (in the UK). The right scale is the aggregated measure of sentiment obtained from Equation 1.

Figure 2 illustrates the comparison of all sentiments over time for each country. 'sentiment1_lmtrans' is obtained from Equation 1 while 'sentiment2_lmtrans' uses as a divisor the number of words in the text. Using the translated version of the text and the LM dictionary in English, 'art_sentiment1' is the difference between positive and negative words divided by the number of positive and negative words and 'art_sentiment2' is the difference between positive and negative words divided by the total number of words. Finally, 'finbert_label_1' is obtained from Equation 2 and 'finbert_label_2' takes the difference between the length of positive and negative sentences.

[Under Construction]

3 The influence of media sentiment on inflation expectations

Households tend to collect their information regarding the economic situation from TV shows, newspapers, social medias, peer-interactions or their personal situation. Therefore, their view on the future path of their economy can be affected by both information they receive from medias and their personal situation. Regarding inflation expectations, we argue that households expected inflation $E(\pi_t)$ is conditional to both the observed economic situation and the medias view on monetary policy so that:

$$E(\pi_{t+1}) = f(Macro_t, E(Macro_{t+1}), Medias_t) \tag{3}$$

Where $Macro_t$ is the observed economic situation, $E(Macro_{t+1})$ is the agent forecast of the economic situation. This forecast can be influenced by the agent experience and by reporting of economic forecasts performed by public or private institutions. $Medias_t$ is the medias reporting toward the inflation dynamic and the central bank. More specifically, as medias tend to be more positive toward the central bank view on the economy and its action, the inflation expectations should be closer to the inflation target of the central bank. Then in a period of low (high) inflation, a positive reporting from the media might tend to increase (decrease) households inflation expectations. We test empirically this relation by estimating:

$$E(\pi_{t+n})_{i,t} = \alpha + \beta^1 \pi_{i,t} + \beta^2 E(\pi_{t+n})^{pro}_{i,t} + \beta^3 Medias_{i,t} + \epsilon_{i,t} \tag{4}$$

where $E(\pi_{t+n})_{i,t}$ is the households expected inflation in a country i for a given month t, $\pi_{i,t}$ is the observed inflation in the country i for a month t. $E(\pi_{t+n})^{pro}_{i,t}$ is the forecast by professional of future inflation. $Medias_{i,t}$ is the information content relayed by the medias regarding the central bank.

Focusing on the Euro Area, we rely on a balanced panel data set of 4 countries (France, Germany, Italy and Spain) over 43 months between April 2020 and October 2023 with both country and month fixed effects. We used the following proxies to estimate Equation 5:

- $E(\pi_{t+n})_{i,t}$: Households inflation expectations from the Eurostat Consumer Expectation Survey. It includes both the mean and median values of the quantitative forecast of inflation levels in 12 and 36 months. We label, for example, the mean survey point of the 12 months inflation forecast $E(\pi)_{i,t}^{mean,12m}$, and the median point of the 36 months forecast $E(\pi)_{i,t}^{median,36m}$. We also included the qualitative measure of inflation forecast from the survey (balance between respondents who anticipate an increase and a decrease of the future inflation level in 12 or 36 months) as $E(\pi)_{i,t}^{net,12m}$ and $E(\pi)_{i,t}^{net,36m}$.

- $\pi_{i,t}$: Annual inflation measured without the Harmonized Consumer Price Index (HICP).

- $E(\pi_{t+n})_{i,t}^{pro}$: Annual inflation forecasts in 12 or 24 months from the ECB Survey of Professional Forecasts (SPF). As this survey is conducted at a quarterly frequency, the values for a given quarter published at a month m is applied to the existing month and next two months $m+1$ and $m+2$.

- $Medias_{i,t}$: Information content related to the ECB obtained through press articles. We use the three measures detailed in the previous section $Sentiment_i^{LM\ trans}$, $Sentiment_i^{LM}$ or $Sentiment_i^{Finbert}$ to proxy for the information conveyed by the medias on the central bank.

We also include the impact of the media communication when combined with macroeconomic variables (GDP and unemployment rates) and the shadow rate as an indicator of the overall monetary policy stance.

The table (2) reports the baseline Panel OLS regression results between the tone of European media press and inflation expectations in next 12 and 36 months, without controls.

Knowing that we calculated ECB media press sentiments using FinBERT algorithm and English LM dictionary on translated journals into English, as the translated LM dictionary on original journals expressed in native languages. It reveals that across all univariate regressions, sentiment index of FinBERT algorithm shows a negative and statistically significant coefficients, which indicates the increase of inflation rates in the future. We deduce that European media tonality remains a relevant feature about the inflation path. The negative impact highlights that consumers anticipate a rise in inflation rate in next months. All specifications in table (3) underlines also that households inflation expectations are being negatively influenced by ECB media communications. Considering previous historical rates can emphasize the pessimism degree about the inflation rise. When combining inflationary indicators, macroeconomic data and ECB decisions, households continue being pessimist about inflation over the next 12 and 24 months. The unemployment in its turn significantly and negatively affect their view of the future of the economy. As it is low, consumers predict a decrease in the production hence the negative impact of GDP. Furthermore, if the shadow rate rises, it may signal anticipated changes in ECB monetary policy by for example lowering the quantitative easing, as a consequence, households may interpret it as the ECB is expecting lofty inflation, guiding them to alter their inflation expectations.

4 Monetary policy, economic situation, medias sentiment and sovereign yields

[Under Construction]

In this work, we also considered additional aspect for analysis in relation to sovereign bonds issued by the four key governments. Specifically, we consider Italian, German, French and Spanish bond yields at 5 and 10 years, designed by $BondYield, 5Y_{i,t}$ and $BondYield, 10Y_{i,t}$. We investigated whether such features are being influenced by the media sentiments, macroeconomic trends and central bank decisions with:

$$Y_{i,t} = \alpha + \beta^1 Medias_{i,t} + \beta^2 Shadow\ rate_t + \beta^3 Controls_{i,t} + \epsilon_{i,t} \tag{5}$$

Where $Controls_{i,t}$ is a set of controls including macroeconomic variables such as the GDP and unemployment rate, Results in table (4) show that the price assets positively and significantly react to the shadow rate, this indicates that when the index rise (tightening monetary policy / less accommodate attitude), the yields rise as well. While the negative and significant coefficients between media news and sovereign bond yields stipulates that the media content leads to an increase of sovereign bonds demand which decreases their yields. This reaction is shaped by how investors perceive economic conditions, monetary policy, risk and inflation in response to the news.

5 Conclusion

Media news are rapidly spreading news about central banks decisions and economic conditions can go viral and potentially alleviate the public concerns. In this paper we shed light on how the European media frames ECB decisions and impacts the public perceptions and awareness. The news seem to portray inflation in a negative light, it leads to rise consumers and investors anxiety and pessimism. We considered different European journals expressed in non-English spanning from 2020 until 2023, a period knows to witness a significant economic, pandemic and geopolitical pressures. The COVID-19 pandemic in early 2020 caused an unprecedented global economic disruption. Also, the conflict between Ukraine and Russia since 2022 added economic challenges. Russia is known to be a major and global oil, gas and various raw materials supplier, and such conflict led to an increase in energy and raw material prices which enhanced inflation around the world. Although we constructed a massive multilingual media sentiments set from April 2004, the forecast schedule of the survey only indicates latest available data from April 2020, hence we set a balanced panel regression analysis with fixed effects until Octobre 2023, which correspond to 172 observations (43 observations per country), in which our media-based measures are used as explanatory features to investigate whether the European media news has influence on the households' expectations across the different European countries.

We also considered a financial aspect, and tested whether the media sentiments which we built can drive the bond yield-based rates expressed in 5 and 10 years ahead.

Our study underlines the negative media perception, leading the consumers and the market participants to anticipate a forthcoming increase in inflation for the upcoming months for the different European countries.

6 Appendix

Table 1: Number of Articles per newspapers

country	Name	Nbr of Articles
Spain	ABC	5610
	El Correo	5584
	El Diaro Vasco	4103
	El Mundo	10340
	El Pais	16197
	La Vanguardia	7749
	ABC	6061
	La Voz de Galicia	3694
Germany	Bild	992
	Die Welt	4393
	Frankfurter Rundschau	2629
	Handelsblatt	11061
	Suddente Zeitung	5793
	Taz	1073
	Spiegel	10679
Italy	Corriere della Sera	8444
	Il Messagero	3398
	Il Resto Del Carlino	1350
	Il Sole	14597
	La stampa	4446
	Nazione	1511
	Republicca	3846
France	La Tribune	15484
	Le Parisien	1647
	L'Agefi	20617
	Le Monde	12607
	Les Echos	23088
	Liberation	3163
	Le Figaro	19147

Figure 2: Media News Sentiments respectively of Italy, Germany, France and Spain

Table 2: Baseline: Estimated Effect of FinBERT Algorithm Tones on Inflation Expectations

	$E(\pi)_{i,t}^{mean,12m}$	$E(\pi)_{i,t}^{median,12m}$	$E(\pi)_{i,t}^{mean,36m}$	$E(\pi)_{i,t}^{median,36m}$	$E(\pi)_{i,t}^{net,12m}$	$E(\pi)_{i,t}^{net,36m}$
$Sent_i^{FinBert}$	-0.5651***	-0.5690***	-0.5418***	-0.5071***	-0.0879	0.4080
	(0.0475)	(0.0553)	(0.0486)	(0.0602)	(0.0950)	(0.0808)
Const	-1.882e-16***	-7.4e-17***	2.742e-16***	5.27e-17*	-9.846e-16***	1.209e-17***
	(1.146e-17)	(1.729e-17)	(1.57e-17)	(3.136e-17)	(2.466e-17)	(1.396e-17)
Country FE	Yes	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	172	172	172	172	172	172
Entity	4	4	4	4	4	4
Time periods	43	43	43	43	43	43
R^2	0.3194	0.3238	0.2936	0.2572	0.0077	0.1665
F Statistic	78.358***	79.956***	69.399***	57.822***	1.3017***	33.350***

Note: $^*p < 0.1$ $^{**}p < 0.05$ $^{***}p < 0.01$

Table 3: Specification: Estimated Effect of FinBERT Algorithm Tones and Control Data on Inflation Expectations

	$E(\pi)_{i,t}^{mean,12m}$	$E(\pi)_{i,t}^{median,12m}$	$E(\pi)_{i,t}^{mean,36m}$	$E(\pi)_{i,t}^{median,36m}$	$E(\pi)_{i,t}^{net,12m}$	$E(\pi)_{i,t}^{net,36m}$
$Sent_i^{FinBert}$	-0.2594***	-0.2372***	-0.3353***	-0.2096**	-0.0275	0.0956***
	(0.0391)	(0.0431)	(0.0453)	(0.0859)	(0.0742)	(0.0347)
$E(\pi_{12m})_{i,t}^{pro}$	-0.3805***	-0.3425***	-0.1323***	-0.0005	-0.5658***	-0.0269
	(0.1195)	(0.1153)	(0.0482)	(0.1282)	(0.1488)	(0.0906)
$E(\pi_{24m})_{i,t}^{pro}$	-0.1551***	-0.2036***	-0.0898	-0.3047***	-0.2980***	-0.0714
	(0.0191)	(0.0283)	(0.0644)	(0.0524)	(0.0470)	(0.0609)
GDP	-0.0058	-0.0005	-0.0044	-0.0053	0.0016	-0.0076
	(0.0043)	(0.0051)	(0.0085)	(0.0062)	(0.0055)	(0.0046)
Unemployment	-0.2764***	-0.2844***	-0.3747***	-0.3196**	-0.2293***	-0.0101
	(0.0569)	(0.0788)	(0.1114)	(0.1253)	(0.0783)	(0.0322)
Shadow Rate	0.1398***	0.1375***	0.0343	0.0218	0.0444***	-0.1682***
	(0.0199)	(0.0258)	(0.0377)	(0.0487)	(0.0155)	(0.0213)
Const	2.7463***	2.8042***	3.2633***	2.7624***	2.0635***	-0.4067
	(0.4423)	(0.6072)	(0.8480)	(0.9163)	(0.6219)	(0.2484)
Country FE	Yes	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	172	172	172	172	172	172
Entity	4	4	4	4	4	4
Time periods	43	43	43	43	43	43
R^2	0.670	0.733	0.452	0.533	0.331	0.562
F Statistic	54.918***	74.192***	22.245***	30.827***	13.362***	34.608***

Note: $^*p < 0.1$ $^{**}p < 0.05$ $^{***}p < 0.01$

Table 4: Baseline: Estimated Effect of FinBERT Algorithm Tones with and without Control Data on Sovereign Bond Yields

	$BondYield, 5Y_{i,t}$	$BondYield, 5Y_{i,t}$	$BondYield, 10Y_{i,t}$	$BondYield, 10Y_{i,t}$
$Sent_i^{FinBert}$	-0.4925***	-0.0457***	-0.5050***	-0.0600***
	(0.0899)	(0.0110)	(0.0875)	(0.0144)
$E(\pi_{12m})_{i,t}^{pro}$		-0.1274***		-0.2092***
		(0.0141)		(0.0158)
$E(\pi_{24m})_{i,t}^{pro}$		0.1104***		0.0995***
		(0.0147)		(0.0175)
GDP		-0.0022		-0.0025
		(0.0013)		(0.0020)
Unemployement		-0.0404		-0.0557
		(0.0390)		(0.0451)
Shadow Rate		0.2566***		0.2671***
		(0.0044)		(0.0058)
Const	-3.999e-17**	1.0990***	1.969e-18	1.2595***
	(1.738e-17)	(0.3186)	(2.142e-17)	(0.3658)
Country FE	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes
Observations	172	172	172	172
Entity	4	4	4	4
Time periods	43	43	43	43
R^2	0.2426	0.967	0.2550	0.966
F Statistic	53.486***	781.477***	57.161***	764.244***

References

Berger, H., Ehrmann, M., and Fratzscher, M. (2011). Monetary policy in the media. Journal of Money, Credit and Banking, 43(4):689–709.

Coibion, O., Gorodnichenko, Y., Kumar, S., and Pedemonte, M. (2020). Inflation expectations as a policy tool? Journal of International Economics, 124:103297.

Conneau, A., Khandelwal, K., Goyal, N., Chaudhary, V., Wenzek, G., Guzmán, F., Grave, É., Ott, M., Zettlemoyer, L., and Stoyanov, V. (2020). Unsupervised cross-lingual representation learning at scale. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pages 8440–8451.

Conneau, A. and Lample, G. (2019). Cross-lingual language model pretraining. Advances in neural information processing systems, 32.

Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.

Duffee, G. R. (2018). Expected inflation and other determinants of treasury yields. The Journal of Finance, 73(5):2139–2180.

Garcia, D. (2013). Sentiment during recessions. The Journal of Finance, 68(3):1267–1300.

Heston, S. L. and Sinha, N. R. (2017). News vs. sentiment: Predicting stock returns from news stories. Financial Analysts Journal, 73(3):67–83.

Huang, A. H., Wang, H., and Yang, Y. (2022). Finbert: A large language model for extracting information from financial text. Contemporary Accounting Research.

Lamla, M. J. and Vinogradov, D. V. (2019). Central bank announcements: Big news for little people? Journal of Monetary Economics, 108:21–38.

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

Liu, Y., Gu, J., Goyal, N., Li, X., Edunov, S., Ghazvininejad, M., Lewis, M., and Zettlemoyer, L. (2020). Multilingual denoising pre-training for neural machine translation. Transactions of the Association for Computational Linguistics, 8:726–742.

Loughran, T. and McDonald, B. (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. The Journal of Finance, 66(1):35–65.

Loughran, T. and McDonald, B. (2015). The use of word lists in textual analysis. Journal of Behavioral Finance, 16(1):1–11.

Picault, M., Pinter, J., and Renault, T. (2022). Media sentiment on monetary policy: determinants and relevance for inflation expectations. Journal of International Money and Finance, 124:102626.

Sanh, V., Debut, L., Chaumond, J., and Wolf, T. (2019). Distilbert, a distilled version of bert: smaller, faster, cheaper and lighter. arXiv preprint arXiv:1910.01108.

Stone, P. J., Bales, R. F., Namenwirth, J. Z., and Ogilvie, D. M. (1962). The general inquirer: A computer system for content analysis and retrieval based on the sentence as a unit of information. Behavioral Science, 7(4):484.

Tetlock, P. C. (2007). Giving content to investor sentiment: The role of media in the stock market. The Journal of Finance, 62(3):1139–1168.

Tetlock, P. C., Saar-Tsechansky, M., and Macskassy, S. (2008). More than words: Quantifying language to measure firms' fundamentals. The Journal of Finance, 63(3):1437–1467.