National Bank of Ukraine

Role of the media in the inflation expectation formation process

Tetiana Yukhymenko
National Bank of Ukraine

September 2022

Content

- Motivation
- Literature overview
- Inflation expectations
- News corpus
- Constructing aggregate news indexes
  - Dictionary-based approach
  - Unsupervised ML approach
- Estimation results
- Concluding remarks

Disclaimer: The views expressed in the research are those of the author and do not necessarily represent the views of the National Bank of Ukraine.

Motivation

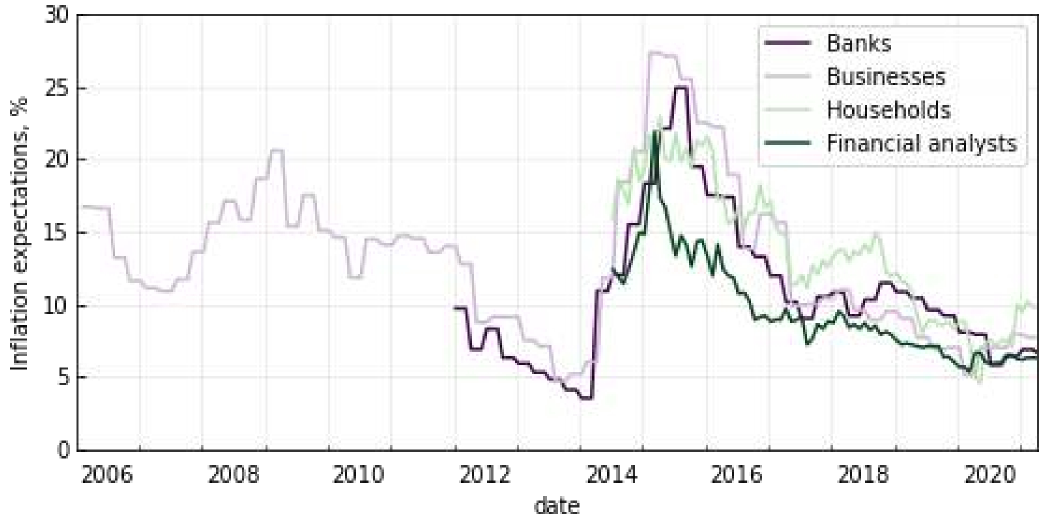
- Anchoring inflation expectations remains a key challenge for central banks, especially in developing countries.
- It can be assumed that respondents are also influenced by tax, tariffs, spending, and monetary and regulatory policy uncertainties. These effects, however, are hard to detect because uncertainty is unobservable.
- The news-based method could be used to investigate the impact of the media environment on the formation of respondents' expectations.

Literature overview

- Information rigidities have a large impact on the macroeconomic variables, thus they should be integrated into modern macroeconomic policies to execute the optimal monetary policy (Coibion and Gorodnichenko, 2012).
- Many modern researchers are trying to investigate the impact of news on the formation of inflation expectations:
  - Dräger and Lamla (2017) found evidence that respondents are more likely to adjust their expectations if they have heard the news on inflation.
  - Bauer (2015) used macroeconomic data surprises cumulated over the monthly or quarterly observation windows as an economic news measure.
  - Larsen et al. (2020) applied machine learning algorithms to a large news corpus and examined the role of the media in the expectation formation process of households.
  - Goloshchapova and Andreev (2017) proposed using machine learning techniques on textual data to measure inflation expectations.
  - Angelico et al. (2021) used a similar approach to build real-time measures of consumers' inflation expectations from tweets.

Data: Inflation expectations

Inflation expectations for the next 12 months, %

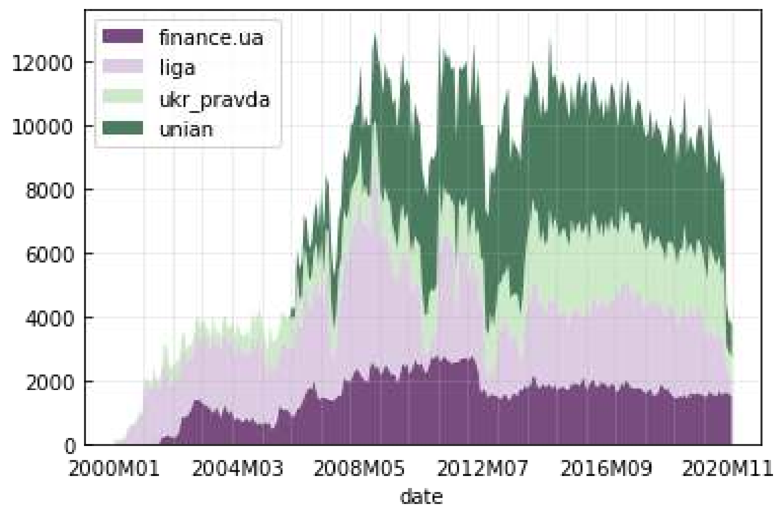


Source: NBU, GfK Ukraine, Info Sapiens.

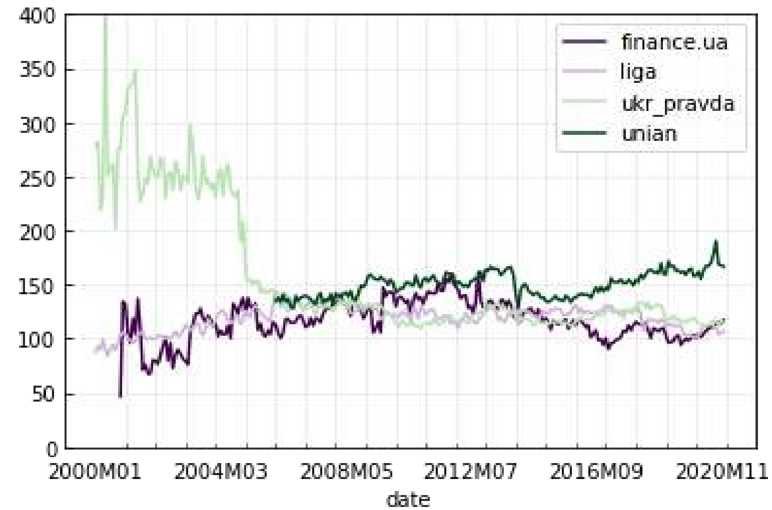
- The National Bank of Ukraine has been running surveys of inflation expectations for the next 12 months for several types of agents: households, banks, businesses, and professional forecasters.

Data: News corpus

Number of news per month



Average size of articles per month



- Our news corpus consists of more than 2 000 000 articles published online covering a sample of 20 years from January 2000 to December 2020.

Constructing aggregate news indexes

- All measurement methods that are based on text-mining can be divided into in two groups:

  1) naïve methods
  2) more complex methods, which are based on machine learning (LDA, BERT etc.)

- News content in our corpus is related mainly to economic, social and political topics

- We do not consider the sentiments of news content in this research

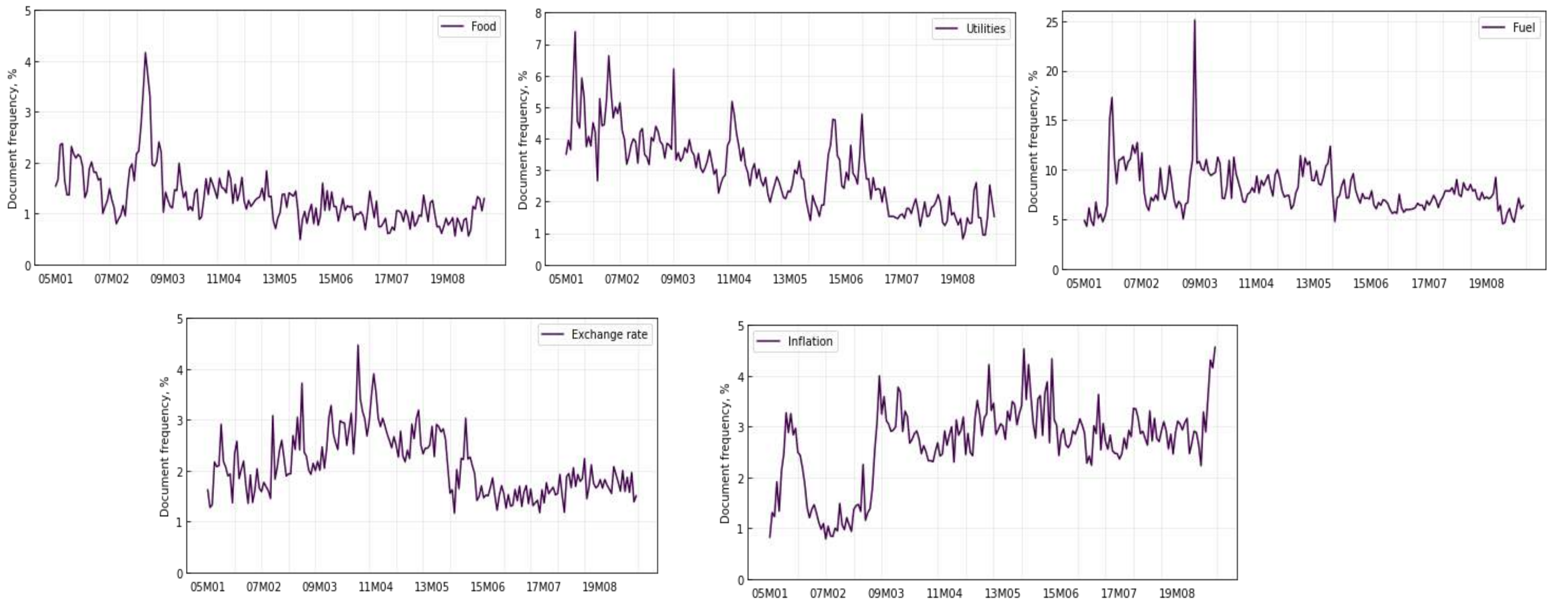
- News indices are calculated as a share of articles related to the topic or more common denotation "document frequency".

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

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

Dictionary-based approach

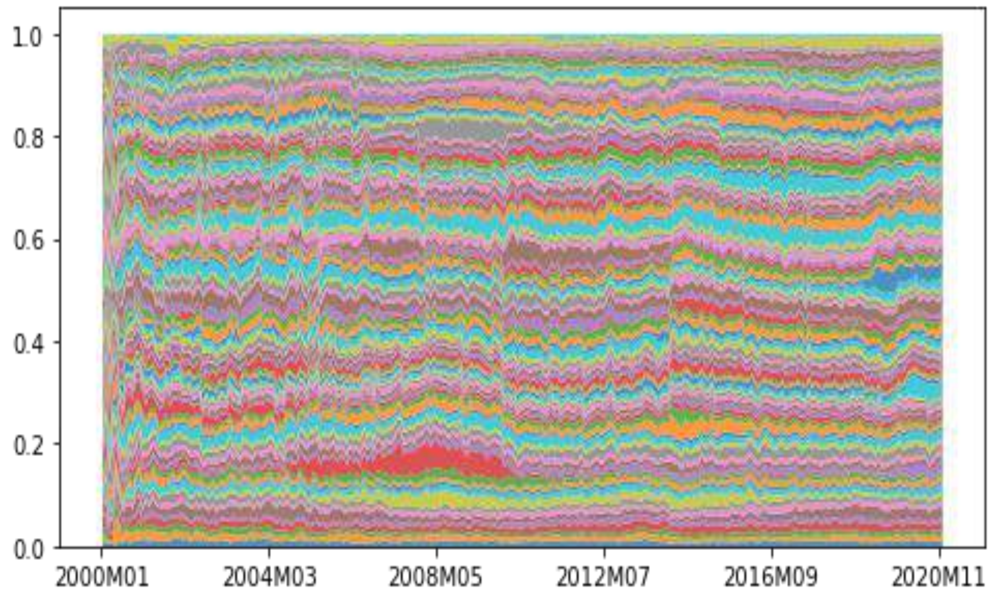
Document frequency of topics relevant to inflation expectations



- The dictionary-based approach is the simplest approach to estimating the impact of news on various macroeconomic indicators.
- Dictionary-based approach to constructing news indices requires good expertise aiming to select relevant keywords. In this case, to determine which prices worry Ukrainians the most, we turn to the consumer basket of the average household.

Unsupervised ML approach

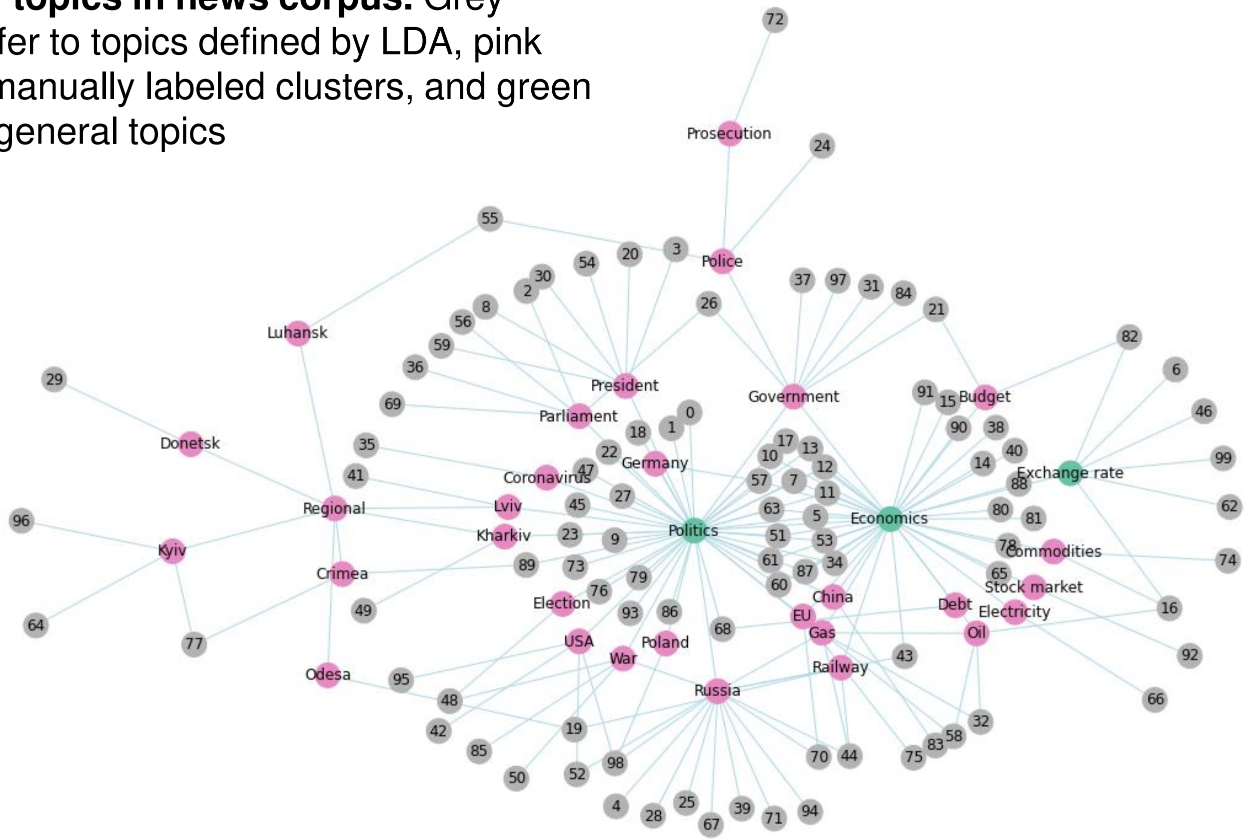
Distribution of topics received with LDA



- 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, the article may contain keywords, but its topic is devoted to a completely different issue.
- The solution here is 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.
- Latent Dirichlet Allocation (LDA) presented by Blei et al. in 2003 nowadays is a very common example of a topic modeling method. The LDA model is an unsupervised learning algorithm that does not need labeled training samples.

Unsupervised ML approach

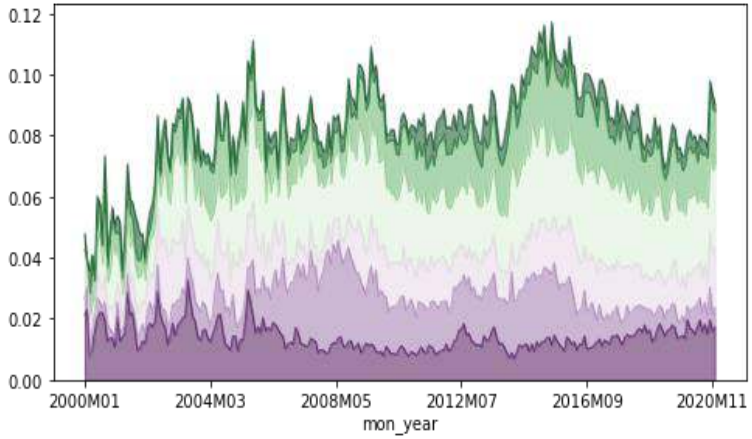
Graph of topics in news corpus. Grey circles refer to topics defined by LDA, pink circles - manually labeled clusters, and green circles – general topics



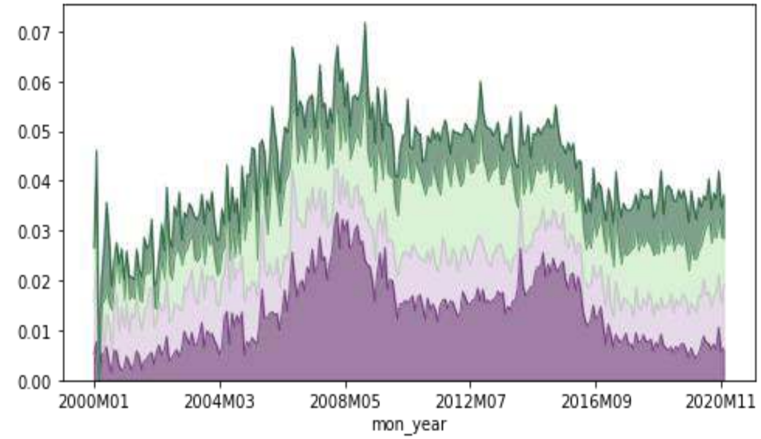
- Most of the news clusters are expectedly attributed to the politics, international relations, parliament and government.
- At the same time, some topics are clearly different and can be referred to the economic topics that may affect inflation expectations. Most news topics do not belong to one but to several clusters.

Unsupervised ML approach

Share of topics related to exchange rate movements , identified by LDA



Share of topics related to oil and gas, identified by LDA



- We built the indexes on the same principle as in the dictionary-based approach (document frequency).
- The popularity of certain topics highly corresponds to the historical development of events

Relationship between news indices and inflation expectations

Formation of inflation expectations

$$E\pi_t = \alpha + \beta\pi_{t-1} + \gamma(\pi_{t-1} - \pi_{t-2}) + \varepsilon \;\longrightarrow\; E\pi_t = \alpha + \beta\pi_{t-1} + \delta df_T^m + \varepsilon$$

Changes of inflation expectations

$$E\pi_t - E\pi_{t-1} = \\ = \alpha + \gamma(\pi_{t-1} - \pi_{t-2}) + \varepsilon \;\longrightarrow\; E\pi_t - E\pi_{t-1} = \\ = \alpha + \gamma(\pi_{t-1} - \pi_{t-2}) + \eta(df_T^m$$

where $E\pi_t$ is expected inflation in period t, π_{t-1} denotes inflation in previous period, and $\pi_{t-1} - \pi_{t-2}$ stands for change in inflation, α, β, γ, η – coefficients of regression, while ε is an error, df denotes document frequency of the news topic m in period T

- I used an extrapolative approach to the formation of inflation expectations
- Annual CPI change is used as a measure of inflation.

Estimation results

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)***	0.289 ***(0.000)***	0.2994 ***(0.000)***	0.297 ***(0.000)***	0.295 ***(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}	1.336 (0.153)	-0.7157 (0.591)	-1.761 (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)*	-0.972 **(0.026)**
	df^m_{t-3}	-1.074 (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	0.843	0.825	0.84	0.837	0.845	R^2	0.57	0.334	0.372	0.34	0.413
	p-value	***(0.000)***	***(0.000)***	***(0.000)***	***(0.000)***	***(0.000)***	p-value	***(0.000)***	**(0.030)**	**(0.015)**	**(0.027)**	***(0.006)***

Respondents	Variables	Without transformations					Variables	1st difference				
		Inflation	Exchange rate	Utilities	Food	Fuel		Inflation	Exchange rate	Utilities	Food	Fuel
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)***	0.14 ***(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}	3.107 **(0.026)**	0.0296 (0.981)	-0.327 (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)	0.375 (0.584)	3.692 **(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)***	-0.137 (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	0.696	0.685	0.746	0.736	0.687	R^2	0.227	0.179	0.219	0.274	0.188
	p-value	***(0.000)***	***(0.000)***	***(0.000)***	***(0.000)***	***(0.000)***	p-value	**(0.016)**	*(0.058)*	**(0.020)**	***(0.004)***	**(0.045)**

Notes: The table shows results of OLS regressions where inflation expectations are dependent variable. Time indicator T of document frequencies is set to t, t-1, t-2 and t-3. Curr. stands for decades of reported month, Prev. – for previous month. Indicators I, II and III following document frequency indices denote number of decades. First figures in cells indicate coefficients, p-values are shown in parentheses (1%, 5% and 10% significance levels highlighted by underlined bold italic (blue), bold (green) and italic (orange) respectively).

Estimation results

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	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)	0.923 (0.239)	0.229 (0.283)
	df^m_{t-1}	-0.109 (0.905)	0.4006 (0.773)	0.5428 (0.469)	1.8611 (0.253)	0.1923 (0.691)	$df^m_{t-1} - df^m_{t-2}$	-0.592 (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)*
	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	0.61	0.739	0.669	0.623	R^2	0.107	0.107	0.059	0.09	0.068
	p-value	***(0.000)***	***(0.000)***	***(0.000)***	***(0.000)***	***(0.000)***	p-value	(0.144)	(0.144)	(0.489)	(0.236)	(0.400)

Respondents	Variables	Without transformations					Variables	1st difference				
		Inflation	Exchange rate	Utilities	Food	Fuel		Inflation	Exchange rate	Utilities	Food	Fuel
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)***	$\pi_{t-1} - \pi_{t-2}$	-0.0152 (0.696)	-0.0159 (0.668)	-0.0117 (0.752)	-0.003 (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)*	0.333 (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)	-0.093 (0.667)
	df^m_{t-2}	0.7433 (0.257)	-0.297 (0.765)	0.0352 (0.939)	1.405 (0.259)	0.2002 (0.628)	$df^m_{t-2} - df^m_{t-3}$	-0.832 **(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)*	2.324 **(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)	-0.769 (0.523)
	R^2	0.675	0.629	0.838	0.66	0.598	R^2	0.164	0.175	0.148	0.023	0.01
	p-value	***(0.000)***	***(0.000)***	***(0.000)***	***(0.000)***	***(0.000)***	p-value	**(0.036)**	**(0.026)**	*(0.058)*	(0.904)	(0.984)

Notes: The table shows results of OLS regressions where inflation expectations are dependent variable. Time indicator T of document frequencies is set to t, t-1, t-2 and t-3. First figures in cells indicate coefficients, p-values are shown in parentheses (1%, 5% and 10% significance levels highlighted by underlined bold italic (blue), bold (green) and italic (orange) respectively).

Estimation results. LDA indices

Topic	Variables	Without transformations: Banks	Businesses	Households	Financial analysts	Variables	1st difference: 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)	1.08 (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)	0.414 (0.455)
	df^m_{t-3}	-0.5554 (0.758)	-2.0173 (0.112)	-1.803 (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	0.846	0.685	0.718	0.621	R^2	0.276	0.215	0.072	0.054
	p-value	***(0.000)***	***(0.000)***	***(0.000)***	***(0.000)***	p-value	*(0.081)*	**(0.022)**	(0.369)	(0.592)
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)	-0.214 (0.608)	$df^m_{t-2} - df^m_{t-3}$	0.5198 (0.529)	0.4751 (0.509)	0.5903 *(0.088)*	-0.702 **(0.025)**
	df^m_{t-3}	-0.481 (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)	-0.229 (0.938)	-0.3152 (0.800)	0.0642 (0.953)
	R^2	0.864	0.756	0.821	0.809	R^2	0.397	0.217	0.147	0.102
	p-value	***(0.000)***	***(0.000)***	***(0.000)***	***(0.000)***	p-value	***(0.009)***	**(0.021)**	**(0.043)**	(0.209)

Notes: The table shows results of OLS regressions where inflation expectations are dependent variable. Time indicator T of document frequencies is set to t, t-1, t-2 and t-3. First figures in cells indicate coefficients, p-values are shown in parentheses (1%, 5% and 10% significance levels highlighted by underlined bold italic (blue), bold (green) and italic (orange) respectively).

National Bank of Ukraine

Concluding remarks

- Using natural language processing and machine learning techniques I cleaned and transformed textual data into news-based quantitative measures reflecting news topics relevant to inflation and inflation expectations.
- I applied two different approaches to filter out the news: a dictionary-based approach and a Latent Dirichlet Allocation (LDA).
- Different news topics have an impact on inflation expectations. For example, I found a strong relationship between inflation expectations of households and financial analysts with news about utilities, while businesses are sensitive to the news about food. Additionally, financial analysts and households have also been sensitive to levels and changes in the frequency of exchange rate news, constructed by LDA.
- As a result, we see that the formation of inflation expectations of different groups of respondents may depend on the media environment, namely both the volume of published articles and changes in this indicator. Different groups of respondents rely on different topics and different periods when estimating future inflation.
- Our results can help understand inflation expectations, especially as anchoring inflation expectations remains a key challenge for central banks. This may, among other things, be important for the central bank's communication policy and help to articulate clear and effective messages

Thank you!