Available online at www.sciencedirect.com

ScienceDirect

The Journal of Finance and Data Science 6 (2020) 31—48
http://www.keaipublishing.com/en/journals/jfds/

Should asset managers pay for economic research? A machine learning evaluation

Krzysztof Rybinski*

Vistula University, Stoklosy 3, 02-787, Warsaw, Poland

Received 12 June 2019; revised 7 August 2020; accepted 13 August 2020
Available online 18 August 2020

Abstract

This paper presents the first-ever comparison of the forecasting power of two types of narratives: articles in a major daily newspaper and regular research reports released by professional forecasters. The applied testing methodology developed in [22] and extended in this paper includes two natural language processing (NLP) techniques — the sentiment analysis and the wordscores model — that are used to convert the text corpora into the NLP indices. These indices are explanatory variables in linear regression, Granger causality test, vector autoregressive model and random forest model. The paper extends this methodology by applying Latent Dirichlet Allocation (LDA) to the newspaper corpus to filter out articles that discuss topics not relevant for economic and financial analysis. The forecasting test is conducted for two major banks in Poland — BZ WBK and mbank and for major daily newspaper Rzeczpospolita, in Polish. It is shown that mbank narratives have the best forecasting power, while BZ WBK and Rzeczpospolita trade second and third place depending on the model applied. In the vast majority of analyzed cases adding an NLP index to the model improves the forecast accuracy. The answer to the title question is — it depends. Before paying for economic research asset managers are advised to apply methods such as presented in this paper to evaluate whether sell-side research offers any forecasting value in comparison with a newspaper.

JEL codes: C54; E47; E52

Keywords: Economic research; Forecasting; Natural language processing; Sentiment analysis; Wordscores

1. Introduction: economic forecasting and natural language processing (NLP)

Many institutions provide economic and financial forecasts regularly. For example, Bloomberg and Reuters collect numerical predictions of professional forecasters — including banks' economist teams — and publish consensus forecasts ahead of the release of major economic data or before the central bank interest rate decision. Professional forecasters are rated based on the quality of their economic and financial forecasts. Best forecasters enjoy high bonuses and, in some cases, even become economic celebrities. Professional forecasting institutions charge a fee for

* https://www.linkedin.com/in/krzysztofrybinski/.
E-mail address: rybinski@rybinski.eu.
Peer review under responsibility of China Science Publishing & Media Ltd.

https://doi.org/10.1016/j.jfds.2020.08.001

their services. Financial Times[1] reported that banks charge asset managers on average 75,000 US dollars annually for access to banks' research, while independent forecasting institutions charge 40,000 US dollars annually. An annual premium subscription of Financial Times costs only 546 euro. This paper answers the question, should asset managers and business leaders pay for economic research, or reading major business newspapers, such as Financial Times, will suffice.

Professional forecasts are usually released in the form of short economic reports that provide narrative justification for a given prediction. For example, the bank research team can write "*we expect the central bank to raise interest rates by* 50 bps *at the next policy meeting amid the unexpected rise of core inflation and signals of increasing wage pressures in many sectors*". Such research report offers not only the numeric forecast (50 bps interest rate increase) but also a narrative, explaining this forecast. Often, daily economic research reports present analysis without specifying the numerical forecast value. For example, it can say "*recent sharp depreciation of the exchange rate may lead to the subsequent rise of consumer inflation in the coming months amid high dollarization of contracts and large import penetration*". Therefore, the narrative of regular economic reports offers a much richer context for business and financial decisions than the "naked" numerical forecasts. It is therefore not a surprise that the number of articles in refereed journals that apply natural language processing (NLP) to financial forecasting exploded from 0 to 5 in years 1998−2008 to just under 30 in 2013−2015 and almost 60 in 2016.[2] The number of citations of such papers was below 20 in years 2005−2011 and rose to 268 and 292 in the years 2017 and 2018, respectively.[3] This research "explosion" led to several review papers that discussed and summarized the literature:[3,4,2,5−10] and.[11] Several key themes, essential for this research effort emerge from these reviews. The LDA[1] models, sentiment analysis[2] and bag of words[3] models are among the most frequently used in the literature, only recently challenged by the rise of deep learning. The vast majority of studies cover the English language text corpora, and while interest in other languages is growing, it is still a tiny part of research in this area. Indices created with the help of the NLP models are used in linear regression or time series models, such as autoregressive moving average (ARIMA) or vector autoregression (VAR). Typically used data sources are very diverse, but there was not a single case of a corpus constructed on the basis of economic research reports produced regularly by a commercial bank or an investment bank economic, equity or fixed income research teams.[4] None of the review articles mentioned applications of the supervised wordscores model that are very often used in the field of political science and allow to position documents on a scale.

In light of the above review of the literature, this paper key contributions are as follows:

- Develops a machine learning methodology allowing to evaluate whether the narrative of the research reports created by professional forecasters helps to predict economic and financial variables in comparison with the texts of newspaper articles.
- Conducts the first ever-published such an evaluation and concludes that some professional forecasters are better than journalists, and some are not.
- Compares the forecasting power of sentiment-based and textscore-based indices and shows that is some tests the rarely used in economic research wordscores algorithm can give better results than trendy sentiment analysis.
- Presented results justify using the developed methodology to improve asset managers' and business leaders' ability to determine which professional forecasters research reports should be read and which should be avoided.
- The analysis is conducted for economic and financial texts in the Polish language, and as shown above, such analysis in languages not spoken in the world major financial centers is rare or non-existent.
- A new data set containing three corpora was created and will be made available for interested researchers. Such data sets are rare, especially in non-core research languages.

[1] LDA - Latent Dirichlet Allocation method.[21,22] LDA is an example of an unsupervised learning algorithm and performs clustering of texts in the corpus based on the likelihood of their major topic.

[2] Sentiment analysis refers to assigning a number to a text, and this number represent the text sentiment. The sentiment can be positive, neutral, or negative on a scale defined by all texts in the corpus.

[3] Bag of words refers to a class of NLP models where text is cleaned and converted to a set (bag) of words. In such representation the order of words in the text does not matter.

[4] Such research is often referred to as the sell-side research, as it is used by banks and brokerage houses to attract attention of their clients, such as asset managers (buy-side), corporations or wealthy private individuals.

Table 1
Description of the monthly economic time series that are analyzed in the paper and their stationarity tests.

Symbol	Name of the series	Type of series	ADF	KPSS
Ip	industrial production	monthly, %yoy	0.097 •	0.1
Rs	retail sales	monthly, %yoy	0.55	0.01**
cpi	Consumer price inflation	monthly, %yoy	0.602	0.014**
ppi	producer price inflation	monthly, %yoy	0.566	0.08 •
pmi	purchasing managers' index	monthly, index	0.070 •	0.1
U	unemployment rate	monthly, percentage	0.809	0.013**
W	Wages	monthly, %yoy	0.552	0.032**
ms	money supply (M3)	monthly, %yoy	0.266	0.01**
Cr	Credit	monthly, %yoy	0.173	0.01**
usd	USD/PLN exchange rate	monthly average, zloty per U.S. dollar	0.021*	0.01*
eur	EUR/PLN exchange rate	monthly average, zloty per euro	0.013*	0.01**
I	main interest rate of the central bank	end of the month, percentage points	0.028*	0.01**
W3m	WIBOR, 3-month money market borrowing rate	monthly average, percentage points	0.023*	0.01**
i2y	Yield of 2-year government bonds	monthly average, percentage points	0.42	0.01**
Oil	Price of the barrel of oil in usd	monthly, in usd	0.309	0.073 •
r3m	Real 3-month interest rate (nominal − CPI)	monthly, in percentage points	0.309	0.01**
r2y	real 2-year interest rate (nominal − CPI)	monthly, in percentage points	0.627	0.079 •

Note: Last two columns presents the *p*-values of ADF and KPSS stationarity tests. In ADF test the null hypothesis states that series is non-stationary, in KPSS test the null hypothesis states that the series in stationary.
**Denotes statistical significance at 0.01 level, *At 0.05 level, • at 0.1 level. The NLP indices are stationary by construction.

Source: Central Statistical Office, NBP.

The rest of this text is organized as follows. Section 2 presents the data and methodology. Section 3 shows the results of applying the LDA model to filtering the newspaper articles. In section 4 we construct sentiment and textscores' indices for all three corpora and analyze their properties. Section 5 presents the analysis of the contemporaneous relationships between the economic/financial time series and the NLP indices. Section 6 verifies whether the NLP indices have predictive power with respect to economic/financial time series. Section 7 discusses the results, concludes, and indicates areas for further research.

2. Methodology and data description

We use five sources of data covering the period of December 2006−January 2018. We collected 39 statements of the Monetary Policy Council of the National Bank of Poland (MPC after that) released on the day of the change in the central bank's main interest rate. MPC statements accompanying no interest rate change meetings were ignored. We scraped from the BZ WBK bank website[5] its daily economic research reports called "*Codziennik*". These are 3−4 pagers discussing political, economic, and financial developments globally and in Poland that occurred in the past 24 h and making short term predictions for events and data releases for the next 24 h. There were 2741 BZ WBK daily reports in the analyzed period. We also scraped daily economic research reports called *"Raport dzienny"*, published by mbank[6] and having similar structure and coverage, collecting 2745 reports in the analyzed period. BZ WBK[7] is the third largest, and mbank the fifth-largest bank in Poland, and their research teams often take leading positions in forecasting competitions.

We also scraped from the Rzeczpospolita daily newspaper's archive[8] articles that contained one of the following phrases and words (in all declinations, including abbreviations and acronyms[9]): National Bank of Poland, Monetary Policy Council, unemployment, budget, inflation, output, GDP, recession, uncertainty. There were 54,489 such articles published from December 2006 to January 2018. All texts were in the Polish language. Finally, we collected monthly

5 https://skarb.santander.pl/skarb/serwis-ekonomiczny/codziennik/codziennik.html.
6 https://www.mbank.pl/serwis-ekonomiczny/raporty/komentarz-dzienny.html.
7 In September 2018 BZ WBK was renamed to Santander Polska after its parent institution. This paper uses the bank former name.
8 https://archiwum.rp.pl.
9 We used regular expressions and 1- and 3-g to find all declination forms.

economic and financial data released by the Polish Central Statistical Office, and the NBP, described in Table 1 below, that also shows their stationarity tests.

We proceeded as follows. Firstly, we conducted a standard text cleaning applied in the "bag of words" models for all four text corpora (NBP statements, BZ WBK and mbank daily economic reports, and Rzeczpospolita articles). We removed stop words[10], punctuation, capital letters, white spaces, and digits. Additionally, we used regular expressions to remove lengthy disclaimers from all daily bank reports. We also used a large Polish language lexicon containing more than 3.3 million word−lemma pairs to replace words with lemmas.[11] Finally, we removed very frequent words and created the document-feature matrices (DFMs).[12] We then applied the LDA model to the Rzeczpospolita newspaper corpus to remove the irrelevant articles. Each document is assumed to be characterized by a set of topics, and the LDA model assigns each document to the most likely topic based on calculated keywords. We named the LDA-identified topics based on the top 20 words identifying each LDA topic. For further analysis, we kept only articles assigned to topics that are relevant for economic and financial analysis. Hence, the final size of the analyzed newspaper corpus was 34,794 articles.

In the next step, we applied sentiment analysis to the lemmatized texts using plWordNet-emo dictionary containing close to 30,000 words, which were manually annotated with sentiment polarity: strongly negative, moderately negative, neutral, moderately positive and strongly positive.[12] Words with strongly negative sentiment polarity were assigned "-2" score, moderately negative "-1" score, neutral zero score, moderately positive "+1" score, and strongly positive "+2" score. Then the sentiment algorithm calculated how many words (lemmas) with identified sentiment polarity appeared in each text. Next, it applied the assigned scores, and calculated the total sentiment score for each Rzeczpospolita article, BZ WBK daily economic report, and mbank daily economic report. The final step was a calculation of monthly averages to obtain monthly sentiment indices for Rzeczpospolita articles and BZ WBK and mbank daily reports.

In this paper, we also use another type of the NLP index, which is an index based on the wordscores model.[13] Interestingly, while the wordscores model is used very often in the political science text mining literature, its applications in economic papers are scarce. It is surprising, as the wordscores model allows for the automatic generation of economically meaningful labels and, consequently, NLP indices that can later be used for econometric forecasting.[13] used the wordscores model to analyze the consistency of ECB communication between 1999 and 2009.[14] applied the wordscores models to 2-g to estimate the forward guidance of major central banks.[15] implemented the wordscores approach to introductory statements of ECB's Governing Council press conferences to estimate a 'shadow prime rate' and its influence on the EURO-STOXX-Banks Future.[16] trained the wordscores model as one of the inputs into the ranking of professional forecasters based on the narratives of their reports.

The final text-processing step was a calculation of the monetary stance indicator for each Rzeczpospolita article, BZ WBK and mbank daily reports, from monetary easing to monetary tightening on a dovish/hawkish scale. We applied the following method introduced in[13] and.[17] In the analyzed period of December 2006−January 2018, 39 decisions changed the main interest rate, 16 decisions increased interested rate, 23 lowered it. Initially, each word (lemma) in every policy statement received the word score of +1 (rate hike) or −1 (rate cut). Then the wordscores algorithm (from R quanteda library) used MPC statements to calculate frequency-weighted scores for words that appeared in these statements and applied these scores to newspaper articles and banks' daily reports. Intuitively, the textscore is a frequency-weighted average of scores of words that appear in the text. If a word appears in both hawkish and dovish policy statements, its score is calculated using relative frequency of appearances. So, for example if a word appeared once in a hawkish policy statement and three times in a dovish statement its score would be (1*(+1) + 3*(-

[10] Stop words are words that appear very often but carry little meaning, such as "the", "and" or "with" in English. We used the list of stopwords in the Polish language available in the lsa library in R (277 words). The library is described here: https://cran.r-project.org/web/packages/lsa/lsa.pdf.

[11] Many text mining papers use stemming to prepare texts for automatic analysis. Because for the Polish language we have a freely available large lemmatizer, instead of stemming we applied lemmatization. The lemmatizer at the time of writing this paper was available at www.lexiconista.com/datasets/lemmatization/. This lemmatizer has been used in several papers published in high impact factor journals. The author can provide the lemmatizer to interested researchers. Additionally, in the case of BZ WBK and mbank daily reports we also removed the names of months, to avoid clustering the reports according to the month of their release.

[12] DFM is a bag-of-words representation of the corpus of texts. Rows in this matrix represent documents and columns represent features (words after cleaning and lemmatizing). An integer value in the cell (i,j) of the DFM indicates how many times j_{th} feature appears in the i_{th} document.

[13] The wordscores model was introduced in.[23] A detailed description of the wordscores algorithm and other text mining techniques can be found in.[24]

1))/4 = - 0.5. Words in newspaper articles and banks' reports that do not appear in any of the MPC policy statements are not used in the textscore calculations.

More formally, wordscores computes the estimated score for a virgin document θ_d (a newspaper article or a bank research report in this paper) as the average of the scores π_w of all the words in the document. Scores are computed from a collection of the reference documents (MPC statements in this paper). When V types of words appear in this collection, and there are W words in d, the textscore becomes[18]:

$$\widehat{\theta}_d = \frac{1}{W}\sum_{w}^{W}\widehat{\pi}_w = \sum_{j}^{V}\widehat{\pi}_j\widehat{P}(w_j|d)$$

where

$$\widehat{P}(w_j|d) = \frac{c(w_j\,in\,d)}{c(d)}$$

and $c(\cdot)$ is the word counting function.

We constructed six NLP monthly indices[14]: sentiment-based and textscore-based for Rzeczpospolita newspaper articles, for BZ WBK and mbank daily economic reports. Then we applied econometric and machine learning techniques to test the predictive power of these indices for selected economic and financial time series. We made sure the analyzed series are stationary by applying first differences, analyzed correlation patterns, applied linear regressions, and Granger causality tests. We also run a machine learning experiment verifying whether VAR models using information from NLP indices exhibit smaller forecast errors than univariate ARIMA models not using such information. We repeated this experiment in a supervised learning framework using random forest models.

3. The results of the LDA analysis

After running the LDA algorithm for Rzeczpospolita corpus for a broad selection of hyper-parameters, we settled for twenty LDA topics. The number of topics closer to ten generated too general and closer to thirty too detailed lists of topics. Other hyperparameters were set to default values of the LDA algorithm in the R topicmodels library. From Rzeczpospolita newspaper we selected articles that mentioned the central bank or ones with keywords typical for economic and financial analysis. But even after applying such a content filter, the keywords describing LDA topics showed that the range of topics remained very wide, as shown in Tables 2 and 3 below. Each table presents the top 10 keywords with the highest likelihood of belonging to a given LDA topic, together with the number of texts allocated to topics. There are ten LDA clusters of articles in Rzeczpospolita that can be identified as relevant for our analysis. We labelled them as: "stock exchange", "firms-investment", "international affairs", "finance", "politics", "public finance", "Poland-EU", "inflation-economy", "privatization", "stocks-bonds-funds".

While all the previous steps were automated, the topic naming and selection of clusters for further analysis were performed by the author, are subject to the author's discretion, and therefore should be justified. Because the NLP indices based on the newspaper corpus will be used to forecast economic and financial variables, some choices were straightforward, such as: "stock exchange", "firms-investment", "finance", "public finance", "inflation-economy" and "stocks-bonds-funds". The topic "international affairs" was included as it contains articles discussing Russia or war in Ukraine, in the economic context. Many past events originating in Russia affected Polish assets risk premia (e.g., Smolensk crash, Crimea war). A similar argument holds for articles about Poland and the E.U. Polish politics many times proved to be very important in shaping economic scenarios. Maybe more controversial is the choice of the privatization topic. In Western countries, privatization has a negligible impact on the economy. Still, in Poland, privatization revenues were an important part of the budget deficit financing, provided a litmus test of the political elite's willingness to reform the country and, at times, moved the exchange rate. For this reason, privatization-related articles were also included in the further analysis.

Most topics that were removed do not require a detailed explanation. There were some doubts regarding three topics, and the decision was made by looking at the sample of documents. The "automotive-finance" articles discuss

[14] Monthly index is a simple average of the scores of all texts that appeared in a given month.

Table 2
Topics 1–10 out of twenty topics generated by unsupervised LDA algorithm for Rzeczpospolita daily newspaper articles, number of articles and top 10 keywords.

Topic name	stock exchange	regulations	energy sector	Firms/ investment	international affairs	leisure	finanse	automotive/ finanse	politics	labour/education
No of articles	3343	2576	3068	3440	3364	3462	2370	1089	3696	2645
Top ten terms	akcja *(share)*	ustawa *(law, act)*	milion *(million)*	firma *(firm, company)*	kraj *(country, nation)*	polska *(Poland)*	bank *(bank)*	waluta *(currency)*	pis *(PiS party)*	praca *(job, work)*
	firma *(firm)*	podatkowy *(tax related)*	miliard *(billion)*	projekt *(project)*	rosja *(Russia)*	polski *(Polish)*	kredyt *(credit)*	euro *(euro)*	partia *(political party)*	osoba *(person)*
	proc *(percent)*	podatek *(tax)*	inwestycja *(investment)*	program *(programme)*	rosyjski *(Russian)*	film *(film, movie)*	finansowy *(financial)*	kurs *(exchange rate)*	polityczny *(political)*	firma *(firm)*
	milion *(million)*	podatnik *(taxpayer)*	energia *(energy)*	milion *(million)*	wielki *(great, big)*	wielki *(great, big)*	bankowy *(banking)*	pojazd *(vehicle)*	polityk *(politician)*	pracownik *(employee)*
	euro *(euro)*	vat *(VAT)*	proc *(percent)*	nowa *(new)*	usa *(USA)*	klub *(club)*	nbp *(NBP, central bank)*	auto *(auto, car)*	polska *(Poland)*	polska *(Poland)*
	usa *(USA)*	przepis *(regulation)*	polski *(Polish)*	polski *(Polish)*	ukraina *(Ukraine)*	mecz *(match, game)*	kredytowy *(credit related)*	nbp *(NBP, central bank)*	polityka *(politics)*	uczelnia *(university)*
	dywidenda *(dividend)*	towar *(good, product)*	energetyczny *(energy sector)*	rynek *(market)*	wojna *(war)*	lato *(summer)*	klient *(client)*	koszt *(cost)*	sprawa *(case)*	profesor *(professor)*
	miliard *(billion)*	koszt *(cost)*	polska *(Poland)*	przedsiębiorca *(entrepreneur)*	sprawa *(case)*	historia *(history, fairytale)*	instytucja *(institution)*	kursowy *(currency related)*	polski *(Polish)*	nauka *(science)*
	sprawa *(case)*	punkt *(item)*	gaz *(gas)*	klient *(client)*	prezydent *(president)*	dziecko *(child)*	depozyt *(deposit)*	cena *(price)*	prezydent *(president)*	polski *(Polish)*
	wynik *(result)*	kwota *(amount)*	firma *(firm, company)*	produkt *(product)*	europejski *(European)*	ostatni *(last)*	rynek *(market)*	obcy *(foreign, alien)*	sejm *(parliament, lower house)*	zmiana *(change)*

Table 3
Topics 11−20 out of twenty topics generated by unsupervised LDA algorithm for Rzeczpospolita daily newspaper articles, number of articles and top 10 keywords.

Topic name	public finance	Poland/EU	labour/social security	health	inflation/economy	privatisation	real estate	stocks/bonds/funds	regions/cities	law/judiciary
No of articles	4345	3458	2146	886	4416	1586	1561	4776	3086	3648
Top ten terms	miliard *(billion)*	kraj *(country)*	praca *(job, work)*	zdrowie *(health)*	proc *(percent)*	sprawa *(case)*	lokal *(premise, flat)*	rynek *(market)*	miasto *(city)*	ustawa *(law, act)*
	finanse *(finance)*	euro *(euro)*	osoba *(person)*	szpital *(hospital)*	wzrost *(growth)*	minister *(minister)*	budynek *(building)*	proc *(percent)*	milion *(million)*	przepis *(regulation)*
	publiczny *(public)*	polska *(Poland)*	pracownik *(employee)*	pacjent *(pacient)*	cena *(price)*	poseł *(deputy, MP)*	cena *(price)*	inwestor *(investor)*	gmina *(county)*	prawo *(legislation)*
	rząd *(government)*	polski *(Polish)*	ubezpieczenie *(social security)*	Lek *(medicine)*	procentowy *(percent)*	rada *(council)*	mkw *(square meter)*	akcja *(share)*	mieszkaniec *(inhabitant)*	sprawa *(case)*
	wydatek *(expenditure)*	europejski *(European)*	emerytura *(pension)*	lekarz *(doctor)*	inflacja *(inflation)*	komisja *(commision)*	dom *(house)*	obligacja *(bond)*	inwestycja *(investment)*	prawny *(legal)*
	Proc *(percent)*	gospodarka *(economy)*	zus *(state social security)*	Proc *(percent)*	stopa *(rate)*	prezes *(chairperson)*	miasto *(city)*	fundusz *(fund)*	droga *(road)*	umowa *(agreement)*
	podatek *(tax)*	strefa *(zone)*	emerytalny *(pension related)*	nfz *(health fund)*	punkt *(percentage point)*	skarb *(state treasury)*	powierzchnia *(area, space)*	papier *(instrument)*	miejski *(city related)*	publiczny *(public)*
	zmiana *(change)*	Unia *(Union)*	proc *(percent)*	medyczny *(medical)*	poziom *(level)*	szef *(chief)*	mieszkaniowy *(housing)*	cena *(price)*	warszawa *(Warsaw)*	zasada *(principle)*
	pkb *(GDP)*	gospodarczy *(economic)*	wynagrodzenie *(wage)*	zdrowotny *(health related)*	pkb *(GDP)*	informacja *(information)*	nowa *(new)*	dolar *(dollar)*	dotacja *(subsidy, grant)*	zmiana *(change)*
	reforma *(reform)*	rynek *(market)*	dziecko *(child)*	leczenie *(treatment)*	gospodarka *(economy)*	prokuratura *(prosecutor)*	budowa *(construction)*	zysk *(profit)*	lokalny *(local)*	sąd *(court)*

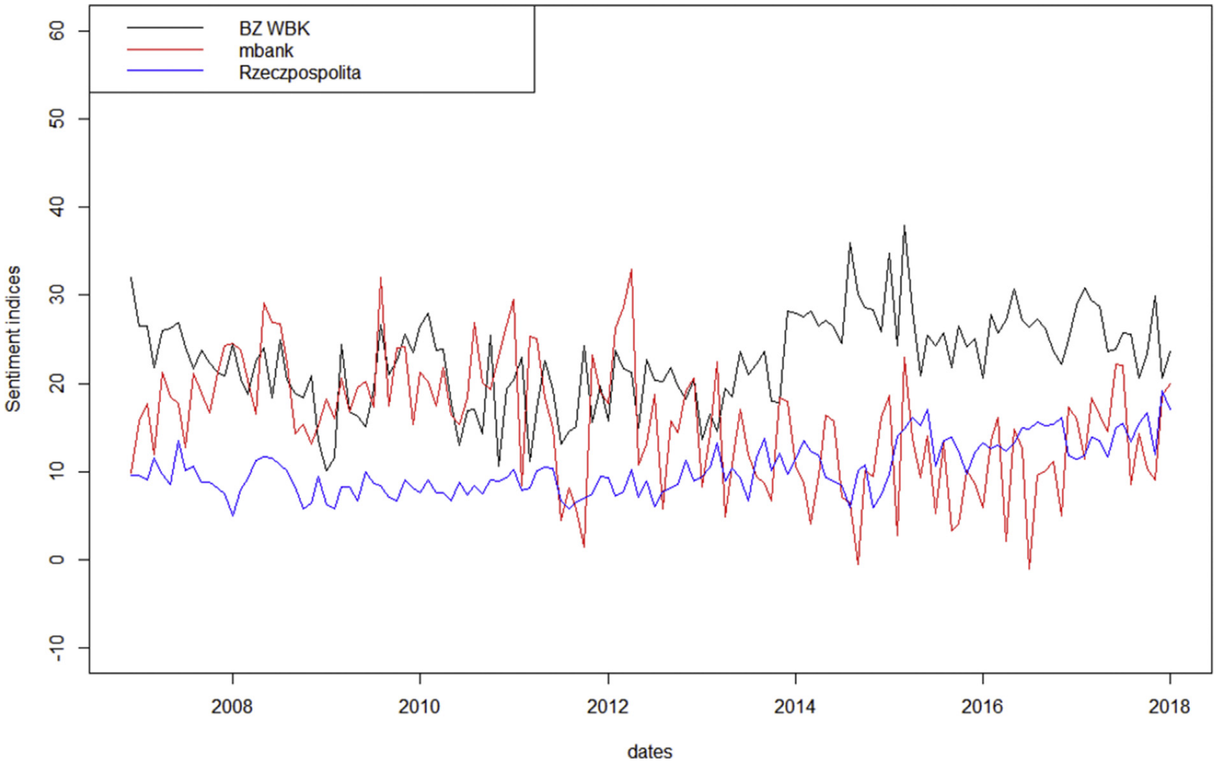

Fig. 1. Sentiment scores for BZ WBK daily reports, mbank daily reports and Rzeczpospolita articles, monthly averages.

mostly personal finance issues, while labor/education articles talk about employee−company relationships (and university-professor in particular) rather than about unemployment. The "regions-cities" cluster talks about local investments, such as road construction and financing these projects by the E.U. structural funds.

We scraped articles from all sections of the Rzeczpospolita daily. The newspaper has a large legal part, that is why the LDA algorithm identified many topics that are related to regulatory or contractual developments in such areas as social security, health, or real estate. The real estate topic would have been relevant in the U.S. Still, in Poland, it is not discussed in the context of an asset bubble. It instead represents a regular reporting of the housing space built, its quality, availability of the government support programs, bank financing conditions, and changing housing prices.

Finally, one can also verify whether the topics were correctly labeled based on the top keywords. While there is some discretion involved, we did not have any significant problems with the proper topic identification. As indicated earlier, the LDA-based filtering reduced the size of the Rzeczpospolita newspaper corpus from 54,489 to 34,794 articles.

4. Properties of sentiment and textscore indices

We defined the NLP indices as follows: "bz" − refers to BZ WBK, "mb" − refers to mbank, and "rp" − denotes Rzeczpospolita; "s" stands for sentiment and "w" indicates textscores computed by the wordscores model. In this section, we examine both the daily and the monthly indices. Monthly indices are calculated as a simple average of the daily index.

Fig. 1 presents monthly sentiment scores for Rzeczpospolita articles after the LDA-filtering, BZ WBK daily reports, and mbank daily reports. It is worth noting that the sentiment for all months is positive, so words with positive sentiment polarity are used much more often than negative words. It is not a feature of the Polish language. There are more words with negative polarity than with positive polarity in the sentiment lexicon. Also, the political sentiment analysis in Poland[19] and[20] showed that often texts that mention the names of politicians have a negative sentiment. The sentiment of the bank reports is above that of the newspaper, although the mbank sentiment index significantly

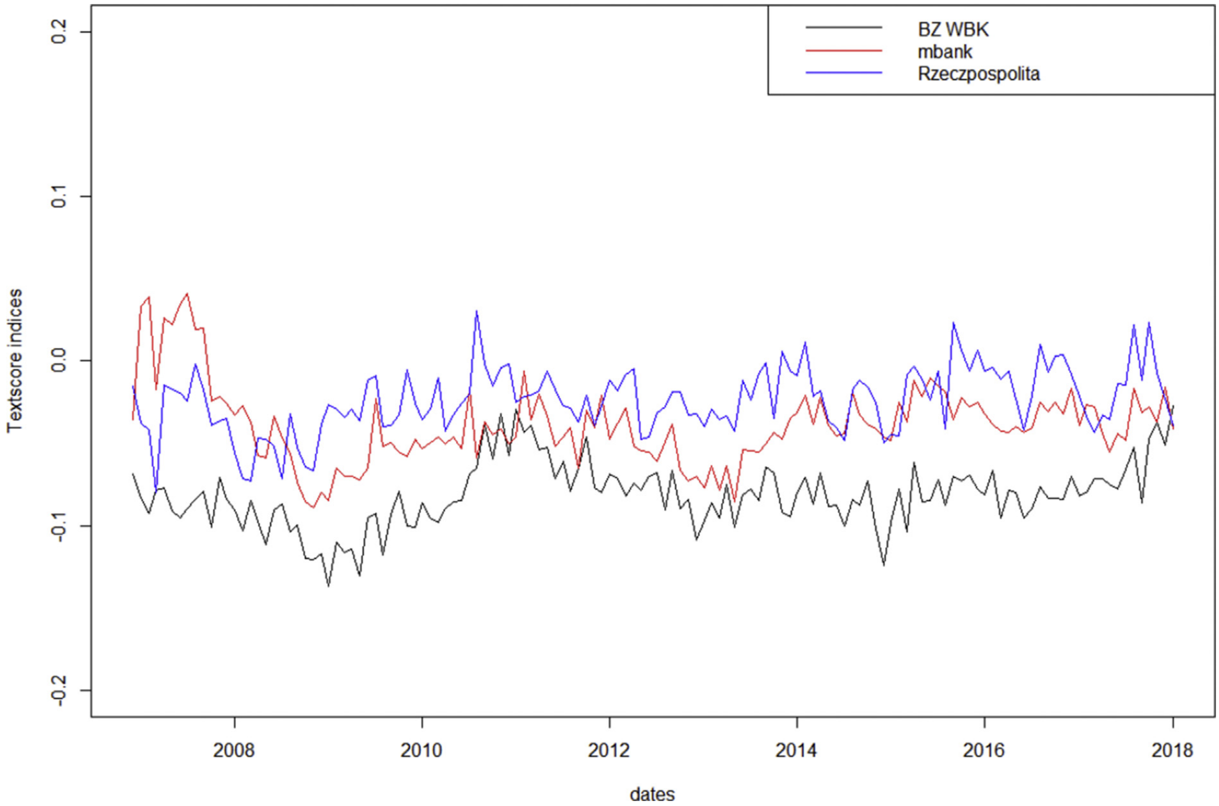

Fig. 2. Wordscores algorithm scale for BZ WBK daily reports, mbank daily reports and Rzeczpospolita articles, monthly averages.

declined after 2013 and in 2016 fell below the Rzeczpospolita sentiment index. While the sentiment indices for both banks exhibit a drop in 2009 (global recession) and 2011 (eurozone crisis), the Rzeczpospolita sentiment index remains relatively stable. It stems from the fact that to calculate monthly sentiment indices, we use, on average, 20 bank daily research reports, that have a similar regular format. In contrast, in the case of Rzeczpospolita, we average sentiment scores for some 226 articles covering different topics, have different lengths and various authors.

There is one more interesting observation. While the sentiment indices for both banks oscillated in a similar range in the 2008—2014 period, they diverged after 2014, with BZ WBK reports exhibiting much more optimism, and mbank sentiment index falling below its 2008—2009 levels. It is unclear why this has happened. The format of the reports remained the same, and chief economists in both banks did not change. It appears that two teams of

Table 4
Correlations between daily and monthly NLP indices calculated for Rzeczpospolita, BZ WBK and mbank.

Daily indices, 95% conf. interval (-0.037, 0.037)						
	bzs	**mbs**	**rps**	**bzw**	**mbw**	**rpw**
bzs	1.000	0.091	0.111	-0.050	0.020	0.045
mbs	0.091	1.000	-0.016	-0.029	-0.220	-0.032
rps	0.111	-0.016	1.000	0.034	0.049	0.162
bzw	-0.050	-0.029	0.034	1.000	0.243	0.096
mbw	0.020	-0.220	0.049	0.243	1.000	0.097
bpw	0.045	-0.032	0.162	0.096	0.097	1.000

Monthly indices, 95% conf. interval (-0.17, 0.17)						
	bzs	**mbs**	**rps**	**bzw**	**mbw**	**rpw**
bzs	1.000	-0.174	0.349	0.000	0.379	0.085
mbs	-0.174	1.000	-0.203	-0.089	-0.075	-0.136
rps	0.349	-0.203	1.000	0.251	0.252	0.322
bzw	0.000	-0.089	0.251	1.000	0.301	0.429
mbw	0.379	-0.075	0.252	0.301	1.000	0.187
rpw	0.085	-0.136	0.322	0.429	0.187	1.000

Note: red color indicates negative and blue color positive correlation. Color intensity reflects the correlation strength.

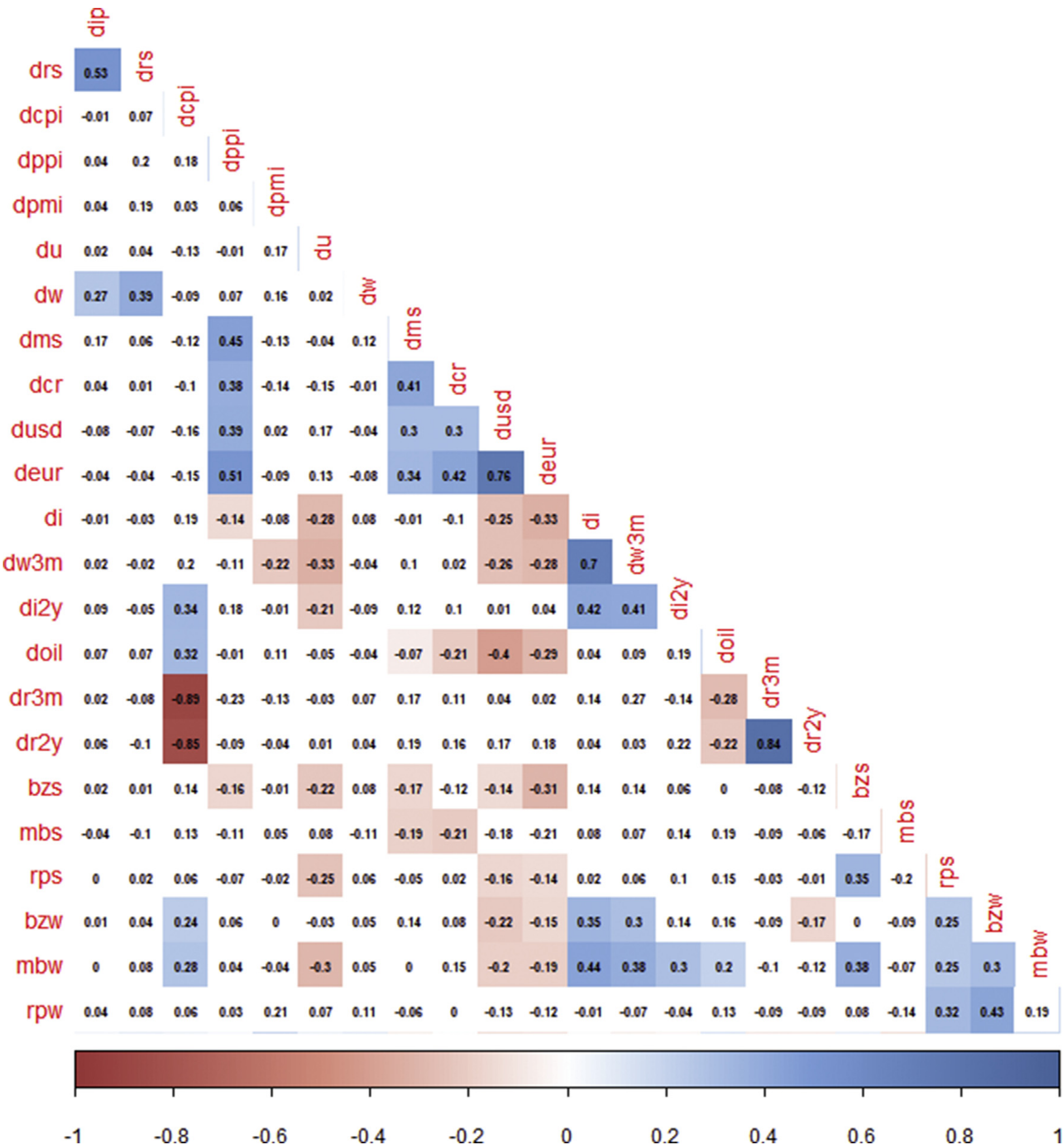

Fig. 3. Correlation between economic, financial time series and NLP indices. Note: The last six lines denote: bz – BZ WBK bank, m – mbank, rp –Rzeczpospolita daily. Suffix “s” –indicates the sentiment-based NLP index and “w” – the wordscores-based NLP index.

professional forecasters, with similar background and experience, despite looking at the same data, one group became optimistic and the other one pessimistic about the economic outlook.

The textscores calculated by the wordscores algorithm show how often bank daily reports and newspaper articles use words (lemmas) that appear in MPC statements accompanying rate hikes (higher scores) or rate reductions (lower scores). As shown in Fig. 2, on average, newspaper articles are more hawkish than bank reports, and BZ WBK corpus is the most dovish one across the entire sample.

The correlation matrix for the NLP indices is shown in Table 4. The following observations can be made. Correlations are smaller for the daily than for the monthly indices due to much higher variability and noise present in the daily data. bzs and bzw are negatively correlated, mbs and mbw are negatively correlated, and rps and rpw are

Table 5
Parameter estimates (β_i) of a multivariate regression of a monthly economic variable (first column) on a given NLP index in the form: $y_t = \alpha + \beta_1 bzs_t + \beta_2 mbs_t + \beta_3 rps_t + \beta_4 bzw_t + \beta_5 mbw_t + \beta_6 rpw_t + \varepsilon_t$.

↓y/x→	Intercept	bzs	mbs	rps	bzw	mbw	rpw	Adj. R^2
Δip	−0.419	0.074	−0.008	−0.115	−4.973	−3.345	20.441	−0.022
Δrs	2.538	−0.027	−0.031	−0.041	1.898	14.594•	13.751	0.011
Δcpi	0.108	0.011•	0.009*	−0.008	4.105*	2.353	−0.931	0.078
Δppi	1.399*	−0.032•	−0.005	−0.030	2.595	2.463	0.553	−0.001
Δpmi	−0.175	0.016	0.013	−0.040	−4.775	−3.878	20.111**	0.032
Δu	0.322	−0.005	0.000	−0.026*	0.557	−1.039	−0.256	0.021
Δw	0.219	0.007	−0.001	0.016	4.248	4.322	0.200	−0.002
Δms	2.362**	−0.051*	−0.038**	−0.005	9.737•	2.507	−9.734•	0.079
Δcr	2.620*	−0.089**	−0.043*	0.023	−1.845	13.469*	−1.373	0.061
Δusd	0.036	−0.003	−0.002•	−0.003	−0.991•	−0.437	0.083	0.041
Δeur	0.097•	−0.005**	−0.002*	−0.001	−0.537	−0.134	−0.164	0.085
Δi	0.260*	0.000	0.000	−0.005	2.148**	2.324***	−1.191•	0.152
Δw3m	0.068	0.001	−0.001	−0.002	1.033	0.915	−1.278*	0.019
Δi2y	0.050	−0.002	0.004	0.005	1.090	2.254**	−1.310	0.066
Δr3m	0.035	−0.010	−0.007	0.008	−2.008	−0.373	−0.780	0.000
Δr2y	−0.118	−0.010	−0.004	0.011	−2.849•	−0.509	−0.642	0.013
No of significant	NA	5	5	1	5	4	4	NA

Note. *** denotes statistical significance at the 0.001 level, ** at the 0.01 level, * at the 0.05 level, • at the 0.1 level. Estimates and statistical significance are reported after applying the Cochrane-Orcutt procedure to correct for autocorrelation. Δ denotes first difference. The estimates in the table show the change of the dependent economic or financial variable when the given NLP index increases by one unit.

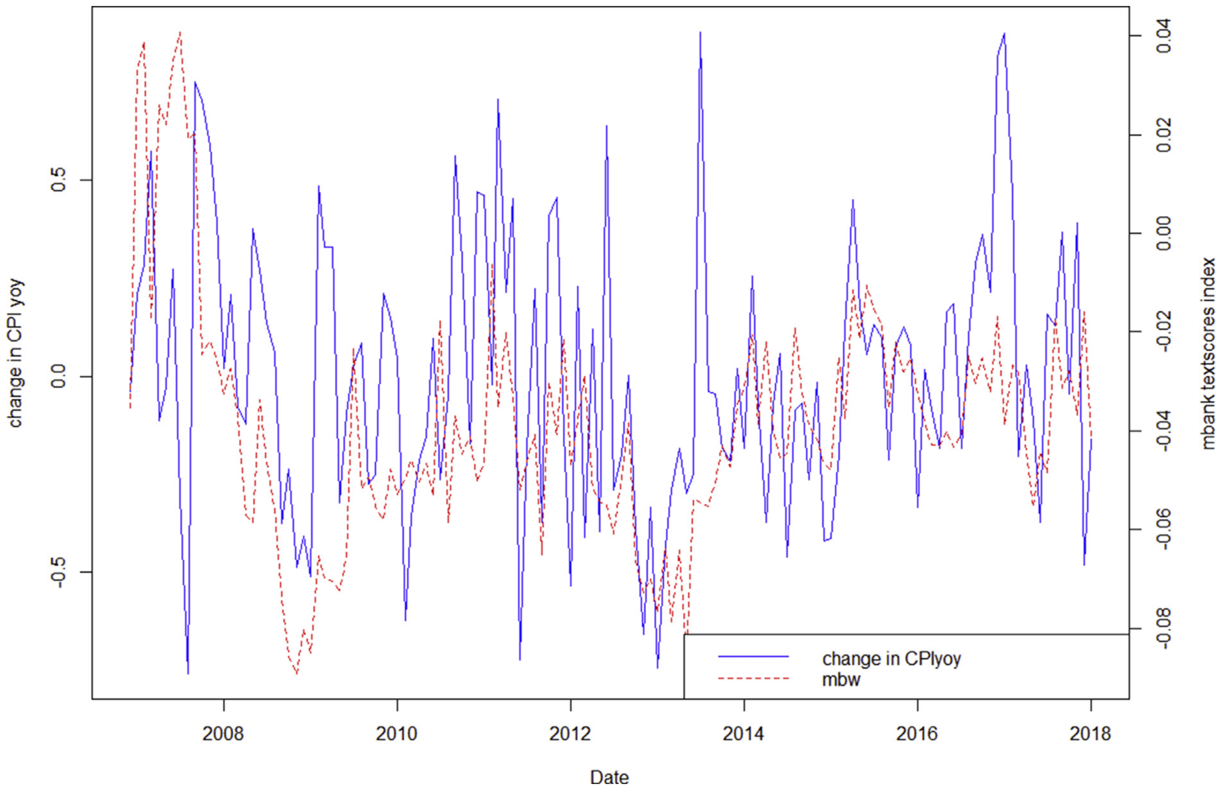

Fig. 4. Co-movement of the changes in yoy inflation and mbank textscores index.

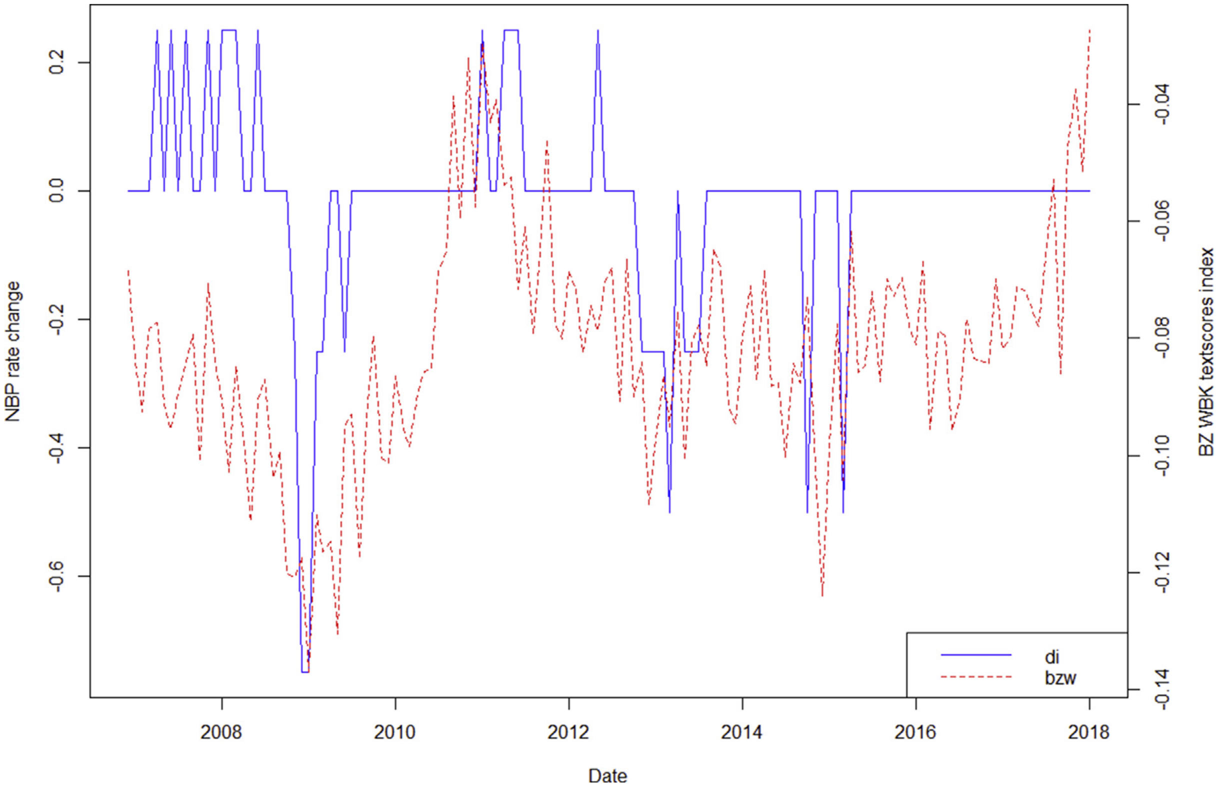

Fig. 5. Co-movement of the changes in the NBP intervention rate and BZ WBK textscores index.

positively correlated. All correlations are significant. For the monthly data, the picture is markedly different. bzs and bzw are uncorrelated, mbs and mbw are also uncorrelated, while rps and rpw remain positively correlated.

It appears that when a daily research report uses more optimistic words (sentiment is higher) is also uses more words that are typical for the MPC statement accompanying an interest cut (textscore is lower). The opposite holds for the newspaper articles that appear on a given day. In this case, a higher sentiment is accompanied by a higher textscore.

To explain these differences, we calculated sentiment for all 39 Monetary Policy Council statements accompanying rate moves issued between 2004 and 2015. For rate hike reports, the min/avg/max sentiment is 16/36.7/173. For rate cut statements, the corresponding values are −14/14.3/40. So, the MPC tends to use a more optimistic language when rates are raised and a more pessimistic style when rates are lowered.[17] showed that the National Bank of Poland influenced the media discourse in areas that fell within its mandate (monetary policy, economic situation, financial stability). It may explain why the positive sentiment/textscores correlation pattern present in MPC statements is also seen in the Rzeczpospolita articles.

But this language influence vanes in the case of bank daily research reports. It is interesting, so we decided to zoom in and look at the bank reports published immediately after the rate hike decision, on day+1 and day+2, and see what the sentiment/textscore correlation for these reports is. Reports published on day+1 exhibit a positive relationship between sentiment and textscore, but it disappears on day+2. So there seems to be some evidence that the MPC influences the bank research narrative, but it lasts only one day. Maybe the reason is that bank economists cite the arguments MPC used to justify the rate decision.

Finally, one should note that while the textscores indices calculated for different text corpora are positively correlated, this is not the case for the sentiment indices. While monthly bzs and rps tend to move together, mbs exhibits a negative correlation with the other two. It should not come as a surprise. In the case of the numerical forecasts, often, some professional forecasters predict a given variable to rise, and others predict a drop. The same turns out to be true for the forecasts based on the NLP analysis of the textual content of the forecasting reports.

Table 6
p-values of the Granger causality tests with lag order of 1, 2 and 3. Reported only values less equal than 0.1, they indicate that a column variable (NLP index) is a Granger cause of a row variable (economic and financial time series).

		bzs	mbs	rps	bzw	mbw	rpw		bzs	mbs	rps	bzw	mbw	rpw		bzs	mbs	rps	bzw	mbw	rpw
Δip	Lag							Lag							Lag						
Δrs	order = 1				0.06		0.05	order = 2				0.06	0.04	0.05	order = 3				0.09	0.02	0.05
Δcpi					0.05	0.01	0.01						0.05	0.03							0.05
Δppi																					
pmi																				0.07	
Δu						0.02			0.03				0.01				0.07			0.01	
Δw					0.10	0.02			0.09			0.02	0.00						0.04	0.00	
Δms																			0.05		0.02
Δcr									0.02			0.02				0.04			0.01		
Δusd					0.04							0.05									
Δeur					0.07	0.08					0.10	0.03						0.09	0.05		
Δi			0.00		0.06	0.00				0.01		0.03	0.00	0.02			0.01		0.03	0.00	0.05
Δw3m			0.05			0.01				0.10		0.02	0.00						0.02	0.00	
i2y						0.05						0.03	0.00	0.04					0.03	0.00	0.02
Δoil														0.10							0.06
Δr3m							0.02							0.03					0.06		0.02
Δr2y					0.01	0.09	0.01					0.01	0.07	0.02					0.02		0.04
No of significant		**0**	**2**	**0**	**7**	**8**	**4**		**3**	**2**	**1**	**9**	**8**	**7**		**1**	**2**	**1**	**10**	**7**	**8**

Table 7
Forecast RMSE for the VAR models and for ARMA models for each monthly time series y_t analyzed. The lowest RMSE model is highlighted in the bold-italic font.

↓y_t	VAR model						ARMA
	bzs	mbs	rps	bzw	mbw	rpw	
Δip	3.633	3.614	3.653	3.881	***3.470***	3.893	3.503
Δrs	2.978	2.888	2.918	3.055	***2.860***	3.015	2.882
Δcpi	0.342	0.345	0.340	0.354	***0.336***	0.336	0.341
Δppi	0.863	0.858	0.867	0.886	***0.853***	0.901	0.871
Δpmi	1.388	1.406	1.350	1.399	1.397	1.380	***1.299***
Δu	0.232	0.229	0.229	0.235	***0.215***	0.238	0.241
Δw	1.175	1.229	1.234	1.254	1.167	1.228	***1.151***
Δms	1.206	1.139	1.091	1.240	1.146	1.140	***1.041***
Δcr	1.465	1.442	1.403	1.476	1.386	***1.385***	1.402
Δusd	0.106	***0.096***	0.101	0.098	0.099	0.104	0.097
Δeur	0.063	0.065	0.068	0.075	0.064	0.064	***0.062***
Δi	0.121	***0.113***	0.121	0.117	0.110	0.122	0.117
Δw3m	0.086	***0.075***	0.077	0.107	0.077	0.076	0.079
Δi2y	0.155	0.142	0.151	0.151	***0.141***	0.146	0.152
Δr3m	0.359	0.353	0.339	0.354	0.342	***0.337***	0.340
Δr2y	0.349	0.352	0.338	0.346	0.333	0.329	***0.323***
No of lowest RMSE	0	3	0	0	6	2	5
Average RMSE	0.907	0.897	0.893	0.939	0.875	0.918	0.869

Table 8
Forecast RMSE for the random forest models for each monthly time series y_t analyzed. In the second column: model without NLP indices as factors, columns 3−5: models with NPL indices based on the given bank research reports BZ WBK (bzs, bzw), mbank (mbs, mbw) and newspaper articles (rps, rpw). The lowest RMSE model is highlighted in bold-italic font.

	NO NLP indices	bzs, bzw	mbs, mbw	rps, rpw
Δip	5.454	5.428	***5.415***	5.475
Δrs	***2.327***	2.376	2.347	2.352
Δcpi	0.268	0.272	***0.264***	0.265
Δppi	0.672	0.688	0.672	***0.660***
Δpmi	***1.233***	1.256	1.279	1.252
Δu	0.214	***0.199***	0.205	0.208
Δw	0.965	0.944	0.934	***0.934***
Δms	0.839	0.836	0.848	***0.831***
Δcr	0.850	0.853	0.851	***0.838***
Δusd	0.084	0.084	***0.082***	0.083
Δeur	0.059	0.058	***0.056***	0.057
Δi	0.070	***0.065***	0.072	0.068
Δw3m	0.046	0.042	***0.042***	0.047
Δi2y	0.103	0.104	***0.099***	0.102
Δr3m	0.280	0.285	0.284	***0.276***
Δr2y	0.279	***0.276***	0.278	0.277
No of lowest RMSE	2	3	6	5
Average RMSE	0.859	0.860	0.858	0.858

5. Contemporaneous relationships between NLP indices and economic and financial time series

After creating the monthly series for sentiment and textscores we examine correlations between the economic and financial variables and the computed scores. All economic and financial time series are in the first differences to eliminate spurious correlations. In Fig. 3, statistically significant correlations are marked with colors, red for negative and blue for positive, color intensity indicates the correlation strength. The last six rows in the heatmap represent NLP variables (sentiment and textscores) for BZ WBK, mbank, and Rzeczpospolita.

Firstly, we note that some economic variables are correlated. The first differences of the industrial output, retail sales, and wages tend to move together. Producer price inflation is positively correlated with exchange rates because export prices invoiced in foreign currencies are part of the PPI. Less obvious is PPI positive correlation with credit, but both have a foreign currency component and respond to exchange rate swings. Credit is positively correlated with money supply, as the two sides of the country's monetary balance sheet should be. The dollar and euro exchange rates tend to move together, which is typical for the floating exchange rate regime when local shocks are much more important than the global ones affecting the euro-dollar exchange rate. The unemployment rate moves inversely to interest rate changes, zloty appreciation against euro coincides with rising short-term money market rate, and nominal interest rates are positively correlated. Rising oil prices are accompanied by a falling dollar and, to a lesser extent, euro against the zloty and by a rising CPI inflation. Real interest rates are by construction negatively correlated with CPI inflation and positively correlated with each other.

However, the purpose of this research is not to investigate dependencies between economic and financial time series. Much more important are correlations reported in the last six rows in Fig. 3. Clearly, the newspaper corpus does not capture the changes in the data nearly as well as the banks' corpora do. The newspaper textscore index correlations are insignificant, and the sentiment index explains only movements in unemployment and exchange rates. BZ BWK indices are correlated with changes in eleven variables, mbank indices with ten variables. Signs of correlations with the same type of the NLP index are consistent across corpora. The rising sentiment is accompanied by falling PPI, unemployment, money supply, credit, and by appreciating exchange rates. Textscore indices rise in months that witness rising CPI, interest rates and oil prices, falling unemployment, and appreciating exchange rates. Correlation signs are in line with economic intuition and theory. For example, rising inflation/falling unemployment call for higher interest rates hence the positive/negative correlation of textscore indices with changes in 12-month CPI/changes in the unemployment rate. Appreciating zloty coincides with a more hawkish tone of the bank reports. It is consistent with the argument present in many MPC statements that the appreciation of the exchange rate increases overall monetary restrictiveness.

It is worth noting that the bank textscores are correlated with CPI inflation and unemployment changes, but largely ignore output data, such as industrial output and retail sales. We acknowledge that it may be partly to the fact that we used annual changes in these series, which are noisy. But another plausible explanation is that inflation targeting MPC convinced bank economists that inflation and unemployment are much more important factors determining monetary policy, than output.

In Table 5, we show the results of multivariate regressions of the economic variables on the six analyzed NLP indices. All economic and financial variables are in first differences (monthly changes). Even after differencing residuals were serially correlated, so we applied the Cochrane-Orcutt procedure to deal with this problem. Table 5 presents estimation results after the Cochrane-Orcutt transformation.[15]

First, we note that R^2 in each model is very low, reaching its high of 15 percent for the changes in the central bank interest rate. It was to be expected as the goal was to verify whether a given NLP index can help to build a good forecasting model, and not create a fully-fledged structural one. As shown in the last row of Table 5, NLP indices based on bank corpora prove to be more useful than the newspaper based indices. BZ WBK indices are significant in ten regressions, mbank indices in nine regressions, and Rzeczpospolita indices only in five regressions. Somewhat surprisingly, a very popular in the forecasting literature sentiment indices are significant eleven times, while less popular wordscores based indices are significant in thirteen regressions.

Figs. 4 and 5 present a co-movement of the changes in inflation and the central bank rate and the NLP indices. These examples are based on regressions with the highest R^2 in Table 5.

6. Can the narrative of the daily bank reports and newspaper articles predict economic and financial variables?

The correlation and regression analyses presented in section 5 confirm the existence of statistically significant contemporaneous relationships for the two sets of variables: the economic/financial times series and the NLP indices. In this section, we investigate whether there are interactions between the lagged values of NLP indices and economic/financial time series. It is accomplished by conducting Granger causality tests[16] and running forecast exercises with VAR and random forest models.

[15] We used cochrane.orcutt function in R.
[16] See.[25]

Statistically significant *p*-values of Granger causality tests are reported in Table 6. NLP indices based on the BZ WBK research reports are significant 30 times, indices based on mbank reports 29 times, and indices based on Rzeczpospolita newspaper 19 times. Similarly to the correlation and regression analysis, bank corpora prove much more useful than the newspaper corpus for the economic and financial forecasting. Sentiment indicators are statistically significant 12 times, and textscores indicators are significant 66 times. It does confirm that much-excited sentiment analysis is less useful in forecasting than hardly noticed in the financial literature wordscores model, with labels defined by the central bank actions and words.

The previous two steps verify whether the narrative of the daily economic research reports or newspaper articles can be used for forecasting. The VAR part of the testing procedure finds the best unrestricted VAR model for the given data and calculates the forecast RMSE gain or loss, resulting in adding the NLP index to the model.

We begin by estimating the ARMA models for all the economic time series, with automatic lag selection based on the AIC criteria and with the maximum lag length set to five. We start after the first three years, although different starting points can be selected. We store the one-month-ahead forecast, add one month to the training data, re-estimate the model, perform prediction, and repeat until the last observation is reached. For all one-period-ahead forecasts, we calculate the forecast RMSE[17] for each series. These ARMA RMSE forecast errors are benchmarks used to verify the forecasting power of the NLP indices.

In the next step, we repeat this exercise with the reduced-form VAR models:

$$\mathrm{Y_t} = \mathrm{A_0} + \mathrm{A_1}\,\mathrm{Y_{t-1}} + \ldots + \mathrm{A_P}\,\mathrm{Y_{t-P}} + \varepsilon_\mathrm{t},$$

We estimate the two-variable VAR models Y = [Δx, nlp], where x is one of the analyzed economic and financial time series in first differences to ensure stationarity and nlp is one of the six analyzed NLP indices. The maximum VAR lag length (P) is set to five, the same as in the ARMA model, and the optimal lag length selected by the AIC criterion. The forecast RMSE for each VAR model is computed in the same way as in the ARMA models and reported in Table 7.

For eleven out of sixteen analyzed time series, the VAR model with one NLP index generated a lower forecast RMSE that the ARMA model did. In these eleven cases, nine were based on mbank reports, two on newspaper articles, and none on BZ WBK reports. The average forecast RMSE calculated for all series is lowest for the ARMA model, followed by VAR with mbw index and VAR with rps index. Unlike in the previous tests this time, the mbank corpus helps to generate the lowest forecast errors, but Rzeczpospolita corpus comes second and scores better than the BZ WBK based NLP indices. The average forecast RMSE calculated for sentiment indices is lower than for the textscores indices.

To conduct the random forest forecasting exercise, we need to transform data to the format suitable for supervised machine learning. While in time series models, we can use the entire series for estimation, the random forest model requires a fixed size input. In the paper, we present a model where the economic and financial series in time *t* (outcome variable) is explained by the lagged economic and financial data in time *t-1*, *t-2*, and *t-3* (features). We split the data into train and test samples, with the first 95 observations going to the train set and the last 36 observations (three years) to the test set. The random forest model is trained on the train set. After training the model, we produce one period ahead forecasts for the test set and calculate the forecast RMSE. We repeat the exercise, but this time including both NLP indices based on a given corpus in the set of features. The results are reported in Table 8. We also tested a higher number of features up to six lags, and the results were similar.

Firstly, we note that forecast errors in Tables 7 and 8 should not be compared, as the RMSE statistics are calculated for different periods. But both tables tell a similar story. In fourteen out of the sixteen time series, lower forecast errors can be achieved by adding the NLP indices. mbank indices score best again, but this time only marginally better than Rzeczpospolita, with BZ WBK taking the last place.

7. Discussion, conclusions, and ideas for further research

In this paper, we analyzed three text corpora: articles from the leading Polish daily newspaper Rzeczpospolita and daily economic reports of two major Polish banks BZ WBK and mbank, to construct two types of monthly NLP

[17] RMSE − Root Mean Square Error $RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}$.

indices for each corpus. The first index measures the average monthly sentiment of texts. The second one measures the position of the texts on the wordscores hawkish/dovish scale determined by the words used by the Monetary Policy Council in its statements following interest rates hikes and cuts.

The goal of this paper was to test whether the sell-side research reports' narratives are more informative about the future economic and financial trends than newspaper articles. The answer depends, however, on the type of the test applied. In correlations and multivariate regression experiments, we showed that bank-based NLP indices perform better than the newspaper-based NLP indices. Because correlations and linear regressions involved no lags, it means that bank reports published in month *m* help predict economic and financial data for month *m*. Such data is often released in month *m+1,* which implies that bank-based NLP indices are more useful in short-term forecasting than the newspaper-based indices. The story changes when we test these NLP indices in the VAR and the random forest frameworks. In such a framework, we predict the economic and financial data for month *m* using texts of the reports and articles published in the previous five months (VAR, maximum lag) or three months (random forest, fixed lag). In both testing frameworks, best results, i.e., lowest forecast RMSE, are achieved by using NLP indices based on the mbank corpus, followed by Rzeczpospolita newspaper and the BZ WBK bank taking the last place.

In light of these results, the answer to the question posed in the title "Should asset managers pay for economic research?" is, unfortunately, "it depends". The narratives of some professional forecasters' reports can be much more informative than the newspaper articles. Still, some research reports may not add any value to forecasting models applying the NLP analysis. It is therefore recommended that before choosing a professional forecasting service, the asset managers should utilize a set of tests developed in this paper to verify the predictive power of a considered source.

We also showed that the sentiment analysis, prevalent in the financial forecasting literature, generates higher forecast errors that the wordscores model with labels generated by the central bank statements and rate decisions.

The value of the narrative of newspaper articles and daily bank reports can be discovered only by using machine learning methods. Consistent human coding of thousands of texts is impossible or very costly, and changes of the sentiment or in the text position on the hawkish/dovish scale from day to day (period to period) are too subtle to be noticed even for a regular and careful reader. Therefore, machine learning models and text mining should become a part of the regular research routine in bank research teams and in any businesses that take important short-term financial decisions. Professional forecasters are advised to add such indices to their short-term forecasting models as it will very likely improve the quality of their forecasts.

One question naturally emerges. We obtained statistically significant results only for some economic and financial time series, so one could argue that these results represent a data mining and could have been achieved simply by chance. However, we note in all cases, the signs of correlation coefficients and estimated regression parameters are in line with economic intuition, for example, rising interest rates coincide with more hawkish texts. Also, adding NLP indices to VAR or random forest models reduces forecast errors across the vast majority of time series.

Due to data availability, this research was based only on one newspaper − Rzeczpospolita - and on two banks: BZ WBK and mbank. A natural extension would be to add other newspapers and research reports of other banks, as well as extend this research to other countries. Institutions uniquely well-positioned to conduct such exercise are banks' economic teams that have access to their own past reports. Although it is likely that results would not be published if reports were found to offer no forecasting value. An even better option would be to ask banks to make their historical daily (periodic) reports publicly available on their websites, as in the case of BZ WBK and mbank. Academic researchers and private sector analysts would be able to evaluate the reports' quality using machine learning methods.

It would also be useful to develop a Polish language sentiment lexicon specific to economic and financial texts[18] and create sentiment indices for both analyzed corpora based on such a specialized dictionary. Our results based on the general-use sentiment lexicon and unigrams proved to have a very weak or no predictive power in the short-run.

One more promising research avenue is to use the created NLP indices in fully-fledged forecasting models, for example, structural econometric models used by the central banks and compare their forecast quality with and without NLP indices used as explanatory variables.

[18] Recently proposed methods allow for automatic generation of such domain specific sentiment lexicons, see[26] and.[27].

Funding details

This work has not been supported by any funding.

Data availability statement

The author used the following data sources for the text corpora analyzed in this paper: NBP's Monetary Policy Statements available at: https://www.nbp.pl/home.aspx?f=/polityka_pieniezna/dokumenty/komunikaty_rpp.html; BZ WBK bank daily economic reports available at: https://skarb.santander.pl/skarb/serwis-ekonomiczny/codziennik/codziennik.htmlmbank daily economic reports available at https://www.mbank.pl/serwis-ekonomiczny/raporty/komentarz-dzienny.html; and Rzeczpospolita daily newspaper articles available at the paid archive: http://archiwum.rp.pl/. All statements, reports and articles converted to the text or Rdata format can be obtained from the author upon request.

Conflicts of interest

The author has none to declare.

References

1. Financial Times. *Banks Charge Asset Managers $75,000 a Year for Research*. Financial Times; 2017, 9 April 2017.
2. Xing FZ, Cambria E, Welsch RE. Natural language based financial forecasting: a survey. *Artif Intell Rev*. 2018;50:49–73.
3. Peji'c Bach M, Krsti'c Z, Seljan S, Turulja L. Text mining for big data analysis in financial sector: a literature review. *Sustainability*. 2019;11:1277.
4. Türegün N. Text mining in financial information. *Current Analy Econ Fin*. 2019;1:18–26.
5. Sun S, Luo C, Chen J. A review of natural language processing techniques for opinion mining systems. *Inf Fusion*. 2017;36:10–25.
6. Loughran T, McDonald B. Textual analysis in accounting and finance: a survey. *J Account Res*. 2016;54(4):1187–1230.
7. Rajput V, Bobde S. Stock market forecasting techniques: literature survey. *Int J Comput Sci Mobile Comput*. 2016;5:500–506.
8. Kumar BS, Ravi V. A survey of the applications of text mining in financial domain. *Knowl Base Syst*. 2016;114(15):128–147.
9. Fisher IE, Garnsey MR, Hughes ME. Natural language processing in accounting, auditing and finance: a synthesis of the literature with a roadmap for future research. *Intell Syst Account Finance Manag*. 2016;23(3):157–214.
10. Ravi K, Ravi V. A survey on opinion mining and sentiment analysis: tasks, approaches and applications. *Knowl Base Syst*. 2015;89:14–46.
11. Nassirtoussi AK, Aghabozorgi S, Waha TY, Ling Ngo DC. Text mining for market prediction: a systematic review. *Expert Syst Appl*. 2014;41:7653–7670.
12. Zaśko-Zielińska M, Piasecki M, Szpakowicz S. A large wordnet-based sentiment lexicon for polish. In: *Proceedings of the International Conference Recent Advances in Natural Language Processing (RANLP'2015)*. 2015:721–730.
13. Jansen DJ, de Haan J. *An Assessment of the Consistency of ECB Communication Using Wordscores*. DNB Working Paper. 2010:259.
14. Bennani H. The art of central banks' forward guidance at the zero lower bound. *Rev Econ Paris*. 2018;69:192 (in French).
15. Hayo B, Kai Henseler K, Rapp MS. Estimating the monetary policy interest-rate-to-performance sensitivity of the European banking sector at the zero lower bound. *Finance Res Lett*. 2018.
16. Rybiński K. Ranking professional forecasters by the predictive power of their narratives. *Int J Forecast*. 2020. available online 16 May 2020.
17. Rybinski K. A machine learning framework for automated analysis of central bank communication and media discourse. The case of Narodowy Bank Polski. *Bank Credit*. 2019;50(1):1–19.
18. Lowe W. Understanding wordscores. *Polit Anal*. 2008;16:356–371.
19. Rybinski K. Sentiment analysis of Polish politicians. *e-Politikon*. 2017;24:162–195.
20. Rybinski K. Political sentiment analysis of press freedom. *Studia Medioznawcze*. 2018;3(74):31–48.
21. Blei DM, Ng AY, Jordan MI. Latent dirichlet allocation. *J Mach Learn Res*. 2003;3:993–1022.
22. Blei D. Probabilistic topic models. *Commun ACM*. 2012;55(4):77–84.
23. Laver M, Benoit K, Garry J. Extracting policy positions from political texts using words as data. *Am Polit Sci Rev*. 2003;97:311–331.
24. Grimmer J, Stewart B. Text as data: the promise and pitfalls of automatic content analysis methods for political texts. *Polit Anal*. 2013:1–31.
25. Granger CWJ. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*. 1969;37:424–438.
26. Tai Y, Kao H. *Automatic Domain-specific Sentiment Lexicon Generation with Label Propagation. The 15th International Conference on Information Integration and Web-Based Applications and Services, Vienna, Austria*. 2013.
27. Hamilton WL, Clark K, Leskovec J, Jurafsky D. Inducing domain-specific sentiment lexicons from unlabeled corpora. In: *Empirical Methods in Natural Language Processing (EMNLP)*. 2016:595–605.