

A Service of

ZBW

Leibniz-Informationszentrum Wirtschaft Leibniz Information Centre for Economics

Heikkinen, Joni; Heimonen, Kari

# Working Paper Media tone: The role of news and social media on heterogeneous inflation expectations

Bank of Finland Research Discussion Papers, No. 8/2024

**Provided in Cooperation with:** Bank of Finland, Helsinki

*Suggested Citation:* Heikkinen, Joni; Heimonen, Kari (2024) : Media tone: The role of news and social media on heterogeneous inflation expectations, Bank of Finland Research Discussion Papers, No. 8/2024, Bank of Finland, Helsinki, https://nbn-resolving.de/urn:nbn:fi-fe2024090569406

This Version is available at: https://hdl.handle.net/10419/302555

#### Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

#### Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



# WWW.ECONSTOR.EU

Bank of Finland Research Discussion Papers 8 • 2024

Joni Heikkinen – Kari Heimonen

Media tone: The role of news and social media on heterogeneous inflation expectations



Bank of Finland Research Bank of Finland Research Discussion Papers Editor-in-Chief Esa Jokivuolle

Bank of Finland Research Discussion Papers 8/2024 5 September 2024

Joni Heikkinen – Kari Heimonen: Media tone: The role of news and social media on heterogeneous inflation expectations

ISSN 1456-6184, online

Bank of Finland Research Unit

PO Box 160 FIN-00101 Helsinki

Phone: +358 9 1831

Email: research@bof.fi Website: www.suomenpankki.fi/en/research/research-unit/

The opinions expressed in this paper are those of the authors and do not necessarily reflect the views of the Bank of Finland.

# Media Tone: The Role of News and Social Media on Heterogeneous Inflation Expectations

Joni Heikkinen\*# Kari Heimonen\*

This version: 04/09/2024

# Abstract

This study investigates the role of media tone on inflation expectations. Examining the relationships between news and the inflation expectations of various U.S demographic groupings, we find that traditional news influences older cohorts, while social media news align more closely with the expectations of younger and more educated groups. Interestingly, social media correspond more closely than traditional news with the expectations of professional forecasters. Our analysis shows that media influences can persist for longer than a year, highlighting the importance of historical inflation data and the gradual adaptation of new information. Additionally, we find that separate media tones for specific news topics such as "Inflation & Fed" and "Healthcare Costs" resonate differently across demographic groups. These insights highlight the nuanced role of media in shaping inflation expectations across demographic segments.

Keywords: Inflation Expectations, Household Heterogeneity, Media Tone, Local Projection, Language Model, Forecasting

**JEL codes:** E3, E31, E32, E52, E58

\*) e-mail: kari.heimonen@jyu.fi

e-mail: joni.heikkinen@bof.fi

Jyväskylä International Macro and Finance Research Group (JyIMaF),

University of Jyväskylä School of Business and Economics (JSBE), PO Box 35, FI-40014 University of Jyväskylä, Finland #) Research Unit, Bank of Finland, Finland

**Acknowledgements:** The JyIMaF research group is thankful for the OP Group Research Foundation's financial support. Mr. Heikkinen is also grateful to the Finnish Cultural Foundation, the Jenny & Antti Wihuri Foundation, the YJS Foundation, and JSBE for their financial support.

\* The authors are grateful for the feedback received from participants at various seminars and conferences, including at the 2024 ICMAIF28 conference at the University of Crete, where Antonio Conti provided valuable comments, as well as at the Bank of Finland and JyIMaF research group seminars, and the 39th Summer Seminar of Finnish Economists. We also appreciate the valuable comments from Maritta Paloviita and Esa Jokivuolle. The views expressed in this paper are those of the authors and do not necessarily reflect those of the Bank of Finland, the European Central Bank, or the Eurosystem.

# 1. Introduction

"The conventional story, which is dominant in central banks around the world, rests on two key premises. First, that inflation expectations... are an important determinant of realized inflation.... The second key premise is that central bank behavior and possibly central bank communications can influence inflation expectations and through them macroeconomic outcomes."

Ben Bernanke's keynote address to the NBER Conference, May 19, 2022

"Inflation feeds in part on itself, so part of the job of returning to a more stable and more productive economy must be to break the grip of inflationary expectations."

Paul Volcker's statement before the U.S. Congressional Joint Economic Committee, October 17, 1979

Between 2021 and 2024, many countries experienced inflation at levels well above central bank targets, raising concerns about the effectiveness of their monetary policies. A key aspect of monetary policy is managing household inflation expectations, which can vary significantly across demographic groups. Such variation in expectations could potentially have a significant impact on the overall effectiveness of policy measures.

Despite the well-documented heterogeneity in household inflation expectations, the influence of traditional news and social media on these expectations remains largely unexplored. Existing studies highlight the difference between consumers and professionals, as well as the impact of demographic factors such as gender, age, education, and cognitive abilities on inflation expectations (D'Acunto et al., 2019, 2021, 2023; Das et al., 2020; Malmendier & Nagel, 2016; Mankiw et al., 2003; Weber et al., 2022). Only a handful of studies specifically examine how various media sources and news topics contribute to this heterogeneity (e.g., Lamla & Maag, 2012; Lamla & Sein, 2014; Larsen et al., 2021).

Our study addresses this gap by investigating the role of media on inflation expectations across various U.S. demographic groups. We differentiate between the influences of traditional news and social media news, noting that social media also serves as a source of news, while considering factors such as education level and age. By employing the Carroll (2003) model, we assess the extent to which "media tone" of media shapes inflation expectations among different households. To achieve this, we use media tone data from Refinitiv MarketPsych Indices.<sup>1</sup> To our best knowledge, this study is the first to apply such data in this context.

For the general public, we utilize the University of Michigan's Inflation Expectations dataset based on its Survey of Consumers. As a proxy for expert expectations, we use the Survey of Professional Forecasters (SPF). Our

<sup>&</sup>lt;sup>1</sup> Refinitiv MarketPsych Indices quantify sentiment and emotional tone from news and social media using natural language processing techniques. It was previously known as the Thomson Reuters MarketPsych Indices (TRMI). See Appendix A for information about the construction of TRMI media tones.

observation period runs from Q2 1998 to Q2 2017, providing a rich dataset for examining the dynamic interaction of media tone and inflation expectations.

Using a two-step methodological approach, we first examine the influence of news and social media on inflation expectations separately using aggregated media tone data. We employ various models (including OLS, local projection and MIDAS models), as well as Granger-causality tests, to assess this aggregate influence.

We then explore the specific news topics that influence inflation expectations. For this, we utilize micro-level news data combined with the Latent Dirichlet Allocation (LDA) topic model and the Financial BERT (FinBERT) language model. These models let us categorize news into distinct topics and form our own micro-based media tones to analyze the influences of various inflation news on different demographic groups. We apply Dynamic Model Averaging (DMA) to further understand how particular news topics affect inflation expectations over time.

Our findings from the first part reveal a degree of variation in the influence of media on different demographic groups. Traditional news sources have a more pronounced influence on older cohorts, while social media news significantly influence the expectations of younger and more educated groups. Additionally, our study reveals the potential of media tones to explain and forecast actual U.S. inflation, highlighting the predictive power of traditional news relative to social media.

We also examine the relationship between media tones and SPF inflation expectations. Our findings indicate that SPF expectations align more closely with social media, particularly reflecting the expectations of younger, highly educated households.

We apply Granger causality tests to evaluate the predictive power of past media tones and SPF values on current inflation expectations. Our results indicate a strong relationship between media tones, particularly social media, and inflation expectations. SPF forecasts are mostly relevant for seniors aged 65 and above. Results suggest that media sources play a more significant role in shaping inflation expectations than SPF forecasts.

Exploring the dynamics of media shocks using a local projection model, we find that the impact of media shocks on inflation expectations is gradual, supporting the notion of slow adaptation to new information. This reaction varies across groups, with professional forecasters showing a prolonged response compared to other groups. Finally, we employ a MIDAS model with media tones to predict public inflation expectations on both quarterly and monthly frequencies. Our results suggest that traditional news is a more reliable predictor for both quarterly and monthly inflation expectations than social media.

In conclusion, the results establish a strong link between news and inflation expectations using aggregate media tones. Our analysis further investigates the influence of specific news topics on inflation expectations. Using LDA and the FinBERT language model, we categorize news into such topics as "Inflation & Fed", "Gas & Food Prices", "Healthcare Costs", "Mortgages & Business", and "Education Cost". These topics exhibit varying levels of influence on different demographic groups. Less-educated individuals have greater exposure to topics such as "Inflation & Fed", "Healthcare Costs", and "Mortgages & Business", while the most elderly cohort display unsurprising sensitivity to the topics of "Healthcare Costs" and "Inflation & Fed".

Employing DMA, "Mortgages & Business" news is found to be particularly important for both younger (18-34) and older (65-97) age cohorts, with younger generations also sensitive to "Gas & Food Prices" and "Healthcare Costs". More volatility in the probabilities is observed for less educated individuals, particularly with topics such as "Mortgages & Business" and "Healthcare Costs".

The rest of the study is structured as follows: Section 2 provides an overview of the relevant literature, Section 3 presents our data in detail, Section 4 introduces our estimation methods and construction of our FinBERT media tones, Section 5 presents the results, and Section 6 concludes.

#### 2. Related literature

Understanding the drivers of inflation expectations is crucial for evaluating the effectiveness of monetary policy since inflation expectations play a crucial role in shaping economic outcomes and the effectiveness of monetary policy (Bernanke, 2007). Notably, inflation expectations can influence households' consumption-savings decisions (i.e., consumption Euler equation) and investment choices i.e., spending on durable goods (Weber et al., 2022). Furthermore, high inflation expectations can also amplify shocks, like the effects of oil shocks on the economy (Bernanke, 2022).

Traditionally, macroeconomics has relied on the Full Information Rational Expectation (FIRE) models, assuming homogeneity in economic agents' expectations, but recent literature reveals significant heterogeneity in households' expectations with upward biases in inflation expectations compared actual inflation figures (D'Acunto et al., 2023). Previous studies show significant dispersion of agents' inflation expectations across demographics such as gender, age, education, income, and cognitive abilities (D'Acunto et al., 2019, 2021, 2023; Das et al., 2020; Malmendier & Nagel, 2016; Weber et al., 2022). This body of work highlights the need to consider demographic factors when analyzing inflation expectations.

Extensive research underscores the importance of demographic factors in shaping inflation expectations. For instance, Cavallo, Cruces & Perez-Truglia (2017) find that some households base their inflation expectations on past supermarket prices rather than official inflation statistics. D'Acunto, Malmendier & Weber (2021) further highlight this connection by revealing a systematic disparity in inflation expectations between genders in which women tend to have higher inflation expectations in households where they handle shopping alone. Malmendier & Nagel (2016) discover that personal experience helps to explain the difference in inflation expectations between different age cohorts. Research further indicates that inflation expectations may differ across education and income levels (Das et al., 2020; Weber et al., 2022) as well as male cognitive abilities (IQ) (D'Acunto et al., 2019, 2022). Survey questions and respondents' economical literacy and understanding significantly impact the inflation expectations formulation (Weber et al., 2022). Existing evidence supports the heterogeneity of inflation expectations among individuals. Our study contributes this literature and examines how age and education influence different groups' inflation expectations in response to new information presented in the media.

The presence of heterogeneity in inflation expectations challenges the assumption of frequent and noiseless updates of beliefs, as assumed by the FIRE paradigm. Notably, households update their inflation expectations by processing new information and reconditioning their future expectations from past values, i.e., learning from their previous events and decisions (Evans & Honkapohja, 2001; Malmendier & Nagel, 2016). However, this updating includes noisy or rigid elements. Coibion & Gorodnichenko (2015) classify households' expectations-forming processes into sticky information models, where collecting new information is expensive and time-consuming resulting in infrequent updates, and noisy information models, where households form expectations from noisy information sets via signal extraction and update their expectations more frequently (Mankiw & Reis, 2002; Coibion & Gorodnichenko, 2015).

Households obtain new information through media sources, including traditional news and social media, which act as intermediaries refining and presenting new information. In the past, traditional news sources like TV and newspapers were primary sources of economic information for households (Blinder & Krueger, 2004), which is a rational choice due to their low cost and time-efficiency (Lamla & Sein, 2014). However, in today's digital age, social media has emerged as a credible source of information for households (Coibion et al., 2022). Our question to be solved deals with the differences between the household's inflation expectations to news and social media depending on their level of education or age.

Carroll (2003) investigates the relationship between news and inflation expectations, introducing a sticky information model wherein some agents absorb information from newspaper articles on inflation. Carroll (2003) proposes an epidemiological model where agents face constant probability absorption rate of news articles. To approximate newspapers, Carroll (2003) utilizes Survey of Professional Forecasters (SPF) data, justifying its use as it represents expert opinions interviewed by news outlets and subsequently report these to households. Findings indicate that increased news coverage positively relates with more accurate inflation expectations, highlighting the importance of information absorption (Carroll, 2003). Our analysis reveals that professional forecasters' inflation expectations differ from those of the general public and are more closely aligned with social media than with traditional news.

To explore effects of media coverage on the discrepancy between German households and professional forecasters' inflation expectations, Lamla & Maag (2012) employ Bayesian learning methods and use a dataset from the media research institute, Media Tenor, which includes coded newspaper and television news. The dataset is used to construct an inflation "tone" to measure whether inflation is perceived to be rising or falling. Their findings reveal that media reporting about inflation significantly affects the disagreement of household inflation expectations, but it does not have the same impact on professional forecasters. (Lamla & Maag, 2012.) According to Lamla & Maag (2012), this discrepancy exists because professionals form their expectations independently of the media coverage of inflation. We contribute to this relationship between the public's inflation expectations and professional inflation forecasts.

Building upon Carroll's (2003) model, Lamla & Sein (2014) combine the "tone" with the sticky and noisy information models, by introducing noise in media coverage that presents a signal extraction problem for agents.

Employing Bayesian learning, they find that larger media coverage leads to more accurate inflation expectations. However, the content and tone of news stories also play a significant role, with biased news stories resulting in less accurate inflation expectations. (Lamla & Sein, 2014.) Our study contributes to this "tone" approach by examining the influence of the versatile and widely covered traditional news and social media news on the inflation expectations of different demographic groups.

Larsen et al. (2021) expand the noisy information model by incorporating information-rigid media, suggesting that more media coverage the less noisy is the information to agents. Using Dow Jones Newswire Archive text data to form a news corpus and examining news topics and tone in relation to households' inflation expectations measured by University of Michigan survey, they identify topics about technology and health as significant influencers of inflation expectations. They also find that many news topics with high predictive power for inflation and inflation expectations and that relevant news stories improve the accuracy of households' inflation expectations. We also detect from our FinBERT media tones that not all topics of inflation have a similar influence on agents' inflation expectations and the significance varies across groups. Along Larsen et al. (2021) also Lamla & Maag (2012) find no relationship between news topics and Survey of Professional Forecasters (SPF). This discrepancy arises because professionals form their expectations independently of the media coverage of inflation. Our evidence somewhat supports this as we find that SPF is more closely aligned with social media.

In a different context, Dräger & Nghiem (2021) find positive relationship between inflation expectations and current consumption, indicating that higher expected inflation is associated with higher planned consumption. Their results suggest that monetary and financial news leads to slightly more accurate inflation predictions and influences both expectation formation and spending decisions. They also study news effects across demographics, revealing that news significantly affected consumers who save, as well as those with lower initial information and lower inflation forecast accuracy. (Dräger & Nghiem, 2021.) We extend this literature and examine how the aggregate media tones and specific news topics influence households with different level of schooling and age.

Furthermore, the credibility and format of news sources can significantly influence inflation expectations, as observed by Coibion et al. (2022). Their survey in the US reveals that households tend to dismiss news articles from US Today articles even though it is a nonpartisan news company compared to those of reported by FED directly. This effect is particularly pronounced among low-income and low-educated individuals, and the study suggests that newspapers are perceived as less credible than government, friends and co-workers, social media, or television. (Coibion et al., 2022.) This highlights the importance of the source of news. We utilize an aggregate media tone, consisting of both traditional news and social media, allowing us to differentiate and aggregate the effects of these sources.

# 3. Data

#### 3.1 Inflation and heterogeneity of inflation expectations

To measure inflation, we rely on the headline inflation measure provided by the US Bureau of Labor Statistics<sup>2</sup>. This metric enables us to accurately capture inflation rates and explore the interactions between inflation expectations, media tone, and economic outcomes.

To investigate the heterogeneity of the public's inflation expectations, we use data from the University of Michigan's Survey of Consumers' inflation expectations survey. This survey provides mean inflation expectations across different education and age groups. By asking respondents about their expected price changes over the next year, we analyze how inflation expectations vary across different demographic groups. Our dataset covers quarterly observations from Q02:1998 to Q02:2017, providing a rich source of data for our analysis.



Figure 1: One-year-ahead inflation expectations by education in the US.

<sup>&</sup>lt;sup>2</sup> We use the "CPI for All Urban Consumers 1-Month Percent Change", aggregated to quarterly levels by averaging over three months.



Figure 2. One-year-ahead inflation expectations by age in the US.

The Figures 1 and 2 indicate significant heterogeneity of inflation expectations across education and age groups. Figure 1 reveals that households with higher levels of education exhibit lower inflation expectations, possibly due to their greater financial literacy and ability to comprehend inflation dynamics. Meanwhile, Figure 2 confirms previous findings by Malmendier and Nagel (2016) that inflation expectations differ among age cohorts, although the variation is not as pronounced as across different levels of education.

#### 3.2 Media tones and the Survey of Professional Forecasters (SPF)

As a measure of media, we utilize the traditional news and social media news tones from Refinitiv (Refinitiv Marketpsych's media tone data), which gives us insights into the media information and tone of economic articles. This dataset is a rich source of information, generated through the textual analysis of English-language financial and economic news from 2000 global media and 800 social media sites. This dataset has been available since 1998 and it is highly dimensional, providing multiple sentiments and emotions that enable us to understand media sentiment towards economic factors in a nuanced manner. A more detailed description of media tone data is in Appendix A. To ensure a comprehensive understanding of long-term trends, we aggregate the daily media tone data into quarterly levels using the three-months average method. Figure 3 displays the media tone data for traditional news and social media.



Figure 3. Media tones of news and social media in the US.

Figure 3 demonstrates significant variation in both media tones across both traditional news and social media news. Notably, one of these largest decreases occurred in 2000 and late 2001, which may reflect the Federal Reserve's monetary policy tightening of 1999-2000 and the following recession in 2001. The other decrease in media tone occurred after the Global Financial Crisis (GFC) of 2008 when the Federal Reserve rapidly decreased the federal funds rate and initiated the unconventional monetary policy in quantitative easing program (QE1). These fluctuations in media tone capture the sentiment within the media regarding economic conditions, which could potentially influence both household behavior and inflation expectations. A higher media tone corresponds to more positive sentiment about the economic conditions reported in the media.

In addition to the media tones, we also employ the Survey of Professional Forecasters (SPF) one-year-ahead inflation expectations. The SPF is a widely recognized and utilized source of data, providing valuable insights into inflation expectations from a professional perspective. SPF is obtained from the Philadelphia Fed database. Figure 4 displays the evolution SFP inflation forecasts. Furthermore, the inclusion of this dataset enables us to compare the professional expectations with media information embedded in the media tones.



Figure 4. SPF one-year-ahead inflation expectations.

The SPF one-year-ahead mean inflation expectations in Figure 4 mirror the media tones depicted in Figure 3. Specifically, both the tone of media and professionals' inflation expectations rise during the 1999-2001 period and subsequently decrease. However, the decrease in expectations occurs later for professionals, reaching its lowest point in late 2004.

Similarly, during the GFC in 2008, the decrease in professionals' inflation expectations and the tone of news move in parallel, which potentially reflects the influence of news on professional inflation expectations. By uncovering similarities in the inflation expectations of professionals and in the tone of the media, our analysis provides a more nuanced understanding of the drivers of inflation expectations.

#### 4. Methods

#### 4.1 Inflation expectations and inflation

We assume that households gain valuable information of the future inflation dynamics from the news and social media. Thus, they believe that the information provided by the media is more accurate than the information they generate from their adaptive rational forecasts based on past inflation. Agents expect that inflation is the fundamental value plus an error term, which captures the unforecastable transitory inflation shocks, both of which are unforecastable. Additionally, inflation beyond the next period is considered a random walk. The transitory shock has no long-run impact, and only the permanent component impacts inflation (Carroll 2003; Blanchard, 2013).

Agents form their inflation expectations at time t for inflation at a future time t + j (where j represents the number of periods ahead) by augmenting their rational expectation inflation with media tone about inflation expectations at time t. Following Carroll (2003, pp. 272-278), we can derive the following Equation 1 for the population meanvalue of their inflation expectations, denoted as  $M_t^e[\pi_{i,t,t+j}]$ :

(1) 
$$M_t^e[\pi_{i,t,t+j}] = \lambda_i M T_t[\pi_{i,t,t+j}] + (1-\lambda_i) M_{t-1}^e[\pi_{i,t,t+j}]$$

Here,  $\lambda_i$  stands for the fraction of the population *i* (where *i* denotes different education levels and age cohorts) who base their yearly inflation expectations in quarterly data from media, which we model with media tone  $MT_t$  of news and social media. In quarterly data the inflation rate for quarter (t + 1) at annual rate is  $\pi_{t,t+1} = 4[\log p(t + 1) - p(t)] = 4\pi_{t+1}$ . The term  $(1 - \lambda_i)$  indicates the share of the population *i* that bases their expectations on period *t* inflation on their views about inflation on period (t - 1) of (t + 1) inflation rate. From there Equation 2 can be derived (Carroll, 2003, pp. 273-278), where the agent's quarterly expectations for yearly inflation are as

(2) 
$$M_t^e[\pi_{i,t,t+4}] = \lambda_i M T_t[\pi_{i,t,t+4}] + (1-\lambda_i) M_{t-1}^e[\pi_{i,t,t+3}]$$

Our aim is to reveal the potential heterogeneity in inflation expectations  $\pi_{i,t,t+4}^{e}$  among different age cohorts and education groups *i*. To achieve this, we employ rigorous econometric methods, estimating the following Equation 3:

(3) 
$$\pi^{e}_{i,t,t+4} = \beta_{1,i} + \beta_{2,i} [MT_{t,t+4}] + \epsilon_{i,t}$$
  
(4)  $\pi^{e}_{i,t,t+4} = \beta_{1,i} + \beta_{2,i} [MT_{t,t+4}] + \beta_{3,i} [\pi^{e}_{i,t-1,t+3}] + \epsilon_{i,t}$ 

where  $\pi_{i,t+4}^{e}$  represents yearly inflation expectations of various demographic groups, and media tone. In top that we also estimate a specification in Equation 4, which is augmented a model with the last period realized inflation figure  $\beta_{4,t-1}[\pi_{t,t+4}]$ . That specification indicates how much of the agents' inflation expectations are generated from the media tone inflation information, agents' lagged inflation expectations and the last period official inflation

figures. Accordingly, it reveals the exposure of inflation expectations to the last period headline inflation and the stickiness of inflation expectations.

It is also important to know whether the media tones are able to forecast the actual future inflation which rationalizes the agents' formation of their inflation expectations based on media tone inflation forecasts. Hence, we investigate the media tones' ability to explain the actual US inflation using the following Equation 5:

(5) 
$$\pi_{t+1} = \beta_1 + \beta_2 [MT_t] + \epsilon_{t+1}$$

where  $\pi_{t+1}$  denotes actual inflation rates at time t. We also apply Equation 6 to conduct one-step-ahead expanding window forecasting exercises where the initial sample is 12 quarters. We analyze forecasting results through visual performance and by their Root Mean Squared Errors (RMSE).

(6) 
$$\pi_t = \beta_1 + \beta_2 [MT_{t-1}] + \epsilon_t$$

In addition, using media tone data, we also study Carroll's (2003) statement regarding the SPF's ability to proxy news to the public. To do this, we employ Equations 7 and 8 to analyze the relationship between the media tones and SPF.

(7) 
$$SPF_t = \beta_1 + \beta_2[MT_t] + \epsilon_t$$
  
(8)  $SPF_t = \beta_1 + \beta_2[MT:News_t] + \beta_3[MT:Social media_t] + \epsilon_t$ 

Furthermore, we assess the relative explanatory power of media tones' inflation information and professional inflation expectations  $SPF_t$  over the inflation expectations  $\pi^e_{i,t+4}$ .

We delve deeper into the transmission of inflation information between media tones and SPF on inflation expectations, illuminating the mechanisms behind the formation of these expectations. Initially, we estimate the contemporaneous relationship between media tones and  $SPF_t$  expectations. Under condition (1), if  $SPF_t$  is a channel for media tones, they should be related, Equation 7. In addition, under condition (2),  $SPF_t$  should also be able to predict  $\pi^e_{i,t+4}$ , as specified in Equation 8.

(9) 
$$\pi_{i,t+4}^{e} = \beta_{1,i} + \beta_{2,i} [SPF_t] + \epsilon_{i,t+4}$$
  
(10)  $\pi_{i,t+4}^{e} = \beta_{1,i} + \beta_{2,i} [MT: News_{t,t+4}] + \beta_{3,i} [SPF_t] + \epsilon_{i,t+4}$ 

Subsequently, under condition (3),  $SPF_t$  should contain the information of "Media tone: News" when both are used to predict  $\pi_{i,t+4}^e$ , as shown in Equation 10. If conditions (1) and (2) hold true, and in Equation 10, if  $\beta_{2,i} =$ 0 and  $\beta_{3,i} \neq 0$ , then  $SPF_t$  acts as the transmission channel for Media tone: News on inflation expectations. We execute these tests in bidirectionally: first, assessing the predictive capability of  $SPF_t$  on Media tone: News, and then evaluating the predictive influence of Media tone: News on  $SPF_t$  respectively. This approach examines the extent to which agents' inflation expectations are derive from the professional forecasters' inflation expectations  $\beta_{3,i}$  and the impact of media  $\beta_{2,i}$  on those expectations.

#### 4.2 Local projection

To analyze the dynamicity of the impact of media on inflation expectations, we employ a linear local projection (LP) model, originally developed by Jordá (2005). This method has gained widespread popularity in macroeconomics, e.g., notably utilized by Auerbach and Gorodnichenko (2013). The LP model allows us to estimate the dynamic effects of a shock on the inflation expectations over specific time horizon. Our approach is inspired by Ramey and Zubairy (2018), who apply somewhat similar model to investigate the effect of US military news shocks on the US economy, among other applications. Our LP model is outlined in Equation 11:

(11) 
$$y_{i,t+h} = \alpha^h + \beta_h [Media \ shock_t] + \phi x_t + u_{t+h}^h$$

Here,  $y_{i,t+h}$  represents inflation expectations of *i* group *h* periods ahead,  $\alpha^h$  is the intercept,  $\beta_h$  is the effect of the media shock at time *t*,  $\phi x_t$  represents control variables and  $u_{t+h}^h$  is the error term. The model has 4 lags in the estimation with 12 quarters (3 years) rolling estimation horizon, zero trend, and a 1.96 confidence interval. We use the media tone from both news and social media to proxy *Media shock*. This allows us to assess the impact of a media shock on the inflation expectations of various groups.

The evolution of the time-path of  $y_{t+h}$  describes information rigidity, as per the sticky information model of Mankiw & Reis (2002) and the noisy information model proposed by Woodford (2003) and Sims (2003). This analysis reveals how strong and rapidly the new inflation information is transmitted to inflation expectations. Agents update their inflation expectations with each estimation period's media tone information, forming a weighted average of their prior beliefs (informed by the lagged period information of headline inflation,  $\pi_{t-1}$ ), a constant reflecting the inflation target, and the latest media shock information.

Following Coibion and Gorodnichenko (2015), we assume inflation follows an AR (1) process  $\pi_t = \rho \pi_{t-1} + \varepsilon_t$ , where  $\varepsilon_t$  is a normally distributed (*i.i.d*) innovation of  $\pi_t$  and  $\rho < 1$ . Consumer cannot observe the overall headline inflation  $\pi_t$  directly but they receive signals about it through  $MT_t = \pi_t + \omega_{it}$ , where  $\omega_{it}$  represents a mean-zero, normally distributed noise (*i.i.d*) across time and agents. Agents form forecasts  $E_{it}\pi_{t+h}$  based on the  $MT_t$  and controls like last period inflation and a constant.

(12) 
$$E_{it}\pi_{t+h} = \alpha^h + \beta_h[MT_t] + \phi[\pi_{t-1}] + u_{t+h}^h$$

In Equation 12,  $\beta_h$  captures new information's impact relative to the previous forecast. A fully accepted media shock implies  $\beta_h = 1$ ; uncertainties or agents related frictions regarding media shock information suggest  $\beta_h < 1$ . Incorrect signals may bias agents' expectations, with no feedback from forecast errors, potentially leading to persistent forecast biases. Our focus is on agents' inflation expectations responses to media shock, considering the information source type.

We investigate how media shock influences both professional and public inflation expectations, recognizing that media coverage might affect these groups differently. We employ two LP model specifications: the first with  $\phi = 0$ , excluding control variables, and the second with  $\phi \neq 0$ , incorporating the last period's realized headline inflation

 $[\pi_{t-1}]$  as a control. This enables us to estimate agent's reactions to media shock and preceding period's headline inflation.

The rapid integration of new media shock information into inflation expectation reveals a high sensitivity of inflation expectations to new information. Conversely, sluggish response suggests rigid adaptation to new information. Understanding the speed at which inflation expectations adjust is crucial for monetary authorities, as it affects how inflation expectations respond to new information and evolve during business cycles.

#### 4.3 Micro-based FinBERT media tones

The media tone variable for traditional news is constructed from thousands of different media sources and providing a single measure of news constituting thousands of individual information pieces related to various terms (see Appendix A). To enhance the robustness of our findings, we create our own micro-based FinBERT media tones from inflation-related news articles, compiled directly from the Lexis Nexis Uni database. Following Gorodnichenko et al. (2023), we utilize Financial BERT (FinBERT) language model, developed by Araci (2019).

To accurately select relevant news from Nexis Uni database, we target articles containing one or more of the following terms in their headlines or main text: *"inflation"*, *"price level"*, *"inflation expectations"*, *"FED"*, *"interest rate"*, *"consumer price index"*, *"CPI"*, *"consumer prices"*, *"energy prices"*, *"food prices"*, *"monetary policy decisions"*, *"monetary policy"*, *"fiscal policy"*, *"wage growth"*, *"healthcare costs"*, *"healthcare expenses"*, *"education expenses"*, *"education costs"*, *"gas prices"*, *"cost of living"*, *"rate of inflation"*, *"inflationary"*, *"purchasing power" or "inflation rate"*. Additionally, we require that one of these terms appear within 100 words of the following keywords: "U.S", "United States", "US", "Federal reserve" or "FED" or that the headline or the main text includes terms like "FED interest rate", "federal reserve", "federal reserve", "federal reserve".

We further refine the selection by specifying the language as "English" and focusing on "Web-based Publications" or "Newspapers or Magazines & Journals" published in "North America, United States". Following these criteria and an after initial cleaning, we compile a comprehensive corpus of 76,750 news articles, covering the periods from December 31, 1997, to February 4, 2018.

To enhance the relevance of our dataset, we clean it by removing articled unrelated to economic inflation, such as those discussing tire or grade inflation, and articles concerning countries other than the US. Subsequently, we transform the news articles into sentences and clean the news corpus by removing emails, phone numbers, timestamps, and other extra characters, such as punctuations. We then select sentences between 10 and 100 words. and remove common stop words.

Finally, we apply the sentences to a Latent Dirichlet Allocation (LDA) topic model, similar to that of Larsen et al. (2021). Utilizing the LDA model, which employs Gibbs sampling and K=15 topics<sup>4</sup>, allows us to systematically extract topics from our news data. The derived LDA topics are illustrated in Figure 5. Essentially, we integrate the

<sup>&</sup>lt;sup>3</sup> Previously Larsen et al. (2021) find that healthcare costs and education expenses news are related to inflation expectations. <sup>4</sup> We conduct a grid search with subsample of our data and calculate the perplexity score for different combinations to find optimal number for K, which is 15.

methodologies of Gorodnichenko et al. (2023) and Larsen et al. (2021) to construct a dataset that accurately reflects the media tone of inflation news in media sources.<sup>5</sup>



Figure 5. News topics from the LDA model.

Figure 5 indicates that the LDA model categorizes the words from the news sentences into 15 distinct topics, highlighting the 10 most frequent words for each. Based on these topics, we select news articles corresponding the most frequent words as follows: topics 10 and 14 are identified as the "Inflation & Fed"; topics 4 and 5 are associated with "Gas & Food Prices"; topic 12 is designated "Healthcare Costs"; topic 3 is labeled "Mortgages & Business"; and topic 11 is classified as "Education Cost". Table 1 lists the distribution of news articles in each topic, sourced from approximately 1,300 different news outlets, including CNN, The Economist, Newsweek, Forbes, and Politico, among others.

 $<sup>^{5}</sup>$  Haavio et al. (2024) similarly combine the LDA and FinBERT methods to create an inflation text tone for studying the ECB's loss function.

	Inflation & Fed	Gas & Food Prices	Healthcare Costs	Mortgages & Business	Education Costs
Number of news					
articles	17 138	10 257	4 378	5 730	5 524

Table 1. Number of news articles in selected topics.

As outlined in Table 1, our FinBERT analysis focuses on 5 distinct news topics ("Inflation and Fed", "Gas & Food Prices", "Healthcare Costs", "Mortgages & Business", "Education Cost"). These news articles are then classified by FinBERT language model following Gorodnichenko et al. (2023). FinBERT, trained on financial news<sup>6</sup>, classifies texts with a Softmax probability of "negative", "positive", or "neutral" (Araci, 2019). Notably, FinBERT offers significant advantage to other language models, such as GPT-models, by enabling classification of large volumes of text data without any rate limits. We determine the tone of news articles in each topic as "positive" if it had the highest probability of being so. Following the approach of Gorodnichenko et al. (2023), we calculate the tone for each date by subtracting the count of negative articles from positive ones and dividing by the total number of articles. The tones for each news article are then normalized by the number of words. Finally, we aggregate these daily FinBERT tones to quarterly levels using a three-month averaging, covering the period from Q2:1998 to Q4:2017. These quarterly FinBERT media tones are illustrated in Figure 6.

According to Figure 6, all our FinBERT media tones are negative, with the mean of the inflation news tones categories approximately -10, indicating that financial news over the 20-year span from 1998 to 2017 was predominantly negative. Notably, news related to "Inflation & Fed" is most negative during the GFC. Similarly, news related to "Gas & Food Prices" shows significant declines during the GFC especially in Q3:2008 and Q4:2008, coinciding with the sharp rise in gas prices in June-August 2008. For instance, in July 2008 the price of one barrel of West Texas Intermediate crude oil was double (140 \$) compared to that of year earlier (70 \$). Consequently, our FinBERT media tones, a measure of inflation news, interprets high inflation as negative news. Table 2 shows the correlation between our measures of FinBERT media tone news. The strongest correlations are observed between "Healthcare Costs" and "Education Costs" (0.53), and between "Inflation & Fed" and "Mortgages & Business" (0.47) news.

<sup>&</sup>lt;sup>6</sup> Training data comes from both Thomson Reuters news stories and Financial PhraseBank dataset from Malo et al. (2014).



Figure 6. Quarterly FinBERT media tones.

	Inflation & Fed	Gas & Food Prices	Mortgages & Business	Healthcare Costs	Education Costs
Inflation & Fed	1	0.41	0.48	0.13	0.2
Gas & Food Prices	0.41	1	0.32	0.25	0.46
Mortgages & Business	0.48	0.32	1	0.21	0.3
Healthcare Costs	0.13	0.25	0.21	1	0.53
Education Costs	0.2	0.46	0.3	0.53	1

Table 2. Correlation matrix.

We further refine our FinBERT media tone by selecting sentences from newspapers that include the word "inflation". Following the initial data cleaning as described previously, we process 68,340 sentences through the FinBERT model. We develop a more focused measure of inflation news tone by calculating difference between positive and negative probabilities assigned by FinBERT, adjusted by the sentence length. Subsequently, we aggregate the tones for each news article, normalize them by the number of words, and then compile these into quarterly levels. The sentence-based FinBERT media tone is presented in Figure 7.



Figure 7. Quarterly sentence-based FinBERT media tone.

The sentence-based FinBERT media tone exhibits the strongest correlation with "Mortgages & Business" reaching 0.559. The other notable correlations include "Inflation & Fed" at 0.474, "Gas & Food Prices" at 0.382, "Healthcare Costs" at 0.265, and "Education Costs" at 0.211.

# 5. Empirical results

We begin by examining the role of media tones on the inflation expectations of different age cohorts. Specifically, the influence of various sources of media tone – news and social media – on these cohorts is detailed in Tables 3, 4, 5, and 6, based on the insights derived from Equations 3 and 4.

	Dependent variable					
-	18-34	35-44	45-54	55-64	65-97	
	(1)	(2)	(3)	(4)	(5)	
News	10.514***	10.210***	10.965***	11.421***	10.948***	
	(2.217)	(2.118)	(2.408)	(2.300)	(2.129)	
Constant	2.403***	2.395***	2.505***	2.529***	2.636***	
	(0.226)	(0.216)	(0.245)	(0.234)	(0.217)	
Observations	79	79	79	79	79	
R <sup>2</sup>	0.226	0.232	0.212	0.243	0.256	
Adjusted R <sup>2</sup>	0.216	0.222	0.202	0.233	0.246	

Table 3. The role of news on inflation expectations of different age cohorts.

	Dependent variable					
_	18-34	35-44	45-54	55-64	65-97	
	(1)	(2)	(3)	(4)	(5)	
AR-1	0.409***	0.491***	0.590***	0.562***	0.517***	
	(0.091)	(0.086)	(0.078)	(0.078)	(0.081)	
News	9.412***	8.576***	9.186***	9.831***	9.333***	
	(2.034)	(1.824)	(1.870)	(1.828)	(1.774)	
Constant	1.109***	0.895***	0.573*	0.639**	0.875***	
	(0.354)	(0.325)	(0.319)	(0.320)	(0.325)	
Observations	78	78	78	78	78	
R <sup>2</sup>	0.387	0.457	0.550	0.551	0.522	
Adjusted R <sup>2</sup>	0.371	0.443	0.538	0.539	0.509	

Table 4. The role of AR-1 and news on inflation expectations of different age cohorts.

Table 5. The role of social media on inflation expectations of different age cohorts.

	Dependent variable						
	18-34	35-44	45-54	55-64	65-97		
	(1)	(2)	(3)	(4)	(5)		
Social media	11.110***	12.772***	10.755***	10.353***	9.800***		
	(2.695)	(2.457)	(2.962)	(2.891)	(2.693)		
Constant	2.551***	2.383***	2.724***	2.822***	2.928***		
	(0.223)	(0.203)	(0.245)	(0.239)	(0.223)		
Observations	79	79	79	79	79		
R <sup>2</sup>	0.181	0.260	0.146	0.143	0.147		
Adjusted R <sup>2</sup>	0.170	0.250	0.135	0.132	0.136		

*Note:* Numbers in parenthesis are standard errors. \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Results in Table 3 suggest that traditional news has a significant role for all age cohorts' inflation expectations. The overall news has the largest association on the inflation expectations of 55-64 age cohort. Since the sizes of coefficients are quite alike, we do not provide strong evidence on heterogenous connections of inflation expectations on news. However, social media, in Table 5, has the largest effects on the younger 18-34 and 35-44 age cohorts. This evidence is somewhat expected as the older generation are more likely exposed to information from traditional news outlets and younger from social media. The constant varies from 2.4 (35-44 years old) up to 2.6 (65-97 years old) suggesting that without the news, different groups' inflation expectations are slightly above the 2 percent inflation target.

Tables 4 and 6 incorporate the AR-1 process, yet the influence of media tones on inflation expectations remains significant. This suggests that media tones effectively convey timely information about future inflation, shaping agents' expectations. Next, we will present the results for the role of media tones on different education groups.

Tables 7 and 8 present the influence of news on different groups of level of education. We have strong evidence of a significant relationship between the news and the expectations of all education groups. While variations in exposure to this information exist among the groups, these differences are relatively minor. Individuals with "Graduate studies" show the highest responsiveness to news, whereas those with only "High school or less" are less responsive. Although the variances across all groups are minimal, the distinction between the highest and lowest education levels is notable. The estimated constant, indicating baseline inflation expectations in the absence of news, is lower for individuals with "Graduate studies", decreasing progressively with lower education levels. The figures lend support that the credibility of the two percent inflation target of the FED, ca. 2% is highest among the "College degree" group, see Table 8.

Tables 9 and 10 display the influence of the social media only. The results of the exposure to social media are slightly different compared to those of the news. Here, the smallest influence is on the group of "Some college" education. Similar to previous, the social media has the highest response among individuals with "Graduate studies". The difference between the "High school or less" and "Some college" is statistically significant. The two higher educated groups also show strongest credibility of the FED two percent inflation target. The significance of the social media on agents' inflation expectations remains valid even though the AR-1 processes are included into the models. Our analysis suggests that regardless of education level, all groups' inflation expectations are influenced by media tone of traditional news and social media news. In summary, media tones are positively associated with inflation expectations, revealing a modest degree of variation in their influence across different educational groups.

_	Dependent variable					
	18-34	35-44	45-54	55-64	65-97	
_	(1)	(2)	(3)	(4)	(5)	
AR-1	0.417***	0.453***	0.601***	0.587***	0.557***	
	(0.093)	(0.086)	(0.081)	(0.083)	(0.086)	
Social media	10.056***	10.701***	9.207***	9.136***	8.528***	
	(2.435)	(2.133)	(2.279)	(2.275)	(2.197)	
Constant	1.212***	1.022***	0.711**	0.790**	0.970***	
	(0.358)	(0.308)	(0.329)	(0.342)	(0.351)	
Observations	78	78	78	78	78	
R <sup>2</sup>	0.358	0.474	0.512	0.488	0.455	
Adjusted R <sup>2</sup>	0.341	0.460	0.499	0.475	0.440	

Table 6. The role of AR-1 and social media on inflation expectations of different age cohorts.

	Dependent variable					
	High school or less	Some college	College degree	Graduate studies		
	(1)	(2)	(3)	(4)		
News	8.955***	11.405***	11.059***	12.598***		
	(2.324)	(2.410)	(2.073)	(2.019)		
Constant	3.214***	2.597***	2.071***	1.771***		
	(0.237)	(0.246)	(0.211)	(0.206)		
Observations	79	79	79	79		
R <sup>2</sup>	0.162	0.225	0.270	0.336		
Adjusted R <sup>2</sup>	0.151	0.215	0.260	0.327		

Table 7. The role of news on different education groups inflation expectations.

Table 8. The role of AR-1 and news on different education inflation expectations.

	Dependent variable					
_	High school or less	Some college	College degree	Graduate studies		
	(1)	(2)	(3)	(4)		
AR-1	0.597***	0.561***	0.494***	0.475***		
	(0.082)	(0.079)	(0.082)	(0.079)		
News	7.733***	9.384***	9.683***	10.387***		
	(1.825)	(1.921)	(1.759)	(1.744)		
Constant	0.893**	0.722**	0.648**	0.563**		
	(0.369)	(0.332)	(0.293)	(0.265)		
Observations	78	78	78	78		
$\mathbb{R}^2$	0.508	0.530	0.509	0.550		
Adjusted R <sup>2</sup>	0.495	0.518	0.496	0.538		

	Dependent variable					
	High school or less	Some college	College degree	Graduate studies		
	(1)	(2)	(3)	(4)		
Social media	9.048***	9.379***	10.413***	12.327***		
	(2.816)	(3.054)	(2.609)	(2.568)		
Constant	3.372***	2.965***	2.325***	2.024***		
	(0.233)	(0.253)	(0.216)	(0.213)		
Observations	79	79	79	79		
R <sup>2</sup>	0.118	0.109	0.171	0.230		
Adjusted R <sup>2</sup>	0.107	0.098	0.161	0.220		

**Table 9.** The role of social media on different education groups inflation expectations.

		Dependent variable					
	High school or less	Some college	College degree	Graduate studies			
	(1)	(2)	(3)	(4)			
AR-1	0.606***	0.599***	0.525***	0.497***			
	(0.083)	(0.082)	(0.086)	(0.086)			
Social media	8.227***	9.084***	9.360***	9.736***			
	(2.177)	(2.339)	(2.172)	(2.207)			
Constant	0.970**	0.787**	0.763**	0.748***			
	(0.374)	(0.354)	(0.312)	(0.282)			
Observations	78	78	78	78			
R <sup>2</sup>	0.488	0.484	0.447	0.473			
Adjusted R <sup>2</sup>	0.474	0.471	0.433	0.459			

Table 10. The role of AR-1 and social media on different education inflation expectations.

## 5.1 Media tone in explaining and forecasting actual inflation

We next explore the capacity of media tones to account for the actual headline Consumer Price Index (CPI) inflation. The findings, derived from Equation 5, are outlined in Table 11.

	Dependent variable			
	Inflation (CPI)	Inflation (CPI)		
	(1)	(2)		
News	3.009***			
	(0.684)			
Social media		3.156***		
		(0.829)		
Constant	-0.111	-0.067		
	(0.070)	(0.069)		
Observations	79	79		
R <sup>2</sup>	0.201	0.158		
Adjusted R <sup>2</sup>	0.191	0.148		

 Table 11. Explanatory power of media tone for actual inflation.

*Note:* Numbers in parenthesis are standard errors, \*p<0.1; \*\*p<0.05; \*\*\*p<0.01.

Table 11 offers an assessment of the power of media tone in explaining contemporary inflation. It shows that both media tone variables are highly significant, and they have a positive relation to the actual US inflation, indicating that the news they convey aligns with the prevailing inflation levels. However, the Adjusted R-squared are relatively low, with the highest R-square, attributed to the news, being 0.191.

Figure 8 displays the actual headline inflation against the forecasts derived from the news. Although the inflation forecasts mimic the headline inflation dynamics, the true headline inflation changes largely exceed the forecasts. The social media inflation forecasts are presented in Figure 9, where the model produces quite similar predictions.

Comparatively, the news forecasting model produces marginally lower RMSE value (0.2156856) than the social media model (0.2192504). The forecasted and the actual values are weakly related; occasionally the social media is better able to capture the headline inflation dynamics (like in 2009), while in other cases, the news is more accurate (as in 2011).



Figure 8. News forecasting inflation.



Figure 9. Social media forecasting inflation.

## 5.2 Media tone and SPF forecasts

Next, we display the relationship between the media tones and the SPF's inflation expectations, i.e., the results for Equation 7 and 8. Table 12 indicates news is efficient in explaining the SPF's inflation forecasts. Interestingly, social media is more significantly related to the SPF expectations than the news.

	Dependent variable			
	SPF	SPF	SPF	
	(1)	(2)	(3)	
News	2.731***		-1.194	
	(0.914)		(1.104)	
Social Media		5.691***	6.671***	
		(0.939)	(1.304)	
Constant	1.913***	1.733***	1.772***	
	(0.093)	(0.078)	(0.085)	
Observations	79	79	79	
R <sup>2</sup>	0.104	0.323	0.333	
Adjusted R <sup>2</sup>	0.092	0.314	0.316	

Table 12. Media tones explaining the SPF one-year-ahead.

*Note:* Numbers in parenthesis are standard errors. \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Carroll (2003) states that SPF reflects informational content of news because professionals are interviewed to by the news outlets. Hence, Carroll expects that the professionals' opinions are transmitted and reflected in the media tone information. Contrary to this expectation, our findings in Table 12 reveal more complex picture. While the coefficient for news is statistically significant, both R-squared and the Adjusted R-squared values are relatively low. Conversely, social media exhibits a larger, statistically significant coefficient and a higher R-squared. With the inclusion of both media tone variables in the model, the statistical significance of news diminishes, indicating that SPF inflation information predominantly relates with social media rather than traditional news.

This suggests a revaluation of Carroll's (2003) hypothesis that SPF expectations directly reflect the inflation information provided by news outlets. Instead, our results imply that SPF inflation expectations are more closely aligned with social media. This could be attributed to the growing influence of social media channels among professionals, making social media more important source of inflation expectations compared to traditional news.

Our earlier results in Table 5 and 6 indicate that social media largely influenced inflation expectations of individuals aged 35-44 with "Graduate studies". These results now extend to suggest that social media also shapes the inflation expectations of professional forecasters. Next, we present our findings for the mechanism of the expectation process, specifically Equations 9 and 10.

		1	Dependent varial	le	
	18-34	35-44	45-54	55-64	65-97
	(1)	(2)	(3)	(4)	(5)
SPF one-year-ahead	0.731**	0.794***	0.667**	0.679**	0.224
	(0.286)	(0.270)	(0.311)	(0.302)	(0.290)
Constant	1.825***	1.650***	2.111***	2.152***	3.203***
	(0.627)	(0.594)	(0.682)	(0.663)	(0.636)
Observations	79	79	79	79	79
R <sup>2</sup>	0.078	0.101	0.056	0.062	0.008
Adjusted R <sup>2</sup>	0.067	0.089	0.044	0.049	-0.005

Table 13. SPF explaining inflation expectations by age cohort.

	Dependent variable					
	High school or less	Some college	College degree	Graduate studies		
	(1)	(2)	(3)	(4)		
SPF one-year-ahead	0.235	0.314	0.759***	1.011***		
	(0.298)	(0.321)	(0.273)	(0.269)		
Constant	3.565***	3.013***	1.485**	0.785		
	(0.655)	(0.705)	(0.599)	(0.590)		
Observations	79	79	79	79		
R <sup>2</sup>	0.008	0.012	0.091	0.155		
Adjusted R <sup>2</sup>	-0.005	-0.001	0.080	0.144		

Table 14. SPF explaining inflation expectations by education group.

*Note:* Numbers in parenthesis are standard errors, \*p<0.1; \*\*p<0.05; \*\*\*p<0.01.

Table 13 proposes a significant relationship between SPF and younger age cohorts, especially those aged 35-44. Table 14 indicates a more profound association between the SPF and more educated households, specifically those with college degrees and graduate studies. These findings suggest that younger, highly educated households share similar views about future inflation as experts do. We then incorporate our traditional news measure from media tone into the model (i.e., Equation 10) and study how this affects our results.

Including news in the model diminished the previously strong significance of the SPF for different age cohorts' inflation expectations, see Table 15. Only the 35-44 age group remained significant with a p-value below 0.1. Similar evidence is presented in Table 16, where incorporating news reduced the significance of the SPF, leaving only graduate studies with a p-value below 0.05. Thus, our evidence suggests that the influence of news is independent of demographics, as every age group is exposed to media tone information. Conversely, the SPF has a more significant relationship with households in the 35-44 age cohorts and those with graduate studies.

	Table 15.	News and	SPF expl	laining	inflation	expectations	by age cohort.
--	-----------	----------	----------	---------	-----------	--------------	----------------

		Dependent variable					
	18-34	35-44	45-54	55-64	65-97		
	(1)	(2)	(3)	(4)	(5)		
News	9.504***	8.973***	10.205***	10.676***	11.533***		
	(2.330)	(2.209)	(2.546)	(2.431)	(2.254)		
SPF one-year-ahead	0.370	0.453*	0.278	0.273	-0.214		
	(0.275)	(0.261)	(0.300)	(0.287)	(0.266)		
Constant	1.696***	1.528***	1.973***	2.006***	3.046***		
	(0.572)	(0.542)	(0.625)	(0.597)	(0.553)		
Observations	79	79	79	79	79		
R <sup>2</sup>	0.244	0.261	0.221	0.251	0.262		
Adjusted R <sup>2</sup>	0.224	0.242	0.200	0.232	0.243		

Table 16. News and SPF explaining inflation expectations by education group.

	Dependent variable					
	High school or less	Some college	College degree	Graduate studies		
	(1)	(2)	(3)	(4)		
News	9.275***	11.770***	10.027***	10.978***		
	(2.468)	(2.559)	(2.173)	(2.067)		
SPF one-year-ahead	-0.117	-0.134	0.378	0.593**		
	(0.291)	(0.302)	(0.256)	(0.244)		
Constant	3.439***	2.853***	1.348**	0.636		
	(0.606)	(0.628)	(0.534)	(0.508)		
Observations	79	79	79	79		
R <sup>2</sup>	0.163	0.227	0.290	0.384		
Adjusted R <sup>2</sup>	0.141	0.207	0.271	0.368		

We then conduct Granger causality tests with our media tones and employ the SPF as a proxy for news, following Carroll's (2003) suggestion. Prior to estimating the Granger causality tests, we assess the stationarity of the datasets using Augmented Dickey-Fuller (ADF) tests. Non-stationary series are differenced, and any missing values are removed to ensure the dataset's consistency and completeness. We perform the Granger causality tests using a one-lag approach, i.e., comparing Equations 13 and 14.

(13) 
$$\pi_{i,t,t+4}^{e} = \beta_{1,i} + \beta_{2,i}[\pi_{i,t-1,t+3}^{e}] + \epsilon_{i,t}$$
  
(14)  $\pi_{i,t,t+4}^{e} = \beta_{1,i} + \beta_{2,i}[\pi_{i,t-1,t+3}^{e}] + \beta_{3,i}[MT_{t,t+3}] + \epsilon_{i,t}$ 

Thus, we examine whether the previous values of news, social media, or SPF have predictive power for the current values of each inflation expectation. This allows us to study whether media information Granger-causes inflation expectations.

	Dependent variable					
	18-34 35-44 45-54 55-64 65-97					
	(1)	(2)	(3)	(4)	(5)	
News	0.011**	0.025**	0.051*	0.006**	0.110	
Social media	0.040**	0.009***	0.192	0.158	0.388	
SPF one-year-ahead	0.536	0.613	0.494	0.216	0.050*	

Table 17. Granger causality test p-values for inflation expectations by age cohort.

*Note:* \*p<0.1; \*\*p<0.05; \*\*\*p<0.01.

Table 18. Granger causality test p-values for inflation expectations by education group.

		Dependent variable					
	High school or less	Some college	College degree	Graduate studies			
	(1)	(2)	(3)	(4)			
News	0.183	0.160	0.005***	0.000***			
Social media	0.332	0.536	0.094*	0.013**			
SPF one-year-ahead	0.139	0.134	0.221	0.550			

*Note:* \*p<0.1; \*\*p<0.05; \*\*\*p<0.01.

The results in Tables 17 and 18 show the F-test p-values for the tests of Granger-causality. Our results suggest that news Granger-causes inflation expectations among some age and educated group of agents. We find the strongest evidence for our news, which demonstrates statistical significance in Granger-causing inflation expectations for the age groups 18-34, 35-44, 45-54, and 55-64.

Our results also show strong evidence of news influencing the inflation expectations of highly educated households, specifically those with "College degree" and "Graduate studies". The social media measure exhibits the most substantial influence on the younger age cohorts (18-34 and 35-44) and on highly educated households (those with "College degrees" and "Graduate studies"). The SPF shows statistically significant Granger-causality with the older age cohort (65-97). Thus, our results support the notion that news and social media Granger-cause changes in households' inflation expectations, particularly among the highly educated.

#### 5.3 Local projection

This section presents our results from a local projection (LP) model, which examines the impact of media shocks on inflation expectations. The model uncovers potential variations in the magnitude, significance, and persistence of reactions to media shocks proxied by news and social media from media tones. We estimate two distinct versions of our LP model. The first version includes each group's lagged inflation expectations as a control variable, along with a constant term. The findings are illustrated in Figures 10 and 11.



Figure 10. LP model's impulse responses to news shocks.

Figure 10 presents the trajectory of professionals' and the public's inflation expectations in response to a news shock over a 12-quarter time horizon. Initially, the impact of a news shock is positive across all groups. This positive effect, however, gradually diminishes, with the most significant decrease occurring around the 10th quarter after the shock.

The model variant that incorporates lagged headline inflation as a control variable assesses the impacts of new information provided via media tone on agents' inflation expectations. We assume that lagged actual CPI headline inflation reflects the true information set on which agents form their inflation expectations before they become exposed to the new media reports about inflation, see Figure 12.

Comparing the impulse responses to news shock without lagged inflation ( $\phi x_t = 0$ ) to those with it ( $\phi x_t \neq 0$ ), we see that the positive news shock typically dies out within five quarters. The inclusion of lagged inflation and prior information seems to enhance the impact and significance of media shocks. This supports the notion that agents adjust their prior expectations – based on the actual last period's inflation figures – with new information provided by the media. Overall, the impulse responses align with our regression analysis findings, showing similar reactions to news shocks across different age and education groups.

Regarding social media shocks (see Figures 11 and 13), the trajectory of social media shocks appears more volatile than that of news shocks. Additionally, incorporating lagged headline inflation as a control variable in the LP estimation seems to reduce the persistency of the social media shock approximately two quarters. Without the headline inflation data as a control, the social media shock dies out within five quarters, whereas with it, their impact fades within three quarters.

Furthermore, using lagged inflation as a control variable ( $\phi x_t \neq 0$ ) in the LP model appears to diminish the average magnitude of media shock responses across various age and education levels. This suggests that the formation of agents' inflation expectations is influenced not only by recent news and social media but also the lagged information (expectations) about the inflation. This implies that the inflation expectations would consist of persistency and memory.



Figure 11. Impulse responses of local projection model to social media shocks.



Figure 12. LP model with actual CPI inflation as control variable (news shock).

![](_page_36_Figure_0.jpeg)

Figure 13. LP model with inflation as control variable (social media shock).

#### 5.4 Estimation with mixed-data sampling (MIDAS)

Media tone data is available on a daily frequency, while the inflation expectations for different groups (age, education) are available only on a monthly or quarterly basis. This calls for estimation with mixed frequency. For instance, Chambers (2016) presents evidence suggesting that the use of an exact discrete time model for mixed frequency data yields estimators with lower bias and RMSEs compared to those based on traditional methods of aggregating data to the same lower frequency. The mixed frequency approach is particularly appealing in understanding the formation of inflation expectations. An economic agent continuously acquires various types of information about current and future inflation, forming their expectations based on this information. However, their expectations about future inflation are requested only on a monthly basis. To this end, we estimate a mixed-data sampling (MIDAS) regression, which can be represented in general equation form in Equation 15.

(15) 
$$y_t = \beta_0 + \sum_{k=0}^{K} \beta_{k+1} m(x_{t-k}, \theta) + \epsilon_t$$

In Equation 15,  $y_t$  is the predicted variable,  $\beta_{k+1}$  are the coefficients for the lagged values of the explanatory variables  $x_{t-k}$  and  $m(x_{t-k},\theta)$  represents the MIDAS lag polynomial applied to x at time t - k. The function m utilizes parameters  $\theta$  to weight the contributions of lagged values of x. The form of m and parameters  $\theta$  vary based on the selected MIDAS weighting scheme, such as the exponential Almon polynomial lag structure. (Ghysels et al., 2007.)

Furthermore, we implement the MIDAS model, utilizing both daily news and social media, to forecast the University of Michigan's monthly inflation expectations. This monthly series is more detailed than its quarterly counterpart, including three demographic categories for both age and education. To properly align our daily data within the MIDAS model, we use only the last 22 days of our daily media tone data from each month. We conduct an out-of-sample forecasting exercise with a recursive window, i.e., a rolling expanding window with an initial sample of 24 months with two lags, and the exponential Almon polynomial lag structure with starting parameter values of 1 and -0.5.

![](_page_38_Figure_0.jpeg)

- Actual - Forecast

Figure 14: MIDAS model recursive forecasts with daily news.

![](_page_39_Figure_0.jpeg)

— Actual — Forecast

Figure 15. MIDAS model recursive forecasts with daily social media.

News		Social media			
Inflation expectations	RMSE	Inflation expectations	RMSE		
Ages 18-34	0.841	Ages 18-34	0.845		
Ages 35-54	0.788	Ages 35-54	0.786		
Ages 55-97	0.770	Ages 55-97	0.785		
High school or less	0.825	High school or less	0.839		
Some college	0.871	Some college	0.878		
College degree	0.718	College degree	0.730		

Table 19. MIDAS models' RMSE scores for monthly inflation expectations.

Figures 14 and 15, and Table 19, illustrate the MIDAS model's recursive forecasts of inflation expectations based on daily news and social media. The estimates reveal that traditional news tends to predict monthly inflation expectations more effectively than social media sources.

Extending our analysis, we also apply the MIDAS model to forecast quarterly inflation expectations, consistent with our methodology in earlier sections. This involves a similar recursive window forecast, beginning with an initial sample of 8 quarters, and employing the MIDAS model with the same lag and start parameters. The results, showcased in Figures 16 and 17, and Table 20, highlight the forecasts for quarterly inflation expectations based on news and social media.

![](_page_40_Figure_4.jpeg)

Figure 16. MIDAS model recursive forecasts for quarterly inflation expectations (news).

![](_page_41_Figure_0.jpeg)

Figure 17. MIDAS model recursive forecasts for quarterly inflation expectations (social media).

News		Social media		
Inflation expectations	RMSE	Inflation expectations	RMSE	
Ages 18-34	0.719	Ages 18-34	0.737	
Ages 35-44	0.672	Ages 35-44	0.666	
Ages 45-54	0.766	Ages 45-54	0.808	
Ages 55-65	0.725	Ages 55-65	0.779	
Ages 65-97	0.665	Ages 65-97	0.725	
High school or less	0.740	High school or less	0.763	
Some college	0.774	Some college	0.826	
College degree	0.662	College degree	0.702	
Graduate studies	0.640	Graduate studies	0.696	

Table 20. MIDAS model RMSE scores for quarterly inflation expectations.

Interestingly, MIDAS estimation indicates comparable dispersion between the two types of media tones, and it suggests that news might be a more reliable predictor of future quarterly inflation expectations than the social media. Having revealed the connection between news coverage and inflation expectations, we will further investigate which news topics influence these expectations in the next section.

#### 5.5 Estimations using FinBERT media tones

Next, we present the estimation of the exposure of inflation expectations to different FinBERT media tones. The complete set of results are in Appendix B, Tables B1–B8. Nevertheless, Table 21 shows the exposure of different groups' inflation expectations to "Inflation & Fed" type of news in the newspapers, magazines, and web-based publications. Results suggest that the youngest and the oldest age groups are without a significant influence on their expectations. The highest exposures to "Inflation & Fed" news are among the 35-44, 45–54 and 55-65 years old. "Gas and Food Prices" has similar evidence on age cohorts' inflation expectations, see Table B1. The highest influence is for the cohort between 55 and 64 years of age. The limited and selected information about the headline inflation increases the size of the coefficients (3.1-3.5) compared to the sizes of constant in estimates of media tone exposure. We interpret that the information provided from the FinBERT tones is limited and hence, does not offer an accurate and comprehensive information about the expected inflation although it seems to be quite able to exposure on age group inflation expectations.

Table 22 presents the estimates of the influence of "Inflation & Fed" on different education groups inflation expectations. Now, we have an interesting result that households with least educated groups, "High school or less" and "Some college", are most exposed to "Inflation Fed" news. According to the Table B1, the late-middle age group (55–64) is most exposed to "Gas and Food Prices". Moreover, the "Gas and Food Prices" has the highest effect on least educated groups, the "High school or less" and "Some college" education group's inflation expectations, see Table B2. If the degree of education has any relationship with level of income one can assume that gas and food prices could play a significant role in their budget, and they closely follow the actual gas and food price dynamics. Hence, any information about the price hikes of "Gas and Food" may affect their inflation expectations.

Inflation expectations among different age groups are highly immune to inflation information stemming from "Mortgage & Business", see Table B3. The least educated group, "High school or less", is the only group whose inflation expectations are influenced by inflation related news about "Mortgage & Business"; see Table B4. It is possible that, as a group of lower incomes and financially restricted agents, they are highly exposed to any price changes in mortgages.

Inflation news related to "Healthcare Costs" has relation to older age cohorts' (45-54, 55-64, 65-97) inflation expectations (Table B5), whereas the two least educated groups ("High school", "Some college") are exposed to healthcare costs news (Table B6). This proposes that both the age and the level of education which typically relates to salary and wealth attract the attention about "Healthcare Costs".

Our FinBERT media tone also includes news regarding the "Education Costs". Results in Table B7 and B8 reveal that none of inflation expectations of different age cohorts or education group are related to "Education Costs" news.

We also conduct estimations using the FinBERT tone, which is constructed from sentences that include the word "inflation", referred to as "Inflation sentences" (see Tables 23 and 24). We observe the most significant negative

correlation in the age groups of 35-44, 45-54, and 65-97. Regarding education levels, the most substantial negative correlations are found for individuals with "High school or less" and "Some college" education.

		Dependent variable					
	18-34	35-44	45-54	55-64	65-97		
	(1)	(2)	(3)	(4)	(5)		
Inflation & Fed	-0.009 (0.008)	-0.020** (0.008)	-0.021** (0.009)	-0.019** (0.008)	-0.010 (0.008)		
Constant	3.125*** (0.272)	2.746*** (0.251)	2.882*** (0.283)	3.018*** (0.278)	3.371*** (0.265)		
Observations	79	79	79	79	79		
R <sup>2</sup>	0.016	0.083	0.076	0.065	0.020		
Adjusted R <sup>2</sup>	0.003	0.071	0.064	0.053	0.008		

Table 21. The role of Inflation & Fed FinBERT tone across age cohorts.

*Note:* Numbers in parenthesis are standard errors, \*p<0.1; \*\*p<0.05; \*\*\*p<0.01.

**Table 22.** The role of Inflation & Fed FinBERT tone across education groups.

	Dependent variable					
	High school or less	Some college	College degree	Graduate studies		
	(1)	(2)	(3)	(4)		
Inflation & Fed	-0.019**	-0.016*	-0.012	-0.011		
	(0.008)	(0.009)	(0.008)	(0.008)		
Constant	3.460***	3.193***	2.772***	2.630***		
	(0.266)	(0.291)	(0.260)	(0.266)		
Observations	79	79	79	79		
R <sup>2</sup>	0.072	0.041	0.027	0.025		
Adjusted R <sup>2</sup>	0.060	0.028	0.015	0.012		

		Dependent variable				
	18-34	35-44	45-54	55-64	65-97	
	(1)	(2)	(3)	(4)	(5)	
Inflation sentences	-0.026*	-0.030**	-0.035**	-0.028*	-0.033**	
	(0.015)	(0.015)	(0.016)	(0.016)	(0.015)	
Constant	3.197***	3.132***	3.270***	3.394***	3.421***	
	(0.151)	(0.144)	(0.161)	(0.158)	(0.146)	
Observations	79	79	79	79	79	
R <sup>2</sup>	0.037	0.051	0.057	0.039	0.059	
Adjusted R <sup>2</sup>	0.025	0.039	0.045	0.027	0.047	

Table 23. The role of sentence-based FinBERT tone across age cohorts.

	Dependent variable					
	High school or less	Some college	Some college College degree			
	(1)	(2)	(3)	(4)		
Inflation sentences	-0.050***	-0.033**	-0.017	-0.013		
	(0.015)	(0.017)	(0.015)	(0.015)		
Constant	3.660***	3.421***	2.993***	2.877***		
	(0.144)	(0.163)	(0.147)	(0.151)		
Observations	79	79	79	79		
R <sup>2</sup>	0.133	0.050	0.017	0.009		
Adjusted R <sup>2</sup>	0.122	0.038	0.004	-0.003		

#### 5.6 Estimation with Dynamic Model Averaging (DMA)

We apply the dynamic model averaging (DMA) approach to our FinBERT tones to identify which tones are important at different times for various inflation expectations. The forecast of the DMA model is given by the following equation:

(16) 
$$y_t^{DMA} = \sum_{k=1}^K \pi_{t|t-1,k} * y_{t,k}$$

In Equation 16 DMA we use posterior probabilities  $\pi_{t|t-1,k}$  to weight model forecasts  $y_{t,k}$  from K number of models to get DMA forecast  $y_t^{DMA}$  (Raftery, Kárny, Ettler, 2010). We can further decompose the posterior probabilities into the following equation:

(17) 
$$\pi_{t|t-1,k} = \frac{\pi^{\alpha}_{t-1|t-1,k} + c}{\sum_{l=1}^{K} \pi^{\alpha}_{t-1|t-1,l} + c}$$

In Equation 17  $\alpha$  is a forgetting factor between 0 and 1, which determines how the past observations are weighted. There is also another forgetting factor  $\lambda$  just below 1, as mentioned by Raftery, Kárny, and Ettler (2010). Moreover, from the DMA model, we can derive the relative variable importance (RVI) using the posterior probabilities (Drachal, 2020).

(18) 
$$RVI(x_i)_t = \sum_{k=1}^K \pi_{t-1|t-1,k} \mathbb{1}_k(x_i)$$

In Equation 18, the posterior probabilities of the models are summed, where  $\mathbb{1}_k(x_j) = 1$  if the k-th model contains  $x_j$  as an explanatory variable and  $\mathbb{1}_k(x_j) = 0$  if it does not (Drachal, 2020). From this, we can analyze the relevance of individual variables, which influence both the selection of models and their weights. This provides evidence of which models are important at various forecasting times.

In our setup, we utilize all of our FinBERT tones ("Inflation and Fed", "Gas & Food Prices", "Healthcare Costs", "Mortgages & Business", "Education Costs", and "Inflation sentences") in the DMA model to forecast inflation expectations. Similar to our previous MIDAS forecasting exercise, we have an initial sample of 8 quarters.

However, in our DMA exercise we set  $\alpha$  value to 0.97 and the  $\lambda$  value to 0.95. The initial variance is set to 5. In this way, we have 64 different models for each set of inflation expectations. Figures 18-21 present the posterior probabilities for the young and old age cohorts and the low and high education groups. The remaining results are in Appendix C, Figures C1-C5. Table 25 displays the DMA model's RMSE scores.

Inflation expectations	RMSE
Ages 18-34	0.793
Ages 35-44	0.779
Ages 45-54	0.871
Ages 55-65	0.854
Ages 65-97	0.748
High school or less	0.787
Some college	0.856
College degree	0.792
Graduate studies	0.821

Table 25. DMA model RMSE scores for quarterly inflation expectations.

As seen in Figures 18 and 19, the intercept is always included for both the young and old age cohorts. For the younger cohorts (ages 18-34), the news related to "Mortgages & Business" was of significant importance from 2002 to 2009. This holds somewhat true for the older age cohorts as well, as their model demonstrates significant probabilities for "Mortgages & Business". It is also worth noting that the younger generations show more significant probabilities for news about "Gas & Food Prices" and "Healthcare Costs" than the older age cohorts.

When comparing the inclusion probabilities between the highly educated and the less educated (Figures 20 and 21), we observe more volatility in the probabilities compared to those of the age cohorts. Specifically, the group with only a high school education shows the most volatile probabilities for "Mortgages & Business", "Healthcare Costs", "Education Costs", and "Inflation sentences". For the highly educated, the probabilities for "Mortgages & Business" are similar to those observed for the young age cohorts.

![](_page_47_Figure_0.jpeg)

Figure 18. DMA model posterior probabilities for 18-34 age cohort.

![](_page_48_Figure_0.jpeg)

Figure 19. DMA model posterior probabilities for 65-97 age cohort.

![](_page_49_Figure_0.jpeg)

Figure 20. DMA model posterior probabilities for "High school or less" education group.

![](_page_50_Figure_0.jpeg)

Figure 21. DMA model posterior probabilities for "Graduate studies" education group.

## 6. Conclusion

In recent times, inflation has surged to levels not seen in over four decades, placing the central bank's mandate of maintaining price stability at the forefront. Understanding and forecasting households' inflation expectations are crucial for policy effectiveness. However, the heterogeneity in these expectations poses additional challenges. This paper explores the influence of media tones, encompassing news and social media, on various demographic groups' inflation expectations in the United States, examining the effects across different age cohorts and educational levels.

Our analysis of US data suggests that traditional news has a more pronounced influence on the inflation expectations of older cohorts, while social media significantly shapes the expectations of younger and more educated groups. Additionally, we find that news is valuable in forecasting headline inflation. Contrary to Carroll's (2003) proposition, our findings indicate that the Survey of Professional Forecasters (SPF) may not serve as an adequate proxy for news, with social media potentially having a closer alignment with SPF expectations. Yet, the importance of SPF, especially among younger, highly educated individuals, should not be overlooked.

Our Local Projection estimates reveal that the effect of Media shock on inflation expectations persists for about five quarters. When accounting for the dynamics with lagged realized headline CPI inflation, the influence of media on expectations extends to approximately two quarters, highlighting the significant role of past inflation in shaping agents' expectations and the sticky nature of new information integration. Across groups, the dynamics of inflation information's influence on expectations show no significant differences.

To delve deeper into the drivers of inflation expectations, we utilize the FinBERT language model and the LDA topic model to categorize news into five distinct topics: "Inflation & Fed", "Gas and Food Prices", "Healthcare Costs", "Mortgages & Business", and "Education Cost". From these we form our own FinBERT media tones. The influence of these topics on inflation expectations varies, with less educated individuals being more sensitive to "Inflation & Fed", "Gas and Food Prices", "Healthcare Costs", and "Mortgages & Business". Among age cohorts, the oldest group is most affected by "Healthcare Costs", while "Inflation & Fed" resonates more with middle-aged individuals.

Our study sheds light for further research too. First, since our analysis focuses on the role of media tone for inflation expectations, a natural extension would be to integrate news into to a Phillips curve model (Bańbura et al., 2021; Ciccarelli and Osbat, 2017; Conti, 2021). Additionally, a large sample size would enable to incorporate state dependence in local projections and reveal for varying inflation environments. Furthermore, joint analysis of media shocks with de-anchoring measures of inflation expectations could also validate our results further.

In conclusion, our research establishes a strong link between media tone and inflation expectations, emphasizing its influence across different demographic segments. These insights suggest that central banks could benefit from adopting tailored communication strategies, particularly by leveraging the influence of social media, to enhance the effectiveness of their monetary policy. The heterogeneity observed in the drivers of these expectations opens up multiple avenues for future research, particularly the learning processes of various demographic groups.

# References

Araci, D. (2019). Finbert: Financial sentiment analysis with pre-trained language models. arXiv preprint arXiv:1908.10063. https://doi.org/10.48550/arXiv.1908.10063

Bańbura, Marta, Danilo Leiva-Leon, and Jan-Oliver Menz, "Do inflation expectations improve model-based inflation forecasts?," Working Paper Series 2604, European Central Bank October 2021.

Bernanke, B. (2007). *Inflation expectations and inflation forecasting* (Speech 306). Board of Governors of the Federal Reserve System (U.S.).

Bernanke, B. (2022).<sup>21</sup>st Century Monetary Policy: The Federal Reserve from the Great Inflation to COVID-19. W. W. Norton & Company.

Blanchard, O. J., & Quah, D. (1989). The Dynamic Effects of Aggregate Demand and Supply Disturbances. *American Economic Review*, 79(4), 655–673.

Blinder, A. S., & Krueger, A. B. (2004). *What Does the Public Know about Economic Policy, and How Does It Know It?* (Working Paper 10787). National Bureau of Economic Research. https://doi.org/10.3386/w10787

Carroll, C. D. (2003). Macroeconomic Expectations of Households and Professional Forecasters. *The Quarterly Journal of Economics*, 118(1), 269–298.

Cavallo, A., Cruces, G., & Perez-Truglia, R. (2017). Inflation Expectations, Learning, and Supermarket Prices: Evidence from Survey Experiments. *American Economic Journal: Macroeconomics*, 9(3), 1–35. https://doi.org/10.1257/mac.20150147

Ciccarelli, Matteo and Chiara Osbat, "Low inflation in the euro area: Causes and consequences," Occasional Paper Series 181, European Central Bank January 2017.

Conti, Antonio M., "Resurrecting the Phillips Curve in Low-Inflation Times," *Economic Modelling*, 2021, 96, 172–195.

Coibion, O., & Gorodnichenko, Y. (2015). Information Rigidity and the Expectations Formation Process: A Simple Framework and New Facts. *American Economic Review*, *105*(8), 2644–2678. https://doi.org/10.1257/aer.20110306

Coibion, O., Gorodnichenko, Y., & Weber, M. (2022). Monetary Policy Communications and Their Effects on Household Inflation Expectations. *Journal of Political Economy*, *130*(6), 1537–1584. https://doi.org/10.1086/718982

D'Acunto, F., Hoang, D., Paloviita, M., & Weber, M. (2019). Cognitive Abilities and Inflation Expectations. *AEA Papers and Proceedings*, *109*, 562–566. https://doi.org/10.1257/pandp.20191050

D'Acunto, F., Hoang, D., Paloviita, M., & Weber, M. (2022). IQ, Expectations, and Choice. *The Review of Economic Studies*, rdac075. https://doi.org/10.1093/restud/rdac075

D'Acunto, F., Malmendier, U., & Weber, M. (2021). Gender roles produce divergent economic expectations. *Proceedings of the National Academy of Sciences, 118*(21), 2008534118–2008534118.

D'Acunto, F., Malmendier, U., & Weber, M. (2023). What do the data tell us about inflation expectations? In *Handbook of economic expectations* (pp. 133–161). Elsevier.

Das, S., Kuhnen, C. M., & Nagel, S. (2020). Socioeconomic Status and Macroeconomic Expectations. *The Review of Financial Studies*, 33(1), 395–432. https://doi.org/10.1093/rfs/hhz041

Drachal, K. (2020). Dynamic Model Averaging in economics and finance with fDMA: A package for R. Signals, 1(1), 4.

Dräger, L., & Nghiem, G. (2021). Are Consumers' Spending Decisions in Line with A Euler Equation? *The Review of Economics and Statistics*, 103(3), 580–596. https://doi.org/10.1162/rest\_a\_00909

Evans, G. W., & Honkapohja, S. (2001). *Learning and Expectations in Macroeconomics*. Princeton University Press. https://www.jstor.org/stable/j.ctt7s6t9

Ghysels, E., Sinko, A., & Valkanov, R. (2007). MIDAS regressions: Further results and new directions. Econometric reviews, 26(1), 53-90.

Gorodnichenko, Y., Pham, T., & Talavera, O. (2023). The voice of monetary policy. American Economic Review, 113(2), 548-584.

Haavio, M., Heikkinen, J., Jalasjoki, P., Kilponen, J., & Paloviita, M. (2024). Reading Between the Lines-Uncovering Asymmetry in the Central Bank Loss Function. Bank of Finland Research Discussion Paper, (6).

Jordà, Ò. (2005). Estimation and Inference of Impulse Responses by Local Projections. *American Economic Review*, 95(1), 161–182. https://doi.org/10.1257/0002828053828518

Lamla, M. J., & Lein, S. M. (2014). The role of media for consumers' inflation expectation formation. *Journal of Economic Behavior & Organization*, 106, 62–77. https://doi.org/10.1016/j.jebo.2014.05.004

Lamla, M. J., & Maag, T. (2012). The Role of Media for Inflation Forecast Disagreement of Households and Professional Forecasters. *Journal of Money, Credit and Banking*, 44(7), 1325–1350. https://doi.org/10.1111/j.1538-4616.2012.00534.x

Larsen, V. H., Thorsrud, L. A., & Zhulanova, J. (2021). News-driven inflation expectations and information rigidities. *Journal of Monetary Economics*, 117, 507–520. https://doi.org/10.1016/j.jmoneco.2020.03.004

Malmendier, U., & Nagel, S. (2016). Learning from Inflation Experiences \*. The Quarterly Journal of Economics, 131(1), 53-87. https://doi.org/10.1093/qje/qjv037

Mankiw, N. G., & Reis, R. (2002). Sticky Information versus Sticky Prices: A Proposal to Replace the New Keynesian Phillips Curve. *The Quarterly Journal of Economics*, 117(4), 1295–1328.

Mankiw, N. G., Reis, R., & Wolfers, J. (2003). Disagreement about Inflation Expectations. *NBER Macroeconomics Annual*, 18, 209–248. https://doi.org/10.1086/ma.18.3585256

Malo, P., Sinha, A., Korhonen, P., Wallenius, J., & Takala, P. (2014). Good debt or bad debt: Detecting semantic orientations in economic texts. Journal of the Association for Information Science and Technology, 65(4), 782-796.

Raftery, A. E., Kárný, M., & Ettler, P. (2010). Online prediction under model uncertainty via dynamic model averaging: Application to a cold rolling mill. Technometrics, 52(1), 52-66.

Ramey, V. A., & Zubairy, S. (2018). Government Spending Multipliers in Good Times and in Bad: Evidence from US Historical Data. *Journal of Political Economy*, *126*(2), 850–901. https://doi.org/10.1086/696277

Sims, C., A. (2003). Implications of Rational Inattention." Journal of Monetary Economics 50 (3): 665–90. https://doi.org/10.1016/S0304-3932(03)00029-1

Weber, M., D'Acunto, F., Gorodnichenko, Y., & Coibion, O. (2022). The Subjective Inflation Expectations of Households and Firms: Measurement, Determinants, and Implications. *Journal of Economic Perspectives*, *36*(3), 157–184. https://doi.org/10.1257/jep.36.3.157

Weber, M., Cansia, B., Ropele, T., Lluberas, R., Frache, S., Meyer, B.,H., Kumar, S., Korodnichenko, Y., Georgaragoros, D., Coibion., O., Kenny, G., Ponce, J. (2023). Tell me something I don't know: learning in low and high inflation settings. NBER Working Paper No 31485.

Woodford, M. (2003). Imperfect Common Knowledge and the Effects of Monetary Policy. In *Knowledge, Information, and Expectations in Modern Macroeconomics: In Honor of Edmund S. Phelps*, edited by Philippe Aghion, Roman Frydman, Joseph Stiglitz, and Michael Woodford, 25–58. Princeton: Princeton University Press.

# Appendix A

# The following texts are excerpted from the Thomson Reuters MarketPsych Indices (TRMI) handbook.

Thomson Reuters MarketPsych Indices (TRMI)

# TRMI Variables

Thomson Reuters MarketPsych Indices (TRMI) are provided by Thomson Reuters. TRMI data are based on word recognition techniques which aim to extract economics, finance, business and psychological relevant information from financial news, social media, conference calls transcripts and executive interviews. Data are available for more than 7,750 companies from over 30 countries and index level data are available to the 15 largest markets from 1998 onwards. Data are also available for 130 countries. Only text written in English is used in the analysis.

The TRMI database uses news from several of the mainstream news sources for the entire historical sample period. After 2005, also internet news content is used for the analysis. Internet news are restricted to top international, regional and business news sources, and leading industry sources. The social media data is less diverse. At the beginning, the sample consists of content collected from discussions in internet forums and after 2008, top 30% of blogs, microblogs, and other social media content (e.g., finance-specific tweets) are also included to the data. The popularity of these sources is measured with incoming links.

The value of the TRMI variable can be non-integer when any of the words/phrases used to form it are described with a "minimizer" ("maximizer"), which reduces (increases) the intensity of the primary word or phrase. For example, in the phrase "less concerned" the score of the word "concerned" is minimized by "less". As a result, the scores of common words/phrases with minor contributions can be minimized while more intense emotions are weighted more heavily.

## Data analysis

Previous studies mostly use a lexical analysis method where an article text is compared to pre-specified context relevant lexicon and their frequencies are subsequently counted. The most significant limitation of this method is that it concentrates only on one dimension of sentiment (positive vs. negative) and ignores other, partly interrelated, dimensions. Also, some open-source sentiment dictionaries might misclassify business-and finance-related terms. TRMI word recognition technique uses extensive customization and curation of lexicons and aims to identify and score hundreds of sentiment dimensions using different grammatical frameworks to different text sources based on their special traits and also considers the structure of the sentences and articles. When applied to text, the confluence of the various text processing described below generates over 400 psychological variables (PsychVars), each with the potential to be applied to a different entity. It should be emphasized that the method aims to identify word interrelationships and calculates the score based these, instead of using frequency analysis on individual words. The following goes through the methods used for word recognition and content quantifying method used in the data formation.

## Source type differences

TRMI data are extracted from both traditional news media and from social media sources which creates challenging environment for word recognition techniques. Thus, the data analysis is customized for each source due to differences in communication styles between social and news media.

Unlike news media, which uses standardized terms and appropriate tone in its contents, social media contains substantial amounts of sarcasm and irony, colloquial meanings, incomplete thoughts, misplaced punctuation, misspellings, non-standard grammar, case insensitivity and crude language. The expressions of emotions between news and social media also vary. In social media, authors tend to use nationally and regionally developed emoticons (e.g. ":)") and acronyms (e.g. "LOL") to express themselves.

In addition, unlike journalists of traditional news media, who are trained to offer multiple perspectives on the underlying story and go through editorial process, authors in social media have a tendency to express their own opinions and emotional states more directly. In the news, the role of journalists is to describe the emotional states of those they are reporting on. As a result, information obtained from social media is typically less inclusive of contrary viewpoints and more emotionally expressive from the first-person perspective than news information.

As a result of all these differences, text analytical models are used to calibrate the text analytics by source type and different models are used for news, social media forums, tweets, SEC filings and earnings conference call transcripts.

# Entity Identification

Entities are identified with a list containing more than 60 000 entity names and their aliases which include e.g., different language specific writing forms of locations and firms.

TRMI uses anti-correlate filters to exclude irrelevant entities. For example, gold and silver are typically mentioned every two years during the Olympic Games but they are not related to gold and silver commodities. Similarly, references to South Korean Won could be either used to refer to victories achieved by South Korean athletes or to South Korea's currency. Also, the entities should have a correct co-references. For example, "breakfast oats" are not counted as a correct reference to commodity Oats unless it is accompanied with key identification correlates such as "prices" or "futures".

# Timing

TRMI has several variables referring to expectations. The text-analytics software is calibrated to identify verb tenses in every phrase and is able to recognize when the references are future-oriented and hence related to expectations. For example, "Optimism" is a difference between references to future-oriented positive and negative comments and for example to "Uncertainty", any references to past uncertainties are excluded from the analysis.

# Modifiers and Negations

Modifier words change the meaning or tone of a phrase or sentence by modifying its impact. For example, words or phrases that increase the significance of an adjective, such as "large", (e.g. "large loss") are multiplicative on the weighting of the modified word whereas minimizers like "a little" and "a handful" multiply the meaning score by 0.5. Modifiers can either increase (maximizers) or decrease (minimizers) the score of a key term and thus affect the scoring of textual meaning. In cases of frequent meanings based on simple word relationships, like the word "new", a multiplier is used to minimize the weight of the meaning. For example, in the case of "new" the multiplier is 0.1 when that meaning is used in the Innovation index.

Data also takes into account the differences between news headlines and article bodies. Headlines have a multiplicative weight of 3, subtitles a weight of 2 and bodies a weight of 1 in the data formation.

In addition to the maximizing and minimizing operations, the data formation takes into account negations. For example, the phrase "I'm not worried about the earnings release" connotes that the author is not afraid, and as a result, the extracted fear score is negated (-1).

## Source Texts

The TRMI measures are from two groups of sources, news and social media, and the data consists of three feeds: news feed, social media feed and an aggregate feed of combined news and social media content. TRMI processes more than two million articles daily and the data are updated minutely. Each minutely value is an average of the past 24 hours of observations about the target entity.

As sources, TRMI uses data from the largest and most important news organizations, global internet news coverage, and a broad and credible range of social media sources. The TRMI news indices are derived from content delivered in Thomson Reuters News Feed Direct and two Thomson Reuters news archives: a Reuters-only one from 1998 to 2002 and one with Reuters and select third-party wires from 2003 onwards. In 2005, Moreover Technologies aggregate news feed was incorporated to the data adding 50 000 internet news sites to the sources. In addition, hundreds of financial news sites and finance specific sources widely read by professional investors are

also included in the data. MarketPsych social media content has been downloaded from public social media sites from 1998 onwards and the data from Moreover Technologies' aggregate social media feed, which is derived from more than four million social media sites, was incorporated to TRMI in 2009.

#### Quantifying the Articles

Each TRMI variable constitutes of a combination of minor components called PsychVars. To form a TRMI variable, the absolute values of PsychVars are first determined using 24 hours of observations. These absolute values are then summed for all constituents to get the Buzz-variable. More specifically, Buzz is calculated as follows:

$$Buzz_t(a) = \sum_{c \in C(a), p \in P} |PsychVar_{c,p,t}|$$

where t denotes time, a is the entity, C(a) is the set of all constituents of a (for example, for indices, C(a) consists of all individual assets comprising the index) and P(s) is the set of all PsychVars relevant to a particular TRMI variable s.

For the other TRMI variables, the values are calculated as follows:

$$TRMI_{s,t}(a) = \frac{\sum_{c \in C(A), p \in P(s)} (I_t(s, p) * PsychVar_{p,t}(c))}{Buzz_t(a)}$$

where  $I_t(s, p)$  defines whether a PsychVar  $p \in P(s)$  is additive or subtractive to a TRMI variable s at time t:

$$I_t(s,p) = \begin{cases} +1, if \ additive \ at \ time \ t \\ -1, if \ subtractive \ at \ time \ t \end{cases}$$

A single PsychVar can contribute to multiple TRMI variables. For example, EarningsUp\_f PsychVar is a constitute of all, EarningsForecast, Sentiment, Optimism and Fundamental Strength TRMI variables.

# Appendix B

	Dependent variable					
	18-34	35-44	45-54	55-64	65-97	
	(1)	(2)	(3)	(4)	(5)	
Gas & Food Prices	-0.012* (0.006)	-0.011* (0.006)	-0.015** (0.007)	-0.019*** (0.007)	-0.009 (0.006)	
Constant	3.127*** (0.176)	3.107*** (0.169)	3.201*** (0.188)	3.159*** (0.179)	3.462*** (0.174)	
Observations	79	79	79	79	79	
R <sup>2</sup>	0.043	0.041	0.057	0.103	0.028	
Adjusted R <sup>2</sup>	0.030	0.029	0.045	0.091	0.015	

Table B1. The role of Gas & Food Prices FinBERT tone across age cohorts.

*Note:* Numbers in parenthesis are standard errors, \*p<0.1; \*\*p<0.05; \*\*\*p<0.01.

Table B2. The role of Gas & Food Prices FinBERT tone across education groups.

	Dependent variable					
	High school or less	Some college	College degree	Graduate studies		
	(1)	(2)	(3)	(4)		
Gas & Food Prices	-0.012*	-0.019***	-0.014**	-0.012*		
	(0.006)	(0.007)	(0.006)	(0.006)		
Constant	3.776***	3.245***	2.790***	2.705***		
	(0.177)	(0.187)	(0.167)	(0.173)		
Observations	79	79	79	79		
R <sup>2</sup>	0.046	0.088	0.066	0.041		
Adjusted R <sup>2</sup>	0.033	0.076	0.054	0.029		

	Dependent variable				
	18-34	35-44	45-54	55-64	65-97
	(1)	(2)	(3)	(4)	(5)
Mortgages & Business	-0.008	-0.019	-0.024*	-0.021	-0.018
	(0.013)	(0.012)	(0.013)	(0.013)	(0.012)
Constant	3.303***	3.101***	3.205***	3.313***	3.433***
	(0.202)	(0.191)	(0.213)	(0.209)	(0.196)
Observations	79	79	79	79	79
R <sup>2</sup>	0.005	0.032	0.042	0.035	0.027
Adjusted R <sup>2</sup>	-0.008	0.019	0.029	0.022	0.014

Table B3. The role of Mortgages & Business FinBERT tone across age cohorts.

Table B4. The role of Mortgages & Business FinBERT tone across education groups
---

	Dependent variable					
	High school or less	Some college	College degree	Graduate studies		
	(1)	(2)	(3)	(4)		
Mortgages & Business	-0.031**	-0.019	-0.012	-0.003		
	(0.012)	(0.013)	(0.012)	(0.012)		
Constant	3.624***	3.410***	2.957***	2.945***		
	(0.196)	(0.217)	(0.194)	(0.199)		
Observations	79	79	79	79		
R <sup>2</sup>	0.077	0.026	0.013	0.001		
Adjusted R <sup>2</sup>	0.065	0.014	0.0004	-0.012		

	Dependent variable						
	18-34	35-44	45-54	55-64	65-97		
	(1)	(2)	(3)	(4)	(5)		
Healthcare Costs	-0.019	-0.026	-0.038**	-0.037**	-0.042**		
	(0.018)	(0.017)	(0.019)	(0.018)	(0.017)		
Constant	3.258***	3.161***	3.237***	3.312***	3.331***		
	(0.174)	(0.166)	(0.184)	(0.179)	(0.165)		
	70	70	70	70	70		
Observations	/9	/9	/9	/9	/9		
$\mathbb{R}^2$	0.014	0.028	0.050	0.050	0.074		
Adjusted R <sup>2</sup>	0.001	0.016	0.038	0.038	0.062		

Table B5. The role of Healthcare Costs FinBERT tone across age cohorts.

Table B6. The role of Healthcare Costs FinBERT tone across education groups.

	Dependent variable					
	High school or less	Some college	College degree	Graduate studies		
	(1)	(2)	(3)	(4)		
Healthcare Costs	-0.044**	-0.045**	-0.019	-0.028		
	(0.018)	(0.019)	(0.017)	(0.017)		
Constant	3.700***	3.312***	2.971***	2.747***		
	(0.169)	(0.184)	(0.167)	(0.169)		
Observations	79	79	79	79		
R <sup>2</sup>	0.077	0.068	0.016	0.032		
Adjusted R <sup>2</sup>	0.065	0.056	0.003	0.020		

	Dependent variable					
	18-34	35-44	45-54	55-64	65-97	
	(1)	(2)	(3)	(4)	(5)	
Education Costs	0.009	-0.004	-0.015	-0.020	-0.009	
	(0.013)	(0.013)	(0.014)	(0.014)	(0.013)	
Constant	3.553***	3.314***	3.317***	3.314***	3.547***	
	(0.226)	(0.217)	(0.242)	(0.234)	(0.221)	
Observations	79	79	79	79	79	
R <sup>2</sup>	0.005	0.001	0.015	0.027	0.006	
Adjusted R <sup>2</sup>	-0.007	-0.012	0.002	0.014	-0.007	

Table B7. The role of Education Costs FinBERT tone across age cohorts.

Table B8. The role of Education Costs FinBERT tone across education groups.

	Dependent variable					
	High school or less	Some college	College degree	Graduate studies		
	(1)	(2)	(3)	(4)		
Education Costs	-0.016	-0.017	-0.009	-0.004		
	(0.013)	(0.014)	(0.013)	(0.013)		
Constant	3.822***	3.427***	2.991***	2.924***		
	(0.226)	(0.244)	(0.217)	(0.222)		
Observations	79	79	79	79		
R <sup>2</sup>	0.019	0.018	0.007	0.001		
Adjusted R <sup>2</sup>	0.006	0.005	-0.006	-0.012		

# Appendix C

![](_page_62_Figure_1.jpeg)

Figure C1. DMA models posterior probabilities for 35-44 age cohort.

![](_page_63_Figure_0.jpeg)

Figure C2. DMA models posterior probabilities for 45-54 age cohort.

![](_page_64_Figure_0.jpeg)

Figure C3. DMA models posterior probabilities for 55-64 age cohort.

![](_page_65_Figure_0.jpeg)

Figure C4. DMA models posterior probabilities for "Some college" education group.

![](_page_66_Figure_0.jpeg)

Figure C5. DMA models posterior probabilities for "College degree" education group.

# Bank of Finland Research Discussion Papers 2024

ISSN 1456-6184, online

1/2024	Gene Ambrocio – Iftekhar Hasan – Xiang Li Global political ties and the global financial cycle
2/2024	Michaela Elfsbacka-Schmöller – Nigel McClung Price stability and debt sustainability under endogenous trend growth
3/2024	Diego Moreno – Tuomas Takalo Stress test precision and bank competition
4/2024	Eleonora Granziera – Pirkka Jalasjoki – Maritta Paloviita The bias of the ECB inflation projections: a state-dependent analysis
5/2024	Maria Eskelinen – Christopher G. Gibbs – Nigel McClung <b>Resolving new Keynesian puzzles</b>
6/2024	Markus Haavio – Joni Heikkinen – Pirkka Jalasjoki – Juha Kilponen – Maritta Paloviita – Ilona Vänni Reading between the lines - Uncovering asymmetry in the central bank loss function
7/2024	Aleksi Paavola – Ville Voutilainen Central bank collateral policy and credit pricing: evidence from Finland
8/2024	Joni Heikkinen – Kari Heimonen Media tone: the role of news and social media on heterogeneous inflation expectations