THE ROLE OF THE MEDIA IN THE INFLATION EXPECTATION FORMATION PROCESS

TETIANA YUKHYMENKO[a]

[a]National Bank of Ukraine
E-mail: tetiana.yukhymenko@bank.gov.ua

Abstract This research highlights the role played by the media in the formation of inflation expectations among various respondents in Ukraine. Using a large news corpus and machine-learning techniques, I have constructed news-based metrics that produce quantitative indicators for texts, which show if the news topics are relevant to inflation expectations. I have found evidence that various news topics may have an impact on inflation expectations, and can explain part of their variance. Thus, my results could help in the analysis of inflation expectations – which is of value, given that anchoring inflation expectations remains a key challenge for central banks.

JEL Codes C55, C82, D84, E31, E58

Keywords inflation expectations, natural language processing, textual data, machine learning

1. MOTIVATION, THEORETICAL FRAMEWORK, LITERATURE OVERVIEW

Anchoring inflation expectations remains a key challenge for central banks, especially in developing economies. The process of forming inflation expectations it relevant to understanding macroeconomic dynamics and for designing optimal policies. A lot of research has been done in the area of inflation expectations, but there is still a great deal of uncertainty and inconsistency about the factors that determine them. The development of modern information technologies enables us to use new approaches to examine the processes that form inflation expectations. In particular, natural language processing and machine-learning tools can provide additional information that was previously inaccessible. They also makes it possible to supplement with new insights the results of existing studies that have become benchmarks in the industry. Many researchers are now turning to more modern data sources and analysis methods, but the field of research remains largely uncharted. In particular, there is still much uncertainty over how to transform unstructured data into economic indicators, how to take into account the tone of indicators, and how to assess their impact on inflation expectations. In addition, such studies have not yet been conducted on Ukrainian data. Thus, the prospect of being able to apply the latest technologies to already conventional approaches was the main motivation for researching the role of the media in shaping inflation expectations in Ukraine.

The rational expectations hypothesis has dominated the macroeconomic literature for many years. However, in a growing body of research, this hypothesis is being modified to account for information rigidities – expectations could be rational, but in a more realistic environment agents may be inattentive to relevant information due to the costs of acquiring and processing such information. The two leading models of information rigidity are the sticky information model of Mankiw et al. (2004) and the noisy information model developed by Woodford (2004) and Sims (2009). Mackowiak and Wiederholt (2009) also did work in this field. Coibion and Gorodnichenko (2012) proved that information rigidities have a large impact on macroeconomic variables, thus they should be integrated into modern macroeconomic policies in order to execute the optimal monetary policy. They also found that despite common wisdom, there is no significant difference in the degree of information consumption across agents – the speed of information processing by consumers is no lower than that by other agents. Among other things, this can be explained by the noisy information model. Similarly, Coibion and Gorodnichenko (2015a) find that the inflation expectations of professional forecasters from the U.S. Survey can be modeled with imperfect information models due to the existence of information frictions. Coibion and Gorodnichenko (2015b) also research economic agents' expectations in Ukraine, based on survey data on inflation and exchange rate expectations. The survey also shows that there is a strong positive correlation between the evolution of Ukrainian economic agents' expectations about inflation, and exchange rates. While some correlation might be expected from the pass-through of exchange rates into prices, a more likely rationale is that the exchange rate is being used as a straightforward proxy by households of

Available at https://doi.org/10.26531/vnbu2022.253.01

broader price movements within the economy, very much like households within the U.S. do with gasoline prices.

It can be assumed that survey respondents are also influenced by uncertainties regarding tax, tariff, spending, monetary and regulatory policy. These effects, however, are hard to detect because uncertainty is unobservable. However, people may obtain their views about the future path of the economy from the news media, directly or indirectly. So, news-based methods could be used to investigate the impact of the media environment on the formation of respondents' expectations.

For example, Carroll (2003) tested an epidemiological model of expectations in which information diffuses over time from professional forecasters to households. Pfajfar and Santoro (2013) complement this model with a measure of the actual perception of new information about prices. As a news metric, they used a question from a survey where participants have to indicate whether they have heard about positive or negative changes. Hearing news related to prices increases the probability of an adjustment in inflation expectations, while the quality of forecasts is not likely to improve. Similarly, Coibion et al (2019) researched how central bank communications impact expectations. Thus, they compare the answers of respondents after receiving eight different forms of information regarding inflation. They concluded that these messages to the public influence expectations by economically significant magnitudes. However, their effectiveness significantly decreases when channeled via news media. Mazumder (2021) proved that newspaper mentions of the Fed bring consumer and professional inflation forecasts closer, although this effect can vary depending on which newspaper it was published in and how the topic was covered by the author. Dräger and Lamla (2017) also found evidence of the impact of the media on the formation of inflation expectations. They analyzed the rotating panel dimension of the microdata in the University of Michigan Survey of Consumers, and found evidence that respondents are more likely to adjust their expectations if they have heard news about inflation.

However, most of these studies imply the use of supplementary questions in the survey, which can be costly. In addition, even if such questions are introduced, the results will not cover previous periods. Thus, measuring the impact of news and constructing relevant indexes requires novel sources of information and processing methods, as well as significant computational resources. Consequently, researchers are replacing these indexes with alternative indicators which could be related to news metrics. For example, Bauer (2015) used macroeconomic data surprises cumulated over the monthly or quarterly observation windows as an economic news metric. Thus, the data are macroeconomic indicators collected from traditional statistical sources, but their interpretation is somewhat different from the usual time series. Bauer found that several different survey measures of inflation expectations respond significantly to macroeconomic surprises. He also concluded that better anchoring of long-term inflation expectations can reduce the sensitivity of inflation expectations to macroeconomic news, and the variability of nominal rates as well. Garcia and Werner (2018) confirmed that early inflation releases had a significant impact on long-term inflation expectations, and that there was a weakening of the anchoring of inflation expectations in the EU in recent years. Nautz et al (2017) also found that euro area inflation expectation anchoring was undermined after the fall of 2011. They discovered that long-term inflation expectations respond significantly to macroeconomic news. As a news metric, a set of macroeconomic variables was used, including CPI, PPI, unemployment, GDP, trade balance, etc. D'Acunto et al. (2017) additionally found a relationship between the frequency and size of price changes.

Larsen et al. (2021) used a more sophisticated approach. They applied machine-learning algorithms to a large news corpus and examined the role of the media in the expectation formation process of households. It turned out that the news topics in the media are a good predictor of both inflation and inflation expectations. They also found that the degree of information rigidity among households varies across time, which can be explained by the relevant media coverage. Angelico et al. (2021) used a similar approach to build real-time measures of consumers' inflation expectations from tweets. They combined unsupervised machine-learning techniques with a dictionary-based approach to construct indices. Twitter-based indicators appear to be highly correlated with traditional measures of inflation expectations while having an advantage in their speed.

In this work, I focus on the analysis of news and its impact on the formation of inflation expectations. To this end, I explore approaches to transforming texts into quantitative indicators which can then be further used in traditional econometric analyses. These indicators should reflect news topics relevant to inflation expectations and accurately capture their intensity. These metrics should also be easily interpretable, as they aim to explain the impact of news on the formation of inflation expectations. All these tasks can be accomplished with text-mining techniques.

All measurement methods that are based on text-mining can be divided into two groups: 1) so-called naïve methods and 2) more complex methods, which are based on machine learning. Naïve methods are parsimonious, easy to use (as they do not require much computational power), and are recognized worldwide due to their simplicity. They are based mostly on term frequency and document frequency. For example, Baker et al. (2016) investigated the relationship between economic policy uncertainty and rates of investment, output, and employment growth. For these purposes, the authors developed an index of economic policy uncertainty (EPU) based on a monthly count of articles that contain specific terms. Their findings demonstrate that the EPU is a reasonable proxy for various types of important macroeconomic variables, and its results are consistent with theories that highlight the negative economic effects of uncertainty shocks.

However, these approaches have the potential to underestimate the actual level of uncertainty, as they require qualitative expertise and human resources. For example, most naïve methods involve dictionary construction. This problem can be resolved with more complex methods based on machine-learning techniques. Despite their relative complexity, machine-learning approaches have greater predictive power than the naïve methods, empirical results show. The fastest and easiest method is to use an unsupervised machine-learning technique.

One of the most popular unsupervised natural language processing tools is the Latent Dirichlet Allocation (LDA), introduced by Blei et al. (2003). This generative statistical model divides a collection of texts into subgroups, with each subgroup being characterized by keywords associated with a topic. This method estimates the likelihood of the probability of the occurrence of words for a different number of topics. The results indicate the most likely number of topics. LDA is an unsupervised machine-learning technique, which does not require a training dataset. However, the model's results are unpredictable and require careful analysis. But while the methodology has been applied heavily in the machine-learning literature and for textual analysis, surprisingly, in economics it has so far only seen a small number of successful applications, e.g., Larsen et al. (2021) or Azqueta-Gavaldon (2017). Tobback et al. (2016) chose a different route and tried to improve the first EPU index designed by Baker through applying supervised machine learning. Thus, they developed a classification model based on support vector machines (SVMs) and labeled articles into two classes – related to economic policy uncertainty or not. Further, they constructed an EPU SVM indicator based on this classification, and include it in different macroeconomic models. This helps improve the accuracy of these models' economic variable forecasts in the short term.

However, unsupervised models, like naïve methods, also have disadvantages – the absence of sentiment analysis. This could be mitigated by machine-learning and lexicon-based techniques that use a predefined vocabulary and assess the relative frequency of sentiment words in the text. For example, Taboada et al. (2011) presented Semantic Orientation CALculator (SO-CAL). This model uses dictionaries of words annotated with their semantic polarity and strength featuring intensification and negation. SO-CAL can be used on completely unseen data. VADER (Valence Aware Dictionary for sEntiment Reasoning) is another successful example of a lexicon-based sentiment analysis tool. To develop it, Hutto and Gilbert (2014) constructed a list of lexical features and combined them with general rules that embody grammatical and syntactical conventions for expressing sentiment intensity. VADER outperformed many other highly regarded sentiment analysis tools. However, the main downside of lexicon-based techniques is the lack of trained dictionaries in languages other than English.

The introduction of a new language representation model called BERT (Bidirectional Encoder Representations from Transformers), developed by Google researchers (Devlin et al., 2018) was a significant breakthrough in sentiment analysis. Like many other recent works in pre-training contextual representations, BERT makes use of an attention mechanism that learns contextual relations between words (or sub-words) in a text. But unlike many other models, it is designed to pre-train deep bidirectional representations from the unlabeled text (treating on both left and right context). As result, BERT can distinguish differences in the usage of even the same word, taking into account the context of the occurrence of a given word. A pre-trained BERT model can be fine-tuned for a wide range of tasks, including classification. Pre-trained versions of BERT are available in a wide range of languages (including Ukrainian and russian).

To narrow down the scope of this paper, I focus on the simpler naive methods and unsupervised machine learning, and leave the sentiment analyses for future research. Starting with the simplest naïve methods, I will continue with more complex machine-learning methods of text classifications such as LDA. Thereafter, I develop an econometric model to assess the impact of the constructed indices on the formation of inflation expectations.

The paper is organized as follows: The next section presents data characteristics divided into two parts – a text corpus of economic news and inflation expectations in Ukraine. Section 3 describes the construction and results of news-based indices and presents their statistical properties. Section 4 analyzes the empirical specifications of the models and describes the results. Finally, Section 5 offers some concluding remarks and future steps. Additional information and results can be found in the Appendices.

2. DATA CHARACTERISTICS

2.1. News Corpus

The general key criteria in the selection of news sources were the availability of a fairly long archive (at least for the last ten years) and the possibility to scrape data from the web, which significantly limited the available list. Also, the newspapers used should have mainly economic orientation and not be subject to the explicit influence of individual political forces. I was guided by a list of the most popular resources, from which I excluded those that did not meet the specified requirements. In particular, I selected Ukrainian Pravda, Liga, and UNIAN, which are included in the list of TOP-50 Ukrainian online media based on Gemius and TNS ratings (at numbers 6, 11, and 20 respectively). In addition, I included data from the site Finance.ua, which is narrowly specialized in economic topics, and was also among the top 200 most popular sites in Ukraine (according to alexa.com). According to Similarweb.com, these sites in total covered about 8.5% of the traffic of online news and media in Ukraine in August-October 2022. However, these news sites do not provide information on the volume of views or coverage of each news item. This information could be useful in determining the strength of the news impact in the form of a multiplier.

The official language in Ukraine is Ukrainian, but the russian language was very popular before before russia's full-scale invasion, so most national media published materials in two languages. At the same time, the natural language processing infrastructure for the russian language was slightly better developed at the time the research was conducted, containing richer libraries, and modules that could improve the results of the research. For this reason, it was decided to process the news articles in russian.

The web-scraped news corpus consists of more than 2 million articles published online covering a sample of 20 years from January 2000 to December 2020. Our dataset contains the full text of the article and the available metadata, which include, for instance, the date, the link, and the title. Some sources also have a subtitle or general topic. All work with textual data, starting from the web-scraping and preprocessing and ending with the index construction was carried out using the Python programming language.

The textual data is presented in a non-traditional format, which makes statistical inference challenging. Thus, it is essential to preprocess the corpus data into a machine-readable format. Preprocessing includes some steps to clean and reduce the raw dataset before analysis.

First, it is essential to lowercase all of the characters to avoid any case-sensitive processing. This should help to clean the dataset at least in two cases:

– words with the uppercase letters may not be detected as a stopword, as all the stopword lists are lowercase. Stopwords are words that have no significant contribution to the meaning of the text. For example, the most common stopwords are conjunctions and prepositions;

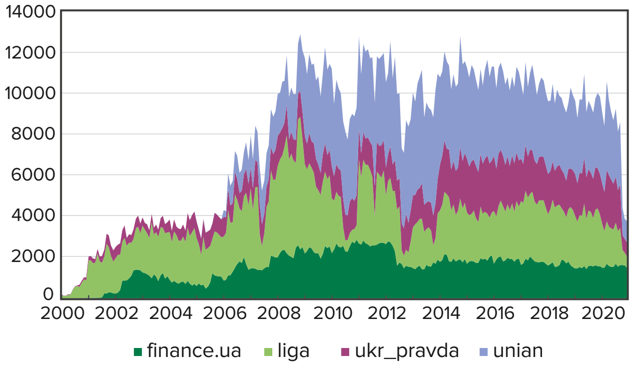
– for grammatical reasons, the same word can be treated differently due to its position in the sentence. For example, the first word of the sentence is always uppercased, ven if it is not a proper name. As a result, the same word could be given two different values.

Second, but fundamental, step in NLP methods is tokenization. Tokenization is a method of breaking a piece of raw text into smaller units called tokens and converting them into a list. Tokens can be words, characters, or subwords. A token is a unit that NLP tools can easily convert to a value suitable for further machine processing.

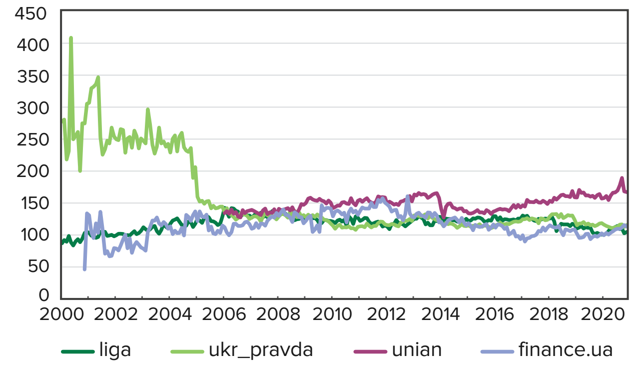
Third, I removed non-essential information like stopwords from the text to simplify data processing. The NLTK library in Python has rich corpora of stopwords in different languages, including russian. Additionally, I removed non-ASCII characters, links, and punctuations by using Regular Expressions (Regex).

The fourth step of text preprocessing is text normalization. The most common normalization techniques for Natural Language Processing are stemming and lemmatization. Stemming is a technique that chops off the ends of words. Due to this approach, words with the same meaning but having some variations according to the context or sentence are normalized. Lemmatization usually refers to a morphological analysis of words, and normally aims to return the base or dictionary form of a word (lemma). russian, as well as Ukrainian, is a morphologically rich language, characterized by free word order and various word forms. Almost all language parts are marked for many characteristics as number, gender, case, tense, aspect, or person, which agreed grammatically with each other (Rozovskaya and Roth, 2019). Therefore, despite the longer processing time and the need for greater computing capacity, lemmatization is preferable to stemming for russian text normalization. For this purpose, I used the morphological analyzer pymorphy2, which returns the dictionary form of a word (Korobov, 2015).

Finally, I received a cleaned and normalized textual dataset consisting of around 300 million words and nearly 800,000 unique tokens. On average, as can be seen in Figure 1, the article size did not change significantly during the observed period (100-140 words per article). However, the number of articles grew considerably from a few articles per month in the early 2000s to between 8,000 and 14,000 per month from 2008, while the distribution of article quantity between online sources also changed (for more details about the news corpus see Appendix C). The increase in the number of online articles in the mid-2000s is related to the rapid growth of internet penetration in Ukraine. Thus, according to the State statistics service of Ukraine, the number of active internet users in Ukraine exceeded 1 million people for the first time in 2007, having gradually increased from 200,000 in 2000. Since that time, the number of active users has skyrocketed to more than 23 million, and accordingly, internet penetration rose to 56% in 2019. This forced news media to move from traditional paper forms to online versions. As a result, the media not only fully transferred their articles online but also expanded the content of websites with additional materials that are not usually placed in newspapers.



(a)



(b)

Figure 1. Number of News (a) and the Average Size of Articles (b) per Month

2.2. Inflation Expectations

The National Bank of Ukraine has been running surveys of inflation expectations for the next 12 months for several types of agents: households, banks, businesses, and professional forecasters. Before the adoption of the inflation targeting regime by the NBU in 2015, Coibion and Gorodnichenko (2015b) widely reviewed these surveys and discussed their limitations. In my paper, I will briefly describe the characteristics of the inflation expectations of all groups of respondents (Figure 2).

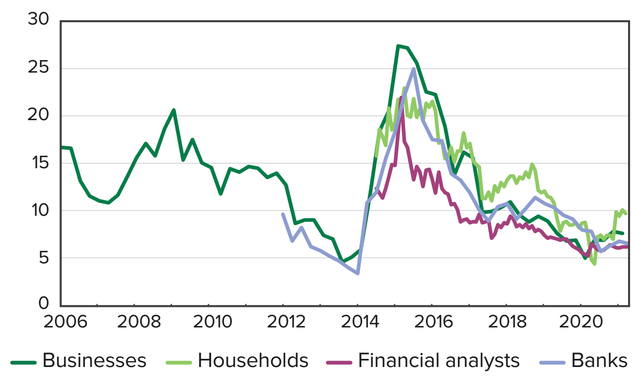


Figure 2. Inflation Expectations for the Next 12 Months, %

Source: NBU, GfK Ukraine, Info Sapiens.

Banks. The survey of banks covers at least 90% of the banking system's assets, excluding insolvent banks and banks in the process of liquidation. The NBU started to survey banks in 2012 and the data is available quarterly. Banks are surveyed during the first weeks of the quarter.

Businesses. This survey includes answers from about 700 non-financial sector enterprises. Enterprises are selected by the quota principle corresponding to the structure of Ukraine's economy, which ensures the representativeness of the sample. The business surveys have been conducted by the NBU quarterly since February 2006, however, they are conducted during the second month of the quarter.

Financial analysts. The NBU commenced monthly surveys of professional forecasters in July 2014 (during the second and third weeks of each month). Since November 2019, the frequency of this survey dropped to eight times a year to match the schedule of Monetary Policy Committee meetings. Responses from financial analysts are collected one week before the meeting date. The number of professional forecasters varies over time – from six to 12.

Households. Simultaneously with the surveys of financial analysts, a survey of households was launched in July 2014. Unlike the other surveys, the household survey is run monthly by the third-party company Info Sapiens (until 2019 it was run by GfK Ukraine). Every second and third week of the month approximately 1,000 consumers are surveyed about their inflation expectations, and on many other different social and economic issues. The sample is nationally representative and changes each month.

Banks, businesses, and households choose an interval of expected inflation for the next 12 months (more details in Appendix D). The resulting estimate is the weighted average of the midpoints of those intervals. The answer "hard to answer" is also available to households, and these answers are excluded from the calculation of average expectations. At the same time, financial analysts provide their discrete inflation forecasts (actual number, not an interval estimate), and their expectations are the simple average of these estimates. The latter can lead to periodic bias, as the number of experts in the survey is not constant.

Table 1 provides a brief statistical snapshot of inflation expectations in Ukraine. Historically, the expectations of professional forecasters have been lower than all other respondents (the mean is 2-4 pp lower than banks, businesses, and household expectations). However, they do not provide much more accurate forecasts, as forecast error fluctuates in different directions (figure 3), similar to other respondents. Thus, the RMSE of expectations of financial analysts is 12.0 p.p., which is higher than households' and businesses' expectations RMSE (11.4 p.p. and 11.3 p.p. respectively for the same period since July 2014). Banks' expectations show the worst forecasting power, with an RMSE of 13.1 p.p.

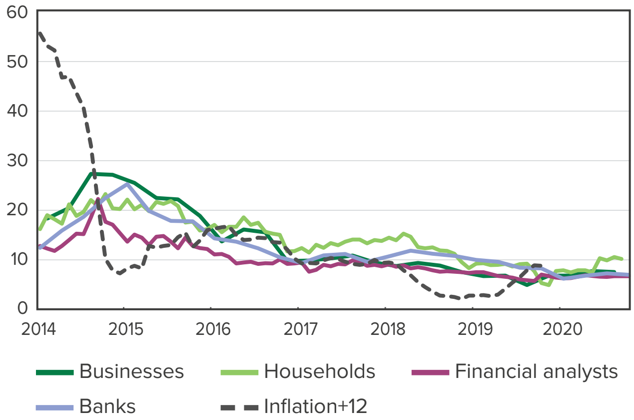


Figure 3. Inflation Expectations for the Next 12 Months and Actual Inflation (+12 months), %

Source: State Statistic Service of Ukraine, NBU, GfK Ukraine, Info Sapiens.

All the expectations have a positive skew, which means that the right tails are quite long. Meanwhile, household expectations are almost symmetrical, having only a small right-skewed tail. The distribution of household inflation expectations is flatter than normal, while all other expectations are more peaked.

3. CONSTRUCTING AGGREGATE NEWS INDEXES

The news content in the corpus is related mainly to economic, social, and political topics. Thus, the sample includes news that is associated not only with inflation developments or expectations (prices, supply of certain goods, tariffs, statistical information, forecasts, etc.). To focus

Table 1. Statistical Properties of Inflation Expectations

	Banks		Businesses		Households	Financial analysts
	Full sample	Since July 2014	Full sample	Since July 2014	Full sample (Since July 2014)	Full sample (Since July 2014)
Count (quarters or months)	38	28	61	27	81	75
Mean, %	10.660	12.120	13.070	13.430	13.780	9.880
std, p.p.	4.990	4.880	5.390	6.920	4.680	3.430
min, p.p.	3.500	5.800	4.700	5.100	4.510	5.340
25%, p.p.	6.890	9.150	9.000	7.800	9.790	7.200
median, p.p.	9.920	10.650	12.760	10.000	13.550	8.800
75%, p.p.	12.000	14.330	15.800	18.650	17.140	12.180
max, p.p.	24.900	24.900	27.300	27.300	22.890	21.900
Skewness	1.074	1.045	0.731	0.811	0.108	1.082
Kurtosis	0.653	0.556	0.132	-0.699	-1.002	0.852

only on the factors determining inflation expectations, I remove news items related to unrelated topics. I apply two different approaches to filter out the news. First, I use a dictionary-based approach to build a set of indexes based on the raw count of news. Second, I implemented a topic analysis using Latent Dirichlet Allocation (LDA) according to Blei et al. (2003).

Both approaches do not take into account the sentiments of news content. However, this may not be a huge problem, as usually the news is biased negatively – Hester & Gibson (2003) found that economic news was written in a negative tone more often than in a positive one. Additionally, they proved that negative news was a significant predictor of consumer expectations about future economic developments. Damstra & Boukes (2018) explain this negative bias of news well, giving a few main reasons:

– free media perform a crucial role in overseeing government, so negative events receive more attention, while positive ones do not meet such a need;

– in the process of judging the newsworthiness of real-world events negativity can be a key value, consequently, a “bad” news story is more likely to be selected by journalists;

– negative events have a stronger news impact than positive ones.

Moreover, Soroka et al. (2019) define the negative tone of news as an essential feature, while good news, in contrast, may be considered as the absence of news. Therefore, to construct simple indexes, it is possible to assume that there is a tendency for news to have a negative impact on perception and expectations.

Of course, a sentiment analysis could be useful to determine the impact of news on economic expectations. In particular, this approach could clean the data series of contradictory events that could have opposite effects. In addition, separating news into positive and negative would help in exploring the possible non-linearity of the impact of multidirectional sentiment. However, to apply this approach, it is important to create a high-quality training dataset, based on clear rules and with the involvement of several independent experts. In addition, sentiment analysis requires significant computing power and time, which may not be reasonable at the initial stages of a study.

3.1. Dictionary-Based Approach

The dictionary-based approach is the simplest approach to estimating the impact of news on various macroeconomic indicators. These indices are calculated as a share of articles related to the topic, commonly denoted as “document frequency”. The intuition behind these indices is that the more alarming the topic, the more articles would be written on the subject – for example in times of crisis.

Document frequency (df) is the fraction of the documents that contain a certain term, and is obtained by dividing the number of documents containing the term by the total number of documents:

$$df(t,D)=\frac{d}{N}, \tag{1}$$

where N is the total number of documents in the corpus D, and number of documents d where the term appears.

The dictionary-based approach to constructing news indices requires expertise in selecting the relevant keywords. In this case, to determine which prices concern Ukrainians the most, I turned to the consumer basket of the average household. Ukrainians spend the most of their income on food. In various periods, the share of spending on food and soft drinks was 40-60% for the period from 2000 to 2020, slightly decreasing in recent years. Accordingly, it is important to select news that contains mentions of basic foods: bread, meat, dairy, vegetables, fruits, etc.

Another essential component of household expenditure is utilities. Although the share of this type of spending is much lower than in many other countries, utility tariffs are important to Ukrainians and are often used as a political football by politicians. They therefore may have a visible impact on expectations. The most important utilities for Ukrainians are electricity and natural gas.

Fuel prices may also have a significant impact on the formation of households’ inflation expectations, even though not all people use private transport. For example, Kilian and Zhou (2020) found several episodes since 1990 in the United States when household inflation expectations growth could almost entirely be explained by hikes in fuel prices. On the one hand, this is due to the ubiquity of gas stations and fuel price boards, which allow their easy use for daily price monitoring. On the other hand, everyone is well aware that fuel is a component of the cost of most goods and services, explicitly or implicitly. In this case, I include news not only about fuel but also about oil, which is a defining cost component of fuel.

As stated in Coibion & Gorodnichenko (2015b) there is a strong positive correlation between inflation expectations and exchange rate developments – especially for households. In this case, logically, not only do exchange rate dynamics affect expectations, but also the coverage of this topic in the media.

In addition, I will analyze the index of news related to the word inflation itself, as such news often contains expert forecasts or analyses of the current situation. According to Zholud et al. (2019), inflation expectations in Ukraine are highly linked to current inflation trends and so have a future-oriented component. Therefore, it is advisable to check the impact of references to inflation in the media on the formation of expectations.

Figure 4 shows the resulting indices calculated using a dictionary-based approach. Since the earliest data on inflation expectations of respondents are from 2006, all the time series of news indices will start from 2005 (one year back to assess lag effects). However, given that the volume of news was much smaller in the early 2000s, this meant only about 10% of the articles had to be removed from the corpus, and there are still about 1.8 million articles left. News related to food has the smallest share among selected topics, while news about fuel appears to be the most important. Also, it can be seen that the document frequency of news related to utilities, in general, decreased over time – except in 2015 when there was a significant jump in the importance of this topic due to utility tariffs being raised to market levels. Interestingly, until 2014 the topic of exchange rate movements was mentioned more often in the news, which is probably due to the greater negative consequences of the sharp devaluation observed in Ukraine at that time with the abandonment of the fixed exchange rate regime. For more information about indices built with a dictionary-based approach, see Appendix E (Supplementary Materials).

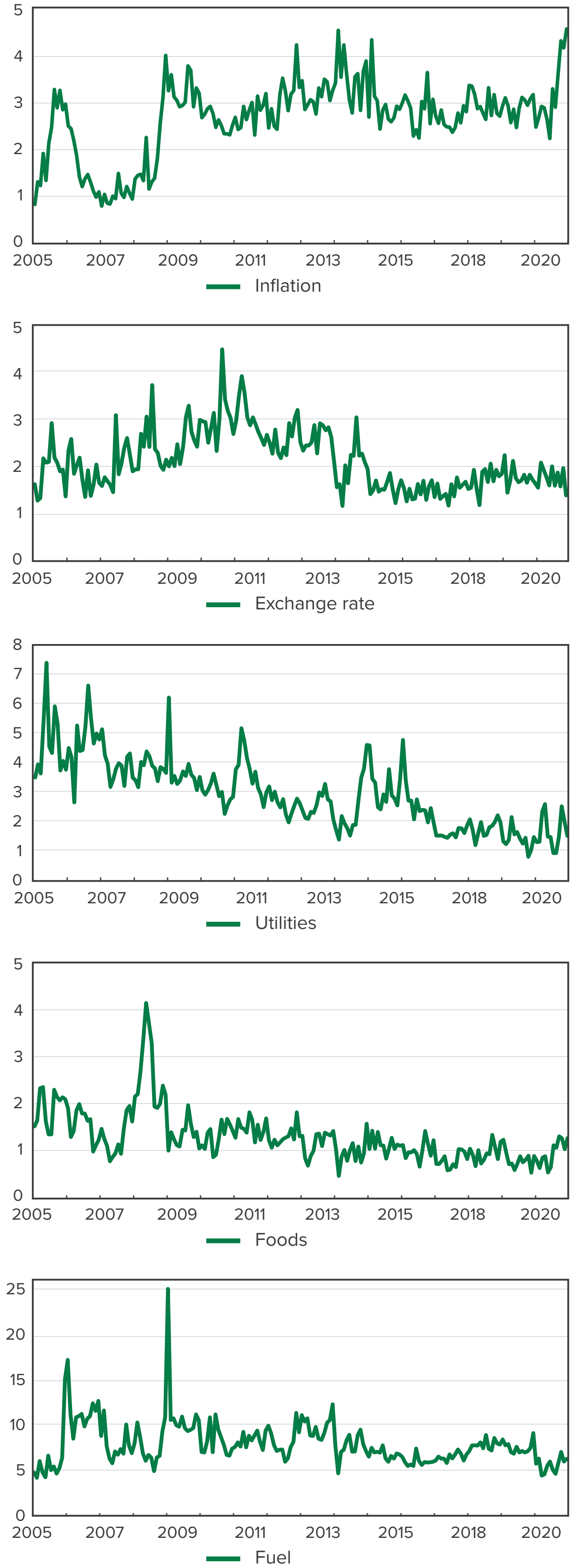


Figure 4. Document Frequency of Topics Relevant to Inflation Expectations

As the inflation expectations of the various respondents are collected at different periods and not evenly throughout the month, the impact of certain short-lived or even discrete news may be extremely important. Thus, some news may last only for several days, and due to the rapid loss of interest in the topic the effect on monthly document frequency can fade away. Therefore, the monthly indices may not reflect the real dynamics of the importance of individual events, and applying indices with a higher frequency may shed light on this issue. To assess this impact, I additionally computed similar indices in decade (10-day) terms for each month – a decade being a third of the month (results are in Appendix E (Supplementary Materials).

It is important to assume the independence of researched variables, which can be indicated by their stationarity. Stationarity is necessary when applying many statistical tools and procedures in a time series analysis. Indeed, if the data was generated by a stationary process, it will have the properties of a sample generated by such a process. According to the Dickey-Fuller test, the monthly time series for the share of utilities, exchange rate, and inflation topics in the news is non-stationary. Resultantly, the relevant indices in a decade-term-only time series for the share of utilities and inflation topics in the news are non-stationary. Additionally, the autocorrelation is high, and it seems that there is no clear seasonality. Therefore, to get rid of the high autocorrelation and to make the entire process stationary, in the same way, I take the first differences.

3.2. Unsupervised ML Approach

One of the important shortcomings of the dictionary-based approach is the availability of quality expertise and the selection of texts based on it. In particular, an article may contain keywords, but its topic may be a completely different issue. For example, the word “fuel” can be attributed to topics related to science and technology or car manufacturing. The solution here is to use unsupervised topic modeling algorithms. These statistical methods analyze the words of the collection of texts and divide them into subgroups, where each subgroup is associated with a set of keywords. Thus, the model finds combinations of words, rather than single ones. In our “fuel” example, articles with word combinations “fuel price” and “rocket fuel” would be distinguished. Most machine-learning models require the use of a part of a data set in which specially trained people classify information according to a predetermined procedure and therefore put labels on data. However, some methods do not need such labeled training samples. Latent Dirichlet Allocation (LDA), presented by Blei et al. in 2003, it today is a very common example of a topic modeling method that uses an unsupervised learning algorithm.

I used an extremely efficient implementation of LDA called LightLDA provided in the nimbusml Python module (Yuan et al., 2015). This state-of-the-art implementation incorporates several optimization techniques and can train a topic model on large document sets much faster. For example, our model produces 100 topics on a 2 million news item dataset in less than an hour, while using the full LDA at this scale takes days. Figure 6 (Appendix B) shows the distribution of topics received with LDA. The popularity of some topics changed over time, while others remained relevant throughout the observation period.

The number of topics in LDA is not fixed and can be set according to the task. I experimented with using a different number of topics. I observed that with a larger number of topics our main results do not change – some topics have very similar content and have to be combined in further analysis. At the same time, the interpretation of a larger number of topics becomes more complicated. With a lower number of topics, it is sometimes difficult to distinguish between different topics that have similar keywords. For example, topics related to exchange rates may include unnecessary information, as some articles contain similar words, but different content.

At this point, human intervention is necessary to analyze and label the topics of the received news clusters. Figure 7 (Appendix B) is a graph showing the relationship between the topics distributed by the LDA. Most of the news clusters are as expected attributed to politics, international relations, parliament, and government. At the same time, some topics are different and can be linked to economic topics that may affect inflation expectations. Most news topics do not belong to one, but to several clusters.

I identify a news cluster related to exchange rate movements, which includes six news topics defined with the help of LDA. As we already know, the situation in the foreign exchange market has some influence on the formation of inflation expectations in Ukraine. I also found a cluster related to commodities, including oil and gas. Additionally, topics describing the electricity market, budget, and government debt can be easily identified. Interestingly, LDA helped identify a topic related to the spread of coronavirus, for which the number of articles unsurprisingly increased from the end of 2019. In addition, LDA has a well-defined topic for the period of the war between russia and Ukraine from 2014 and subsequent years. However, LDA did not group articles related to food prices, utility tariffs, etc., in separate recognizable topics, which can be explained by their having similar structures, as well as their share of the news being relatively low. Increasing the number of topics does not fix this. In Figure 8 (Appendix B) I provide wordclouds for a few of the most relevant topics.

The popularity of certain topics closely corresponds to the historical development of events. In particular, the share of articles on the hryvnia exchange rate, seen in Figure 5, increased in 2008, when the hryvnia depreciated rapidly amid the global financial crisis. The next peak was observed in 2014-2015, when, due to the war between russia and Ukraine and the loss of control of a part of Ukraine's territory, the economy suffered a significant blow. At this time, the hryvnia also depreciated rapidly. But with the transition to a floating exchange rate and stabilization in the foreign exchange market, interest in this topic in the news began to wane.

News about gas and oil behaved in a similar way. Thus, in 2006-2008, the gas issue was extremely important for Ukraine against the background of difficult relations with russia. Problems with gas supply were repeated in 2014. In contrast, lower energy prices have contributed to less coverage of these topics in subsequent years.

I built the indexes the same way as in the dictionary-based approach, using equation 1 of document frequency. Thus, monthly indices were calculated to assess long-term impact, while decadal indices were calculated to assess short-term media shocks that may be important during the time of the inflation expectation survey, but then fade over the month. More details about the statistical characteristics of the indices constructed by LDA division can be found in Appendix F (Supplementary Materials). According to

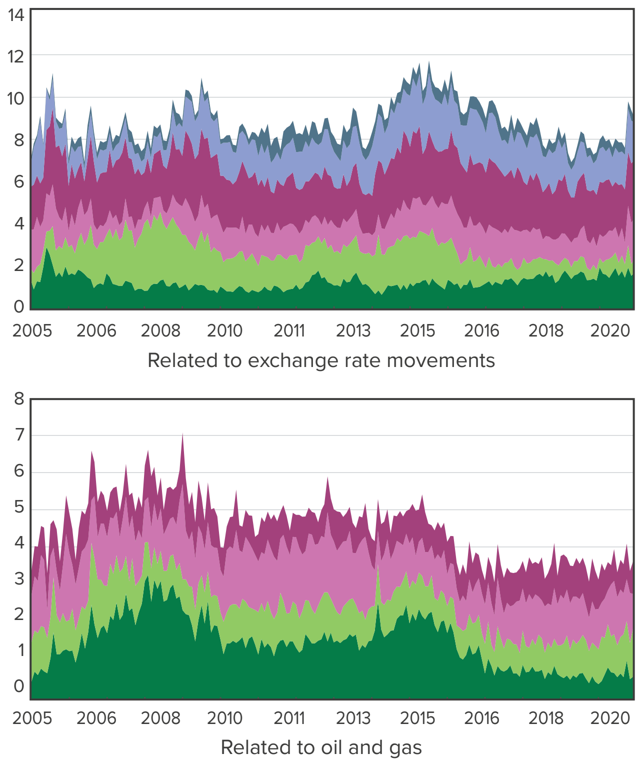


Figure 5. Share of Topics, Identified by LDA, document frequency, %

the Dickey-Fuller test, monthly time series for a share of energy news are non-stationary, while the share of news on exchange rate movements is stationary. Decade time series can be considered stationary with a 95% probability.

4. ESTIMATION RESULTS

As noted in previous sections, inflation expectations are largely shaped by past inflation (Zholud et al., 2019; Coibion and Gorodnichenko, 2015b). Therefore, for the analysis I used an extrapolative approach to the formation of inflation expectations (Lines and Westerhoff, 2010):

$$E\pi_t = \alpha + \beta\pi_{t-1} + \gamma\left(\pi_{t-1} - \pi_{t-2}\right) + \varepsilon, \qquad (2)$$

where $E\pi_t$ is expected inflation in period t, π_{t-1} denotes inflation in the previous period, and $\pi_{t-1} - \pi_{t-2}$ stands for the change in inflation, α and γ – are coefficients of regression, while ε is the error. I use annual CPI change as a measure of inflation.

In this research, I assume that the formation of inflation expectations (equation 2) is influenced by the media environment rather than actual changes in inflation:

$$E\pi_t = \alpha + \beta\pi_{t-1} + \delta df_T^m + \varepsilon, \qquad (3)$$

where df denotes document frequency of the news topic m in period T. T may be equal to t when testing the impact of news on the formation of inflation expectations in the same month a survey is conducted. However, some surveys are conducted at the beginning of the month, therefore, I test the impact of the frequency of news publications during the previous three months on the formation of inflation expectations. Accordingly, T can be equal to t – 1, t – 2, and t – 3. I test the monthly and decade frequency of T, as the inflation expectations survey is not conducted for a whole month, but for shorter periods. In addition, these periods also vary for different respondents. As quarterly surveys are not conducted throughout the quarter, I use matching months instead of aggregating news indices at the quarterly level. For example, bank surveys take place in the first month of the quarter, thus the same month for the news index was used as the base month.

I also test another variation of extrapolative inflation expectations, assuming that the respondents' expectations change in response to changes in current inflation. In this case, the formula of inflation expectations looks like this:

$$E\pi_t - E\pi_{t-1} = \alpha + \gamma\left(\pi_{t-1} - \pi_{t-2}\right) + \varepsilon. \quad (4)$$

I expanded formula 4 with changes in media environments in changes in current inflation:

$$E\pi_t - E\pi_{t-1} = \alpha + \gamma\left(\pi_{t-1} - \pi_{t-2}\right) + \eta\left(df_T^m - df_{T-1}^m\right) + \varepsilon. \quad (5)$$

In this case, all our variables are stationary and we can be sure that their properties do not change over time.

I also assume that the impact of the constructed news indices on inflation expectations is linear, so to estimate this effect I use the OLS regression.

I start with analyzing the impact of news on the formation of inflation expectations using the dictionary-based approach. Table 2 presents the coefficients and p-values (in brackets) of news indices built with a dictionary-based approach in OLS regressions of the inflation expectations of different groups of respondents. This table shows two different approaches: without transformations (equation 3), using all variables as they were computed, and the first differences of all variables, presenting an extrapolation of the change of inflation expectations (equation 5). R^2 for the first type of relationship is as expected much higher than for estimates of the first difference. However, the relatively low R2 for such studies is quite normal and typical for studies of human behavior (King, 1986).

As can be seen from Table 2 (Appendix A), all types of inflation expectations are dependent on current inflation trends as the coefficients are statistically significant. At the same time, only banks and businesses tie changes in inflation expectations to recent changes in inflation, while the relationship between changes in household and financial analyst expectations with recent inflation dynamics is insignificant. This is in line with the opinion that well-anchored long-term inflation expectations should not change in response to news about macroeconomic indicators, in particular inflation (Galati et al., 2011). However, it is too early to talk about anchored inflation expectations, given the difference between the central bank's inflation target and inflation expectations. Therefore, in this case, the result could be due to information rigidity.

The banks' inflation expectations are virtually independent of the current media environment on inflation. Most indicators are not statistically significant or contradict economic logic. For example, the banks' inflation expectations are negatively correlated with food news with a 90% probability. That is, the more this topic is talked about in the media, the faster it reduces the inflation expectations of banks. This might be explained by the tone and content of the news. However, without a more detailed study of the content of this news, this is impossible to determine.

At the same time, it is interesting that banks change their inflation expectations under the influence of changes in the information environment around utility tariffs and fuel, as well as news about inflation in previous periods. The rather significant lag of 2-3 months can be explained by the time needed for the preparation of macroeconomic forecasts, which are the basis for the answers in the survey.

Similar to banks, businesses' expectations may be significantly affected by news about past inflation trends, and by utilities. However, businesses are the most sensitive to the news about food. This can be explained by a high share of agriculture, food industry, retail, and wholesale trade (related to food) enterprises among the surveyed ones, which also corresponds to the structure of the economy of Ukraine. Food news is also an important factor in the estimation of changes in businesses' inflation expectations.

Households' expectations may be the most sensitive to the number of news items related to utility tariffs in the reported period and the previous quarter. This can primarily be explained by the high importance of utility tariffs for Ukrainian households. Thus, a significant share of tariffs are regulated by the government or local authorities, and changes in tariffs cause a substantial negative reaction from society. The share of utility tariffs in the CPI is relatively low, which is largely due to the non-monetary subsidies that were in place in previous periods. However, despite this, the average Ukrainian utility tariffs are some of the most important topics related to inflation, which is confirmed, among other things, by the results of our analysis. Households are also slightly sensitive to information about food in previous periods. Interestingly, these results are not confirmed by changes in households' inflation expectations. Thus, citizens change their estimates of future inflation under the influence of changes in the information field about the exchange rate in the previous three months, while the change in the importance of other topics has little effect on expectations dynamics.

The expectations of professional forecasters respond best to information on utility tariffs in the reporting and previous months, as well as on the exchange rate in the previous three months. In this case, financial analysts respond to both the amount of information, and its change. This may reflect approaches to forecasting for such analysts. As usual, changes in the exchange rate and the expected changes in utility tariffs have the greatest impact on the revision of forecasts.

I also test the hypothesis that shorter-term trends in the media environment may better explain the process of the formation of inflation expectations among different respondents. This is in line with the fact that most surveys are conducted in a shorter period than a month. To this end, I use decadal indices of frequency for mentions of these topics. Shocks in the news that last for several days can fade within a month, due to the rapid loss of interest in the topic, and therefore the monthly indices may not reflect the real dynamics of the importance of individual events. Thus, it is important to apply indices with a higher frequency. Going to the decade level, I get a mixed frequency in the OLS, so to switch to one frequency, I just used matching by month. Thus, I compare news indices separately for the first, second, and third decades of the reporting month with inflation expectations for the same reporting month. The procedure was repeated for individual news indices for the three decades of the previous month against the inflation expectations of the reporting month, as respondents also responded to the dynamics of the media environment in previous periods.

Table 3 (Appendix A) presents the results of OLS estimations of the impact of decade news indices on the formation of inflation expectations. As in the previous case, I add the latest available inflation data, which is published with delay, so I use actual inflation in period t-1.

Here we have a few interesting outcomes that deviate from our monthly estimations. For example, banks may be sensitive to the document frequency of news about the exchange rate in all decades of the previous month, while the monthly indices do not show this relationship. At the same time, food news may be more important in the last decade of the previous month, although monthly indices show significance for the current month. Businesses respond more to the news about utilities and fuel in the first decade of the reporting month. As with monthly indices, decade indices on utility tariffs may affect the formation of the inflation expectations of households. In this case, the first decades of the reporting and previous months are the most important. The expectations of financial analysts also proved to be most dependent on the frequency of news in the first decades of the reporting and previous months. However, in addition to utilities, they may follow the news about inflation (official figures are only published in the first decade of the month) and about fuel.

Another important opinion concerns the fact that the expectations of banks and enterprises are collected once a quarter. Therefore, the period for assessing the impact of news on inflation expectations was increased by applying a three-month moving average. This is especially important considering that the coefficients for the monthly indices are very volatile and can even change sign, depending on the applied lag. Table 4 (Appendix A) presents the results of OLS estimations of the impact of quarterly news indices (3-month moving average) on the formation of inflation expectations.

As can be seen, the hypothesis that banks and enterprises follow longer trends is mostly not confirmed. At the same time, the long-term change in the information space about inflation and the exchange rate is related to the change in inflation expectations of banks, and the change in the volume of food news affects the inflation expectations of enterprises. However, in both cases, this impact is limited to 1-2 quarters.

I repeated a similar procedure to reveal the impact of indices built by the LDA on the formation of inflation expectations. Table 5 (Appendix A) presents the coefficients and p-values (in brackets) of news indices built by the LDA in OLS regressions of inflation expectations of different groups of respondents. Similar to simple indices I tested two different approaches: without transformations, using all variables as they were computed, and the first difference of all variables, presenting an extrapolation of the change of inflation expectations.

According to the results of the regressions, I observe a weak correspondence between the news about energy and utility tariffs that were determined by LDA, and the formation of inflation expectations. Inflation expectations of households, as well as changes in business expectations, demonstrate significance in the reporting month at the 5% and 10% levels respectively, but the sign of the coefficients of these variables contradicts economic logic, which can be associated either with emotional coloring or a reflection of a coincidence of circumstances.

Instead, the situation is somewhat different with the exchange rate news set obtained by LDA. The frequency of such news in the reporting period was significant for the formation of expectations of businesses, households and financial analysts. For households and financial analysts, these indices were also important in recent months. Financial analysts and households were also sensitive to changes in the frequency of exchange rate news. However, households change their expectations in response to more recent developments, while financial analysts respond to a longer period.

Similar to simple indices, I identified short-term spot effects on the formation of inflation expectations by estimating decade indices. Table 6 (Appendix A) represent the results of this estimation.

Interestingly, for some groups of respondents, there is a clear relationship with the indices in the periods when the surveys are conducted. For example, bank surveys are usually conducted at the beginning of the month, and sometimes even cover the last week of the previous month. News about energy in the last decade of the previous month and in the first decade of the reporting month turned out to be significant. There is a similar situation with businesses and households. At the same time, the sign of the coefficients needs further study in terms of sentiment. Inflation expectations of businesses are formed under the influence of news about the exchange rate in the first decade of the reporting month, while all other respondents follow the news for previous periods.

I repeated the same procedure for determining the longer-term impact of news on the formation of inflation expectations using the current three-month average for banks and corporates. Table 7 (Appendix A) shows the main results of the estimations. However, the results indicate the absence of any long-term impact of news on the formation of inflation expectations. Only the expectations of enterprises have a significant connection with the change in the frequency of news about the exchange rate in the current quarter.

Thus, the formation of inflation expectations among different groups of respondents may depend on the media environment, namely both the volume of published articles and changes in this indicator. It is important to note that different groups of respondents may rely on different topics and different periods when estimating future inflation. It can also be seen that recent news, published during the month and even the decade preceding the survey, is mostly more important in shaping inflation expectations than older news. This may, among other things, be important for the central bank's communication policy.

5. CONCLUSIONS

In this paper, I describe the process of analyzing textual data to determine the role of news in the formation of inflation expectations among various types of respondents in Ukraine. I have scraped a news corpus from four Ukrainian online newspapers listed in the most popular online media in Ukraine, which mainly have an economic focus. Using natural language processing and machine-learning techniques, I have cleaned and transformed the textual data into news-based quantitative measures reflecting news topics relevant to inflation and inflation expectations.

I apply two different approaches to filter out the news: a dictionary-based approach and Latent Dirichlet Allocation

(LDA). Both approaches disregard the sentiments of news content, which I leave for future research. I compute all news indices as a "document frequency" following the intuition that the more alarming the topic, the more articles would be written on the subject.

I assume that the impact of the constructed news indices on inflation expectations is linear and estimate this effect with OLS regression. I have tested the impact on the level of inflation expectations, as well as on the change thereof. I have found evidence that different news topics may have a different impact on the inflation expectations of various groups. For example, I find there is a strong relationship between the inflation expectations of households and financial analysts with news about utilities, while businesses are sensitive to news about food. Additionally, financial analysts and households are also sensitive to levels and changes in the frequency of exchange rate news, as shown by LDA.

I also test the hypothesis that shorter-term trends in the media environment may better explain the formation of the inflation expectations of different respondents, as document frequency may vary during the month and the impact of the short-term news may fade away. I prove that for some groups of respondents there is a clear relationship with the indices within the periods when the surveys are conducted. I also show that recent news is mostly more important in shaping inflation expectations than older news.

As a result, the formation of the inflation expectations of different groups of respondents may depend on the media environment, namely the volume of published articles and changes in this indicator. Different groups of respondents rely on different topics and different periods when assessing future inflation. I also find that some events contradict economic logic, which could be a question for future research. In particular, an important issue is the impact of news indices in different periods (i.e. during periods of stability, accelerating inflation, or disinflation). Other future research questions may include an assessment of the tone of news, the relationship of the news indices to other macroeconomic indicators, as well as the predictive power of such indices. Another important issue may be the examination of nonlinearities. In particular, the impact of news may differ depending on the level of current inflation (e.g., economic agents may pay more attention to the news when inflation is high and vice versa) and the monetary policy regime.

These results complement previous studies on the formation of rational inflation expectations. In other words, the overall level of inflation expectations is generally determined by past inflation, and small fluctuations may well be explained by other factors, including the influence of the media environment. The results of this research can aid in understanding inflation expectations, which is important given that anchoring inflation expectations remains a key challenge for central banks. This may, among other things, be important for the central bank's communications policy, and help it to both articulate clear and effective messages and design optimal policy.

REFERENCES

Angelico, C., Marcucci, J., Miccoli, M., Quarta, F. (2021). Can we measure inflation expectations using Twitter? Bank of Italy Working Papers, 1318. Rome: Bank of Italy. Retrieved from https://www.bancaditalia.it/pubblicazioni/temi-discussione/2021/2021-1318/en_tema_1318.pdf

Azqueta-Gavaldón, A. (2017). Developing news-based Economic Policy Uncertainty index with unsupervised machine learning. Economics Letters, 158, 47–50. https://doi.org/10.1016/j.econlet.2017.06.032

Baker, S. R., Bloom, N., Davis, S. J. (2016). Measuring economic policy uncertainty. The Quarterly Journal of Economics, 131(4), 1593–1636. https://doi.org/10.1093/qje/qjw024

Bauer, M. D. (2015). Inflation expectations and the news. International Journal of Central Banking, 11(2), 1-40. Retrieved from https://www.ijcb.org/journal/ijcb15q2a1.pdf

Blei, D. M., Ng, A. Y., Jordan, M. I. (2003). Latent Dirichlet allocation. Journal of Machine Learning Research, 3, 993–1022. Retrieved from https://dl.acm.org/doi/10.5555/944919.944937

Carroll, C. D. (2003). Macroeconomic expectations of households and professional forecasters. The Quarterly Journal of Economics, 118(1), 269–298. https://doi.org/10.1162/00335530360535207

Coibion, O., Gorodnichenko, Y. (2012). What can survey forecasts tell us about information rigidities? Journal of Political Economy, 120(1), 116–159. https://doi.org/10.1086/665662

Coibion, O., Gorodnichenko, Y. (2015a). 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. (2015b). Inflation expectations in Ukraine: A long path to anchoring? Visnyk of the National Bank of Ukraine, 233, 6–23. https://doi.org/10.26531/vnbu2015.233.006

Coibion, O., Gorodnichenko, Y., Weber, M. (2019). Monetary Policy Communications and their Effects on Household Inflation Expectations. NBER Working Paper Series, 25482. Cambridge: National Bureau of Economic Research. https://doi.org/10.3386/w25482

D'Acunto, F., Malmendier, U., Ospina, J., Weber, M. (2019). Exposure to daily price changes and inflation expectations. CESifo Working Paper, 7798. Munich: CESifo. Retrieved from https://www.cesifo.org/en/publications/2019/working-paper/exposure-daily-price-changes-and-inflation-expectations

Damstra, A., Boukes, M. (2018). The economy, the news, and the public: A longitudinal study of the impact of economic news on economic evaluations and expectations. Communication Research, 48(1), 26–50. https://doi.org/10.1177/0093650217750971

Devlin, J., Chang, M.-W., Lee, K., Toutanova, K. (2018). BERT: Pre-training of deep bidirectional transformers for language understanding. Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 1, 4171–4186. http://doi.org/10.18653/v1/N19-1423

Dräger, L. Lamla, M. J. (2017). Imperfect information and consumer inflation expectations: Evidence from microdata. Oxford Bulletin of Economics and Statistics, 79(6), 933–968. https://doi.org/10.1111/obes.12189

Galati, G., Heemeijer, P., Moessner, R. (2011), How do inflation expectations form? New insights from a high-frequency survey. BIS Working Papers, 349. Basel: Bank for International Settlements.

Garcia, J. A. Werner, S. (2018). Inflation news and Euro area inflation expectations. IMF Working Papers, 167. Washington: International Monetary Fund. https://doi.org/10.5089/9781484363010.001

Hester, J. B., Gibson, R. (2003). The economy and second-level agenda setting: A Time-series analysis of economic news and public opinion about the economy. Journalism & Mass Communication Quarterly, 80(1), 73–90. https://doi.org/10.1177/107769900308000106

Hutto, C .J., Gilbert, E. (2014). VADER: A parsimonious rule-based M\model for sentiment analysis of social media text. Proceedings of the International AAAI Conference on Web and Social Media, 8(1), 216–225. https://doi.org/10.1609/icwsm.v8i1.14550

Kilian, L., Zhou, X. (2020). Oil prices, gasoline prices and inflation expectations: A new model and new facts. CEPR Discussion Paper, 15168. London: Centre for Economic Policy Research. Retrieved from https://cepr.org/publications/dp15168

King, G. (1986). How not to lie with statistics: avoiding common mistakes in quantitative political science. American Journal of Political Science, 30, 666. https://doi.org/10.2307/2111095

Korobov, M. (2015). Morphological analyzer and generator for russian and Ukrainian languages. In: Khachay, M., Konstantinova, N., Panchenko, A., Ignatov, D., Labunets, V. (eds) Analysis of Images, Social Networks and Texts. AIST 2015. Communications in Computer and Information Science, 542. Springer. https://doi.org/10.1007/978-3-319-26123-2_31

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

Lines, M., Westerhoff, F. H. (2010). Inflation expectations and macroeconomic dynamics: the case of rational versus extrapolative expectations. Journal of Economic Dynamics and Control, 34(2), 246–257. https://doi.org/10.1016/j.jedc.2009.09.004

Maćkowiak, B., Wiederholt, M. (2009). Optimal sticky prices under rational inattention. The American Economic Review, 99 (3), 769-803. https://doi.org/10.1257/aer.99.3.769

Mankiw, N. G., Reis, R.F., Wolfers, J. (2003). Disagreement about inflation expectations. NBER Macroeconomics Annual, 18, 209–248. Cambridge: National Bureau of Economic Research. https://doi.org/10.1086/ma.18.3585256

Mazumder, S. (2021). The reaction of inflation forecasts to news about the Fed. Economic Modelling, 94, 256–264. https://doi.org/10.1016/j.econmod.2020.09.026

Nautz, D., Pagenhardt, L., Strohsal, T. (2017). The (de-)anchoring of inflation expectations: New evidence from the euro area. The North American Journal of Economics and Finance, 40, 103–115. https://doi.org/10.1016/j.najef.2017.02.002

Pfajfar, D., Santoro, E. (2013). News on inflation and the epidemiology of inflation expectations. Journal of Money, Credit and Banking, 45(6), 1045–1067. https://doi.org/10.1111/jmcb.12043

Rozovskaya, A., Roth, D. (2019). Grammar error correction in morphologically rich languages: The case of russian. Transactions of the Association for Computational Linguistics, 7, 1–17. https://doi.org/10.1162/tacl_a_00251

Sims, C. (2009). Inflation expectations, uncertainty and monetary policy. BIS Working Papers, 275. Basel: Bank for International Settlements. Retrieved from https://www.bis.org/publ/work275.htm

Soroka, S., Fournier, P., Nir, L. (2019). Cross-national evidence of a negativity bias in psychophysiological reactions to news. Proceedings of the National Academy of Sciences of the United States of America, 116(38), 18888–18892. https://doi.org/10.1073/pnas.1908369116

Taboada, M., Brooke, J., Tofiloski, M., Voll, K., Stede, M. (2011). Lexicon-based methods for sentiment analysis. Computational Linguistics, 37(2), 267–307. https://doi.org/10.1162/COLI_a_00049

Tobback, E., Naudts, H., Daelemans, W., Junqué de Fortuny, E., Martens, D. (2016). Belgian economic policy uncertainty index: Improvement through text mining. International Journal of Forecasting, 34(2), 355–365. https://doi.org/10.1016/j.ijforecast.2016.08.006

Woodford, M. (2004). Inflation targeting and optimal monetary policy. Federal Reserve Bank of St. Louis Review, 86(4), 15-41. St. Louis: The Federal Reserve Bank of St. Louis. Retrieved from https://files.stlouisfed.org/files/htdocs/publications/review/04/07/Woodford.pdf

Yuan, J., Gao, F., Ho, Q., Dai, W., Wei, J., Zheng, X., Xing, E.P., Liu, T., Ma, W. (2015). LightLDA: Big topic models on modest computer clusters. WWW'15: Proceedings of the 24th International Conference on World Wide Web, 1351–1361. https://doi.org/10.1145/2736277.2741115

Zholud, O., Lepushynskyi, V., Nikolaychuk, S. (2019). The Effectiveness of the Monetary Transmission Mechanism in Ukraine since the Transition to Inflation Targeting. Visnyk of the National Bank of Ukraine, 247, 19–37. https://doi.org/10.26531/vnbu2019.247.02

APPENDIX A. TABLES

Table 2. Relationship between Monthly News Indices and Inflation Expectations

Respondents	Variables	Without transformations					Variables	1st difference				
		Inflation	Exchange rate	Utilities	Food	Fuel		Inflation	Exchange rate	Utilities	Food	Fuel
Banks	π_{t-1}	0.2949*** (0.000)	0.2890*** (0.000)	0.2994*** (0.000)	0.2970*** (0.000)	0.2950*** (0.000)	$\pi_{t-1} - \pi_{t-2}$	0.1047** (0.022)	0.1545*** (0.005)	0.1642*** (0.003)	0.1658*** (0.003)	0.1737*** (0.001)
	df^m_t	-1.2029 (0.194)	0.9028 (0.510)	-0.7928 (0.356)	-4.0589* (0.099)	0.0736 (0.824)	$df^m_t - df^m_{t-1}$	0.6139 (0.379)	1.5717 (0.189)	0.096 (0.903)	1.6328 (0.460)	0.3832 (0.220)
	df^m_{t-1}	1.1547 (0.360)	-0.7359 (0.662)	2.2787* (0.099)	2.9647 (0.104)	-0.3926 (0.346)	$df^m_{t-1} - df^m_{t-2}$	1.5263 (0.123)	-2.2497 (0.153)	1.2997 (0.306)	0.8267 (0.629)	-0.1084 (0.772)
	df^m_{t-2}	1.3360 (0.153)	-0.7157 (0.591)	-1.7610 (0.143)	-0.4362 (0.840)	-0.7258 (0.113)	$df^m_{t-2} - df^m_{t-3}$	1.614** (0.031)	0.5906 (0.624)	-2.2488** (0.037)	-3.6356* (0.065)	-0.9720** (0.026)
	df^m_{t-3}	-1.0740 (0.264)	-0.6057 (0.675)	0.7886 (0.345)	-1.4298 (0.534)	0.7137* (0.100)	$df^m_{t-3} - df^m_{t-4}$	-1.037 (0.158)	-0.5096 (0.689)	0.9001 (0.246)	2.9091 (0.167)	0.8387** (0.033)
	C	6.4354* (0.093)	9.5205*** (0.000)	5.7224*** (0.000)	10.1442*** (0.001)	9.7139*** (0.000)	C	-8.3529*** (0.006)	0.9782 (0.546)	-0.4893 (0.664)	-2.2835 (0.388)	-1.1227 (0.573)
	R^2	0.8430	0.8250	0.8400	0.8370	0.8450	R^2	0.5700	0.3340	0.3720	0.3400	0.4130
Businesses	π_{t-1}	0.3452*** (0.000)	0.3698*** (0.000)	0.3173*** (0.000)	7.0696*** (0.000)	0.3552*** (0.000)	$\pi_{t-1} - \pi_{t-2}$	0.1434*** (0.003)	0.1344*** (0.004)	0.1467*** (0.002)	0.1326*** (0.003)	0.1400*** (0.003)
	df^m_t	-0.1827 (0.843)	1.2093 (0.245)	-0.2132 (0.768)	0.3255*** (0.000)	0.2268 (0.451)	$df^m_t - df^m_{t-1}$	1.0633 (0.122)	0.9183 (0.232)	-0.8341 (0.153)	0.9607 (0.277)	0.0557 (0.799)
	df^m_{t-1}	-2.0502* (0.073)	-0.5729 (0.672)	1.5272** (0.030)	3.3463*** (0.005)	0.2061 (0.293)	$df^m_{t-1} - df^m_{t-2}$	0.3751 (0.645)	0.0937 (0.926)	1.1851** (0.037)	-0.5685 (0.659)	0.0689 (0.631)
	df^m_{t-2}	3.1070** (0.026)	0.0296 (0.981)	-0.3270 (0.720)	-5.3502*** (0.002)	0.0973 (0.767)	$df^m_{t-2} - df^m_{t-3}$	-0.3006 (0.768)	-0.3382 (0.720)	-0.2001 (0.787)	2.7441* (0.054)	0.0326 (0.910)
	df^m_{t-3}	-0.6576 (0.541)	0.9159 (0.384)	0.3750 (0.584)	3.6920** (0.050)	-0.1574 (0.582)	$df^m_{t-3} - df^m_{t-4}$	-0.6195 (0.442)	-0.4283 (0.576)	-0.2405 (0.664)	-2.7642** (0.012)	-0.3174 (0.203)
	C	8.1511*** (0.000)	4.9233*** (0.010)	5.0653*** (0.000)	-0.1370 (0.924)	5.5163** (0.020)	C	-1.5599 (0.178)	-0.6747 (0.578)	-0.0132 (0.988)	-0.7796 (0.349)	1.1141 (0.489)
	R^2	0.6960	0.6850	0.7460	0.7360	0.6870	R^2	0.2270	0.1790	0.2190	0.2740	0.1880
Households	π_{t-1}	0.2314*** (0.000)	0.2412*** (0.000)	0.1257*** (0.000)	0.1901*** (0.000)	0.2328*** (0.000)	$\pi_{t-1} - \pi_{t-2}$	0.0403 (0.417)	0.0368 (0.440)	0.0538 (0.795)	0.0486 (0.315)	0.0363 (0.456)
	df^m_t	0.3198 (0.708)	0.7091 (0.609)	1.2605** (0.034)	2.2806 (0.146)	0.2741 (0.536)	$df^m_t - df^m_{t-1}$	0.6496 (0.106)	0.9103 (0.155)	-0.3622 (0.268)	0.9230 (0.239)	0.2290 (0.283)
	df^m_{t-1}	-0.1090 (0.905)	0.4006 (0.773)	0.5428 (0.469)	1.8611 (0.253)	0.1923 (0.691)	$df^m_{t-1} - df^m_{t-2}$	-0.5920 (0.169)	-0.5324 (0.406)	0.6644 (0.125)	-0.3849 (0.632)	0.0145 (0.951)
	df^m_{t-2}	0.9379 (0.320)	-0.4846 (0.725)	-0.1165 (0.876)	2.8517* (0.082)	0.3577 (0.463)	$df^m_{t-2} - df^m_{t-3}$	0.6681 (0.136)	-0.7833 (0.223)	-0.6464 (0.135)	1.3106 (0.106)	0.1944 (0.401)
	df^m_{t-3}	1.1311 (0.209)	1.0076 (0.461)	1.5535** (0.011)	1.7897 (0.280)	-0.0602 (0.890)	$df^m_{t-3} - df^m_{t-4}$	-0.0787 (0.861)	1.2966** (0.040)	0.3707 (0.261)	-1.3967* (0.081)	-0.3634* (0.080)

Table 2 (continued). Relationship between Monthly News Indices and Inflation Expectations

Respondents	Variables	Without transformations					Variables	1st difference				
		Inflation	Exchange rate	Utilities	Food	Fuel		Inflation	Exchange rate	Utilities	Food	Fuel
Households	C	3.3627 (0.305)	7.1208** (0.011)	4.9512*** (0.000)	2.2403 (0.289)	4.7291* (0.092)	C	-1.9764 (0.234)	-1.5633 (0.198)	-0.1267 (0.795)	-0.5184 (0.607)	-0.5716 (0.672)
	R^2	0.626	0.6100	0.7390	0.6690	0.6230	R^2	0.1070	0.1070	0.0590	0.0900	0.0680
Financial analysts	π_{t-1}	0.1624*** (0.000)	0.1704*** (0.000)	2.1542*** (0.000)	0.1344*** (0.000)	0.1667*** (0.000)	π_{t-1} - π_{t-2}	-0.0152 (0.696)	-0.0159 (0.668)	-0.0117 (0.752)	-0.0030 (0.940)	-0.0024 (0.953)
	df^m_t	0.1562 (0.794)	0.0426 (0.967)	0.0672*** (0.000)	2.1366* (0.086)	0.1165 (0.729)	df^m_t - df^m_{t-1}	0.5383* (0.100)	0.3330 (0.520)	0.3792 (0.148)	0.4323 (0.538)	0.0671 (0.708)
	df^m_{t-1}	0.9969 (0.133)	-0.1376 (0.892)	1.0333*** (0.004)	1.7494 (0.156)	-0.1388 (0.728)	df^m_{t-1} - df^m_{t-2}	0.6199* (0.086)	-0.8697* (0.092)	-0.4401 (0.215)	-0.3148 (0.646)	-0.0930 (0.667)
	df^m_{t-2}	0.7433 (0.257)	-0.297 (0.765)	0.0352 (0.939)	1.4050 (0.259)	0.2002 (0.628)	df^m_{t-2} - df^m_{t-3}	-0.8320** (0.021)	0.0794 (0.875)	0.8993*** (0.010)	-0.2915 (0.672)	0.1157 (0.595)
	df^m_{t-3}	1.2712* (0.052)	2.3240** (0.022)	0.7893* (0.077)	1.3447 (0.295)	0.1429 (0.677)	df^m_{t-3} - df^m_{t-4}	0.4078 (0.284)	1.5838*** (0.002)	-0.6829** (0.013)	0.6863 (0.333)	0.0082 (0.964)
	C	-2.1422 (0.355)	3.7398* (0.058)	1.2051*** (0.001)	1.1574 (0.491)	4.8133** (0.033)	C	-2.2581* (0.097)	-1.9795** (0.039)	-0.4214 (0.296)	-0.5979 (0.513)	-0.7690 (0.523)
	R^2	0.6750	0.6290	0.8380	0.6600	0.5980	R^2	0.1640	0.1750	0.1480	0.0230	0.0100

Notes: The table shows the results of OLS regressions where inflation expectations are the dependent variable. The time indicator T of document frequencies is set to t, t-1, t-2, and t-3. The first figures in the cells indicate coefficients and p-values are shown in parentheses *p<0.1; **p<0.05; ***p<0.01.

Table 3. Relationship between Decadal News Indices and Inflation Expectations

Respondents	Variables	Inflation		Exchange rate		Utilities		Food		Fuel	
		Curr.	Prev.	Curr.	Prev.	Curr.	Prev.	Curr.	Prev.	Curr.	Prev.
Banks	π_{t-1}	0.3025*** (0.000)	0.3007*** (0.000)	0.2962*** (0.000)	0.2924*** (0.000)	0.2924*** (0.000)	0.2767*** (0.000)	0.2982*** (0.000)	0.3023*** (0.000)	0.2959*** (0.000)	0.2892*** (0.000)
	df^m I	-0.2913 (0.719)	1.0018 (0.272)	-0.5795 (0.593)	-1.7788** (0.030)	0.3746 (0.652)	0.3437 (0.609)	-0.3568 (0.783)	-0.766 (0.452)	-0.5005 (0.287)	0.0193 (0.961)
	df^m II	-0.7864 (0.142)	0.8323 (0.346)	-0.3309 (0.742)	3.6288*** (0.007)	0.2311 (0.732)	0.8281 (0.291)	-1.3411 (0.334)	-0.2261 (0.849)	-0.1320 (0.648)	-0.3379 (0.225)
	df^m III	0.5547 (0.255)	-0.2232 (0.778)	0.4306 (0.680)	-2.2436** (0.012)	-0.0498 (0.932)	-0.2502 (0.738)	-0.7180 (0.637)	1.8529* (0.098)	0.1892 (0.568)	-0.0686 (0.799)
	C	8.6576*** (0.002)	2.2805 (0.494)	8.2388*** (0.000)	8.2280*** (0.000)	6.0084*** (0.000)	5.2585*** (0.000)	9.7445*** (0.000)	6.6276*** (0.000)	10.4832*** (0.000)	10.3404*** (0.000)
	R^2	0.8250	0.8250	0.8140	0.8670	0.8160	0.8280	0.8200	0.8300	0.8230	0.8290

Table 3 (continued). Relationship between Decadal News Indices and Inflation Expectations

Respondents	Variables	Inflation		Exchange rate		Utilities		Food		Fuel	
		Curr.	Prev.	Curr.	Prev.	Curr.	Prev.	Curr.	Prev.	Curr.	Prev.
Businesses	π_{t-1}	0.3429*** (0.000)	0.3468*** (0.000)	0.3645*** (0.000)	0.3590*** (0.000)	0.3285*** (0.000)	0.3165*** (0.000)	0.3300*** (0.000)	0.3378*** (0.000)	0.3517*** (0.000)	0.3527*** (0.000)
	df^m I	-0.0008 (0.999)	-0.5610 (0.409)	-0.0527 (0.940)	0.7648 (0.331)	0.9796* (0.052)	0.7777 (0.142)	1.3742 (0.104)	1.1391 (0.228)	0.7220** (0.019)	0.4026 (0.146)
	df^m II	-0.7326 (0.355)	0.0077 (0.991)	1.1940 (0.167)	-0.7624 (0.299)	-0.0830 (0.908)	0.5752 (0.206)	0.5473 (0.558)	-0.2134 (0.865)	-0.1173 (0.682)	0.0089 (0.955)
	df^m III	0.9513 (0.167)	0.1996 (0.723)	-0.0056 (0.995)	1.2836 (0.110)	0.2448 (0.725)	0.1232 (0.812)	-0.2146 (0.831)	-0.5889 (0.632)	-0.0325 (0.906)	0.0459 (0.853)
	C	8.0788*** (0.000)	9.5355*** (0.000)	5.9074*** (0.001)	5.8469*** (0.001)	5.7110*** (0.000)	4.7702*** (0.000)	6.5813*** (0.000)	8.1705*** (0.000)	3.9404* (0.051)	4.8362*** (0.006)
	R^2	0.6660	0.6590	0.6860	0.6890	0.7250	0.7480	0.6880	0.6640	0.6980	0.6900
Households	π_{t-1}	0.2332*** (0.000)	0.2338*** (0.000)	0.2273*** (0.000)	0.2341*** (0.000)	0.1569*** (0.000)	0.1519*** (0.000)	0.2257*** (0.000)	0.2177*** (0.000)	0.2307*** (0.000)	0.2337*** (0.000)
	df^m I	0.6075 (0.470)	0.3696 (0.641)	-2.1987** (0.031)	-0.6561 (0.507)	1.9786*** (0.003)	2.3171*** (0.001)	0.7161 (0.587)	1.3905 (0.280)	0.9265** (0.017)	0.6657 (0.098)
	df^m II	-0.5300 (0.317)	-0.4296 (0.421)	0.8797 (0.287)	0.6005 (0.512)	0.6423 (0.249)	0.3836 (0.485)	2.0484* (0.088)	1.5500 (0.192)	0.3043 (0.305)	0.4797* (0.083)
Households	df^m III	0.3656 (0.425)	0.3585 (0.484)	1.4200 (0.159)	0.7847 (0.463)	0.1211 (0.810)	0.2490 (0.621)	0.2029 (0.851)	0.4941 (0.682)	-0.4318 (0.120)	-0.4599 (0.136)
	C	8.7718*** (0.001)	9.1639*** (0.000)	10.3816*** (0.000)	8.9515*** (0.000)	5.5579*** (0.000)	5.2581*** (0.000)	7.0805*** (0.000)	6.7930*** (0.000)	4.4817 (0.074)	5.1578*** (0.043)
	R^2	0.6090	0.6050	0.6480	0.6140	0.7090	0.7180	0.6280	0.6270	0.6470	0.6380
Financial analysts	π_{t-1}	0.1604*** (0.000)	0.1623*** (0.000)	0.1652*** (0.000)	0.1662*** (0.000)	0.1028*** (0.000)	0.0922*** (0.000)	0.1508*** (0.000)	0.1492*** (0.000)	0.1655*** (0.000)	0.1668*** (0.000)
	df^m I	1.692*** (0.006)	1.4931** (0.011)	-1.1937 (0.134)	-0.8612 (0.263)	1.6138*** (0.001)	1.8996*** (0.000)	1.7747* (0.095)	1.4521 (0.131)	0.6686** (0.033)	0.5142* (0.100)
	df^m II	-0.1233 (0.775)	-0.0997 (0.805)	0.5855 (0.359)	0.7450 (0.299)	0.7149* (0.069)	0.5090 (0.175)	0.4292 (0.646)	0.3017 (0.733)	0.0583 (0.796)	0.1352 (0.544)
	df^m III	0.0259 (0.939)	0.4572 (0.218)	0.9114 (0.275)	0.4700 (0.568)	-0.0186 (0.958)	0.2341 (0.506)	0.7651 (0.354)	1.1162 (0.228)	-0.2818 (0.190)	-0.3239 (0.172)
	C	2.5993 (0.138)	1.7719 (0.325)	6.7942*** (0.000)	6.6295*** (0.000)	3.2151*** (0.000)	2.7488*** (0.000)	4.3551*** (0.000)	4.5878*** (0.000)	3.9331* (0.051)	4.7127** (0.023)
	R^2	0.6340	0.6420	0.6210	0.6110	0.7480	0.7690	0.6270	0.6230	0.6230	0.6130

Notes: The table shows the results of OLS regressions where inflation expectations are the dependent variable. The first figures in cells indicate coefficients and p-values are shown in parentheses *p<0.1; **p<0.05; ***p<0.01. Curr. stands for decades of reported month, Prev. – for the previous month. Indicators I, II, and III following document frequency indices denote the number of decades.

Table 4. Relationship between Quarterly News Indices and Inflation Expectations

Respondents	Variables	Without transformations					Variables	1st difference				
		Inflation	Exchange rate	Utilities	Food	Fuel		Inflation	Exchange rate	Utilities	Food	Fuel
Banks	π_{t-1}	0.2962***	0.2820***	0.2832***	0.3004***	0.2898***	$\pi_{t-1} - \pi_{t-2}$	0.0623	0.0903**	0.1041**	0.1014**	0.0895**
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)		(0.154)	(0.033)	(0.015)	(0.020)	(0.033)
	df^m_t	-1.7001	1.1527	0.1101	1.3103	-0.0158	$df^m_t - df^m_{t-1}$	1.6473**	1.5458*	-0.4047	0.3991	0.2467
		(0.170)	(0.467)	(0.918)	(0.571)	(0.973)		(0.017)	(0.094)	(0.516)	(0.766)	(0.368)
	df^m_{t-1}	0.7511	-1.7379	1.1679	-1.4879	-0.4129	$df^m_{t-1} - df^m_{t-2}$	-0.7856	-3.4065***	1.6354	-1.1473	-0.6507
		(0.669)	(0.440)	(0.557)	(0.672)	(0.615)		(0.425)	(0.010)	(0.165)	(0.576)	(0.176)
	df^m_{t-2}	1.5584	-0.1744	-1.2997	-0.6431	-0.4109	$df^m_{t-2} - df^m_{t-3}$	0.0814	1.3132	-2.1069*	-0.0197	0.1333
		(0.669)	(0.938)	(0.512)	(0.855)	(0.616)		(0.933)	(0.322)	(0.075)	(0.992)	(0.782)
	df^m_{t-3}	-1.1049	-0.5921	0.8456	0.0465	0.3926	$df^m_{t-3} - df^m_{t-4}$	0.2624	0.6112	0.9212	0.7249	0.4089
		(0.394)	(0.704)	(0.421)	(0.984)	(0.399)		(0.719)	(0.508)	(0.140)	(0.582)	(0.139)
	C	8.6963***	9.9577***	5.5146***	7.9478***	10.5852***	C	-3.6686***	-0.1601	-0.1348	0.0122	-1.0700
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)		(0.006)	(0.781)	(0.764)	(0.987)	(0.174)
	R^2	0.8120	0.8220	0.8170	0.8100	0.8260	R^2	0.159	0.1140	0.0800	0.0590	0.1340
Businesses	π_{t-1}	0.3567***	0.3775***	0.3229***	0.3442***	0.3681***	$\pi_{t-1} - \pi_{t-2}$	0.1549***	0.1513***	0.1549***	0.1451***	0.1468***
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)		(0.000)	(0.000)	(0.000)	(0.001)	(0.000)
	df^m_t	-1.3567	1.2189	0.6571	1.4312	0.3357	$df^m_t - df^m_{t-1}$	0.7995	0.8543	-0.3296	1.9479***	-0.0014
		(0.305)	(0.364)	(0.428)	(0.387)	(0.245)		(0.166)	(0.162)	(0.416)	(0.009)	(0.991)
	df^m_{t-1}	0.2246	-0.788	0.7613	-1.2567	0.1356	$df^m_{t-1} - df^m_{t-2}$	-0.7318	-1.3392	1.3800*	-3.1614**	0.1989
		(0.914)	(0.691)	(0.596)	(0.665)	(0.757)		(0.421)	(0.137)	(0.051)	(0.015)	(0.319)
	df^m_{t-2}	1.4987	-0.2494	-0.3459	-0.6994	-0.0457	$df^m_{t-2} - df^m_{t-3}$	0.8714	0.6937	-1.1423	1.8089	-0.3072
		(0.470)	(0.899)	(0.807)	(0.809)	(0.916)		(0.335)	(0.442)	(0.107)	(0.159)	(0.120)
	df^m_{t-3}	0.0660	1.5895	0.4552	2.1718	0.1126	$df^m_{t-3} - df^m_{t-4}$	-0.8554	-0.0955	0.0518	-0.4610	0.0751
		(0.961)	(0.233)	(0.576)	(0.186)	(0.676)		(0.144)	(0.875)	(0.898)	(0.524)	(0.541)
	C	7.4593***	4.5674***	4.6576***	6.6463***	4.0647***	C	-0.2782	-0.2835	0.0782	-0.2130	0.2446
		(0.000)	(0.000)	(0.000)	(0.000)	(0.001)		(0.492)	(0.498)	(0.798)	(0.475)	(0.637)
	R^2	0.6760	0.6950	0.7460	0.6870	0.6960	R^2	0.1100	0.0880	0.1110	0.1110	0.0940

Notes: The table shows the results of OLS regressions where inflation expectations are the dependent variable. Time indicator T of document frequencies is set to t, t-1, t-2, and t-3 and corresponds to quarters. The first figures in cells indicate coefficients and p-values are shown in parentheses *p<0.1; **p<0.05; ***p<0.01.

Table 5. Relationship between Monthly News Indices Built by LDA and Inflation Expectations

Topic	Variables	Without transformations				Variables	1st difference			
		Banks	Businesses	Households	Financial analysts		Banks	Businesses	Households	Financial analysts
Energy	π_{t-1}	0.2707*** (0.000)	0.3437*** (0.000)	0.2051*** (0.000)	0.1626*** (0.000)	$\pi_{t-1}-\pi_{t-2}$	0.1731*** (0.003)	0.1598*** (0.001)	0.0532 (0.291)	-0.0016 (0.970)
	df^m_t	0.4646 (0.766)	-1.0941 (0.413)	-2.7212** (0.037)	-0.3453 (0.754)	$df^m_t - df^m_{t-1}$	0.7627 (0.622)	-1.8396* (0.059)	-1.1436 (0.111)	0.3425 (0.564)
	df^m_{t-1}	-1.8021 (0.296)	2.4938* (0.072)	-1.5848 (0.196)	-0.3938 (0.708)	$df^m_{t-1} - df^m_{t-2}$	-1.0257 (0.538)	1.2913 (0.197)	1.0800 (0.116)	0.2248 (0.693)
	df^m_{t-2}	-0.9249 (0.405)	0.5825 (0.657)	-1.7524 (0.151)	-0.4304 (0.677)	$df^m_{t-2} - df^m_{t-3}$	-0.2834 (0.809)	0.2077 (0.833)	0.0709 (0.916)	0.4140 (0.455)
	df^m_{t-3}	-0.5554 (0.758)	-2.0173 (0.112)	-1.8030 (0.148)	-1.5595 (0.165)	$df^m_{t-3} - df^m_{t-4}$	-0.1006 (0.956)	-0.8172 (0.374)	0.2491 (0.249)	-1.0749* (0.070)
	C	15.9023*** (0.000)	8.9195** (0.044)	32.4739*** (0.000)	14.6931*** (0.000)	C	1.9027 (0.579)	3.4848 (0.209)	-0.7805 (0.729)	0.1552 (0.936)
	R^2	0.8460	0.6850	0.7180	0.6210	R^2	0.276	0.2150	0.0720	0.0540
Exchange rate	π_{t-1}	0.2207*** (0.000)	0.2349*** (0.000)	0.0553** (0.028)	0.0443** (0.021)	$\pi_{t-1}-\pi_{t-2}$	0.1404** (0.011)	0.1296*** (0.005)	0.0376 (0.422)	-0.0148 (0.703)
	df^m_t	0.2808 (0.622)	1.9069*** (0.008)	1.2814*** (0.004)	0.7794** (0.022)	$df^m_t - df^m_{t-1}$	0.25 (0.667)	0.6776 (0.232)	0.6895** (0.020)	-0.0101 (0.968)
	df^m_{t-1}	1.0477 (0.254)	-0.6672 (0.334)	0.0629 (0.904)	0.9172** (0.032)	$df^m_{t-1} - df^m_{t-2}$	0.9799 (0.265)	-0.1744 (0.770)	-0.7571** (0.030)	0.6726** (0.031)
	df^m_{t-2}	0.7299 (0.397)	0.4674 (0.571)	0.7218 (0.165)	-0.2140 (0.608)	$df^m_{t-2} - df^m_{t-3}$	0.5198 (0.529)	0.4751 (0.509)	0.5903* (0.088)	-0.7020** (0.025)
	df^m_{t-3}	-0.4810 (0.520)	0.7308 (0.322)	1.0341** (0.022)	0.6904** (0.047)	$df^m_{t-3} - df^m_{t-4}$	-1.4949** (0.033)	-0.9657 (0.109)	-0.4956* (0.085)	0.0254 (0.917)
	C	-5.9118 (0.203)	-11.2109** (0.015)	-15.1982*** (0.000)	-10.7207*** (0.000)	C	-2.2814 (0.441)	-0.2290 (0.938)	-0.3152 (0.800)	0.0642 (0.953)
	R^2	0.8640	0.7560	0.8210	0.8090	R^2	0.397	0.2170	0.1470	0.1020

Notes: The table shows the results of OLS regressions where inflation expectations are the dependent variable. Time indicator T of document frequencies is set to t, t-1, t-2, and t-3. The first figures in cells indicate coefficients and p-values are shown in parentheses *p<0.1; **p<0.05; ***p<0.01.

Table 6. Relationship between Decade News Indices Built by LDA and Inflation Expectations

Respondents	Variables	Banks		Businesses		Households		Financial analysts	
		Curr.	Prev.	Curr.	Prev.	Curr.	Prev.	Curr.	Prev.
Energy	π_{t-1}	0.2788*** (0.000)	0.2715*** (0.000)	0.3554*** (0.000)	0.3647*** (0.000)	0.20600*** (0.000)	0.2151*** (0.000)	0.1588*** (0.000)	0.1605*** (0.000)
	df^m I	-2.2081** (0.038)	-1.2279 (0.176)	0.8904 (0.394)	-0.1089 (0.914)	-2.8889*** (0.004)	-2.5453** (0.013)	-1.3360 (0.103)	-1.3255* (0.098)
	df^m II	0.1589 (0.875)	0.6438 (0.413)	-1.9179** (0.043)	0.3250 (0.702)	-1.1732 (0.163)	-0.9812 (0.243)	-0.5387 (0.478)	0.1262 (0.858)
	df^m III	-0.1123 (0.891)	-1.8735** (0.017)	1.0673 (0.102)	1.2336 (0.136)	-1.1909* (0.078)	-0.9189 (0.197)	0.0937 (0.882)	-0.4895 (0.384)
	C	13.8197*** (0.001)	14.9455*** (0.000)	8.3880** (0.025)	4.0943 (0.210)	25.1004*** (0.000)	22.6818*** (0.000)	12.0712*** (0.000)	11.8456*** (0.000)
	R^2	0.8380	0.8650	0.6870	0.6730	0.6890	0.6650	0.6130	0.6150
Exchange rate	π_{t-1}	0.2496*** (0.000)	0.2161*** (0.000)	0.2851*** (0.000)	0.3141*** (0.000)	0.1237*** (0.000)	0.1121*** (0.000)	0.0774*** (0.000)	0.0704*** (0.000)
	df^m I	0.8894* (0.072)	0.8441** (0.049)	1.0070** (0.014)	-0.0078 (0.987)	1.8598*** (0.000)	1.9123*** (0.000)	1.5874*** (0.000)	1.2134*** (0.000)
	df^m II	0.8664** (0.038)	0.4398 (0.371)	-0.0442 (0.917)	0.3258 (0.463)	0.5687* (0.058)	0.7552** (0.012)	0.5167** (0.016)	0.5048** (0.022)
	df^m III	-0.6396 (0.110)	0.3476 (0.424)	0.7195 (0.132)	0.4304 (0.328)	-0.2831 (0.373)	-0.4095 (0.184)	-0.3247 (0.132)	0.0570 (0.824)
	C	-2.2563 (0.566)	-6.3928 (0.126)	-5.2484 (0.155)	2.4127 (0.520)	-7.5191*** (0.006)	-8.4340*** (0.003)	-7.5665*** (0.000)	-7.4620*** (0.000)
	R^2	0.8490	0.8640	0.7340	0.6820	0.7520	0.7560	0.7960	0.7720

Notes: The table shows the results of OLS regressions where inflation expectations are a dependent variable. The first figures in cells indicate coefficients and p-values are shown in parentheses *p<0.1; **p<0.05; ***p<0.01. Curr. stands for decades of reported month, Prev. – for previous month. Indicators I, II and III following document frequency indices denote number of decades.

Table 7. Relationship between Quarterly News Indices Built by LDA and Inflation Expectations

Topic	Variables	Without transformations		Variables	1st difference	
		Banks	Businesses		Banks	Businesses
Energy	π_{t-1}	0.2675*** (0.000)	0.3566*** (0.000)	$\pi_{t-1} - \pi_{t-2}$	0.1846*** (0.001)	0.1466*** (0.002)
	df^m_t	-1.3148 (0.744)	1.1802 (0.661)	$df^m_t - df^m_{t-1}$	0.3972 (0.914)	-2.3983 (0.206)
	df^m_{t-1}	-2.4590 (0.609)	1.6579 (0.695)	$df^m_{t-1} - df^m_{t-2}$	-5.1964 (0.261)	4.6212 (0.137)
	df^m_{t-2}	1.2399 (0.760)	-2.7387 (0.503)	$df^m_{t-2} - df^m_{t-3}$	2.3077 (0.550)	-3.1690 (0.275)
	df^m_{t-3}	-0.5603 (0.775)	0.4886 (0.857)	$df^m_{t-3} - df^m_{t-4}$	2.1470 (0.248)	-0.0908 (0.962)
	C	16.6815*** (0.000)	6.7721 (0.148)	C	0.9729 (0.758)	3.1107 (0.271)
	R^2	0.8440	0.6620	R^2	0.3410	0.1980
Exchange rate	π_{t-1}	0.2194*** (0.000)	0.2357*** (0.000)	$\pi_{t-1} - \pi_{t-2}$	-0.7182** (0.021)	0.1224*** (0.008)
	df^m_t	1.8980 (0.214)	2.8326* (0.059)	$df^m_t - df^m_{t-1}$	1.8418 (0.207)	2.6938 (0.022)
	df^m_{t-1}	-0.9758 (0.752)	-2.1270 (0.399)	$df^m_{t-1} - df^m_{t-2}$	-0.0609 (0.983)	-3.3550* (0.098)
	df^m_{t-2}	1.3292 (0.665)	0.1740 (0.941)	$df^m_{t-2} - df^m_{t-3}$	0.1379 (0.961)	1.8010 (0.341)
	df^m_{t-3}	-0.6878 (0.639)	1.4948 (0.375)	$df^m_{t-3} - df^m_{t-4}$	-1.8472 (0.156)	-1.1601 (0.388)
	C	-5.8439 (0.247)	-10.7359** (0.025)	C	0.1227 (0.807)	0.0370 (0.990)
	R^2	0.8590	0.7430	R^2	0.4320	0.2410

Notes: The table shows results of OLS regressions where inflation expectations are the dependent variable. Time indicator T of document frequencies is set to t, t-1, t-2 and t-3 and corresponds to quarters. The first figures in cells indicate coefficients and p-values are shown in parentheses *p<0.1; **p<0.05; ***p<0.01.

APPENDIX B. FIGURES

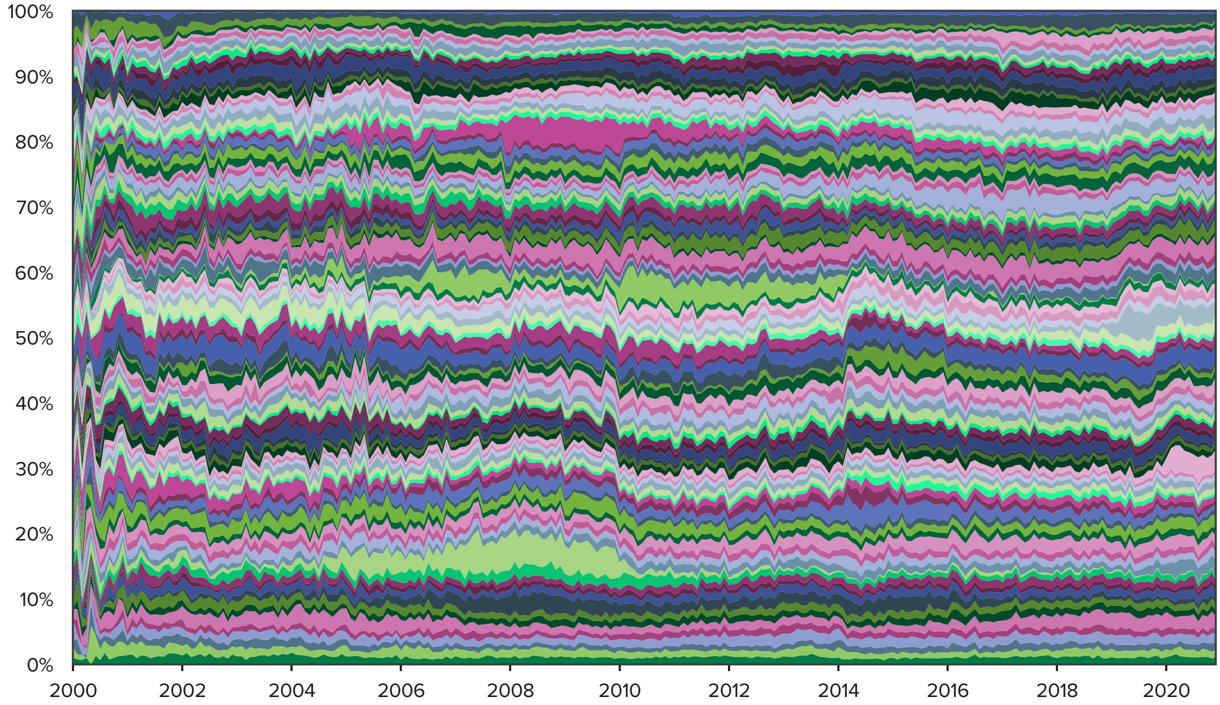


Figure 6. Distribution of Topics Received with LDA

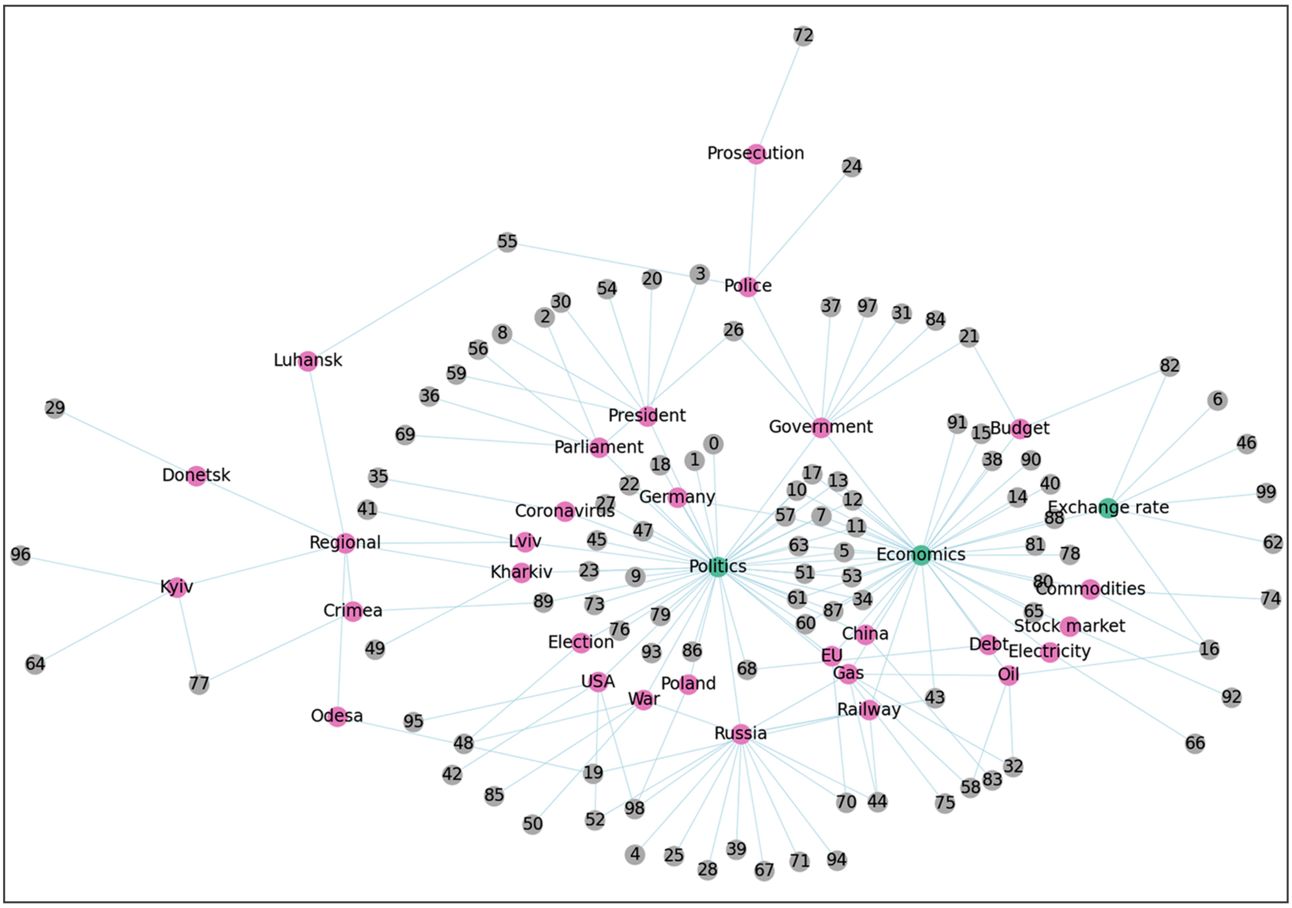


Figure 7. Graph of Topics in News Corpus. Grey Circles Refer to Topics Defined by LDA (pink circles - manually labeled clusters, and green circles – general topics)

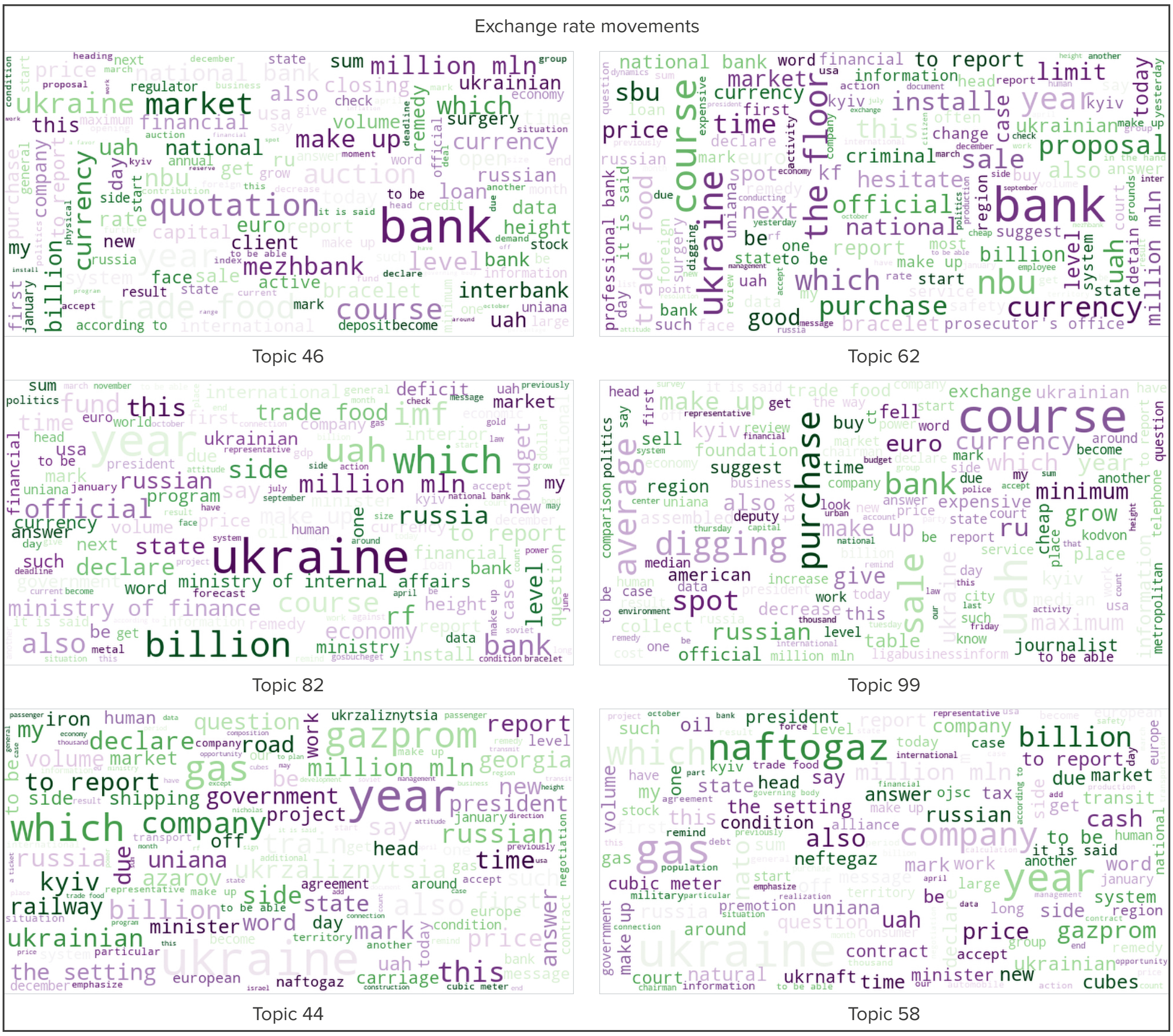


Figure 8. Wordclouds for Selected Topics

APPENDIX C. NEWS CORPUS

Originally, the news corpus consisted of 2,030,000 unique articles. However, after cleaning and filtering out items with various types of errors (parsing errors when web-page tags are wrongly placed, empty pages, corrupted symbols etc.), the number of articles decreased by 50,000 items. As this was only 2.5% of the total number of articles, I consider such a reduction quite acceptable, and that it will not affect the overall result.

Table 8. Article Size in News Corpus (after cleaning)

	finance.ua	liga	ukr_pravda	unian	Total
count	389,951	620,655	339,275	634,832	1,985,143
mean	120.6	121.5	131.0	151.5	132.5
std	96.8	79.6	94.2	123.9	102.1
min	0	2	4	3	0
25%	63	76	83	88	78
50%	99	108	115	127	113
75%	151	149	157	182	162
max	3,832	11,540	5,842	3,986	11,540
skewness	3.305	24.465	9.911	7.766	10.072
kurtosis	32.299	2,456.736	228.865	109.719	394.412

The articles in the corpus differ not only in content but also in writing style, size as measured by word count, and other features (Table 8). Expectedly, different sources of information have some dissimilarities in how the news is written, which is, for example, revealed in the article size. Unian has the largest articles on average, while finance.ua has the smallest articles. At the same time, the sizes of the articles from all sources are very close.

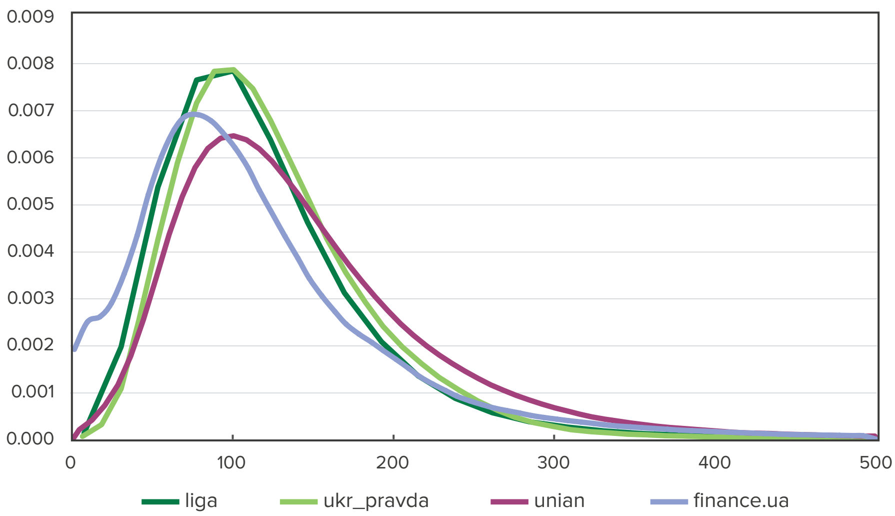


Figure 9. Distribution of Article Size by Sources

The distribution of article sizes for all news sources (figure 9) is highly asymmetrical. All values of the skewness are positive, and the tail of the distribution is longer towards the right-hand side of the curve. Articles from Liga are the most skewed. At the same time, distributions of article length are leptokurtic, which means they are tall and thin, and so near the mean. For example, the number of articles with a length of more than 500 words is less than 15,000, which is only 0.7% of the corpus. The number of articles with extremely small size is also negligible (around 2.5% of the total number).[1]

[1] The average sentence ranges from 15 to 20 words, I consider that the smallest article consists of a topic and one sentence = 30 words.

APPENDIX D. INFLATION EXPECTATIONS SURVEY DESIGN

In contrast to financial analysts, who are asked to answer open questions, banks, businesses, and households are asked to pick from a set of inflation intervals, for example:

"Inflation over the next twelve months will be:

a) less than zero ("prices will fall"),
b) between 0 and X percent,
c) between X and 2X percent,
d) between 2X and 3X percent,
e) between 3X and 4X percent,
f) over 4X percent.

In this example, inflation expectations would be computed by the formula:

$$E\pi = w_a \cdot \left(-\frac{X}{2}\right) + w_b \cdot \frac{X}{2} + w_c \cdot \frac{X+2X}{2} + w_d \cdot \frac{2X+3X}{2} + w_e \cdot \frac{3X+4X}{2} + w_f \cdot \left(4X + \frac{X}{2}\right), \tag{6}$$

where w is the share of respondents who pick the respective interval. Size of X as well as number of intervals is not fixed and changes over time to match the normal distribution of answers. Thus, in 2015 inflation in Ukraine accelerated drastically, so the maximum bracket was expanded to 50% and respondents selected from 12 intervals. Following disinflation in 2020 maximum bracket was decreased to 10% and number of intervals was cut to eight.

Since January 2018, the surveys of households have also included a question about inflation perceptions. Once a year, consumers are asked to answer one open question about perceived inflation over the previous 12 months. Additionally, households are also asked to pick answers from an interval question once a quarter. The construction of this question is similar to the inflation expectations question.