Abstract

During the last several years a lot of academic researchers consider the possibility of predicting the economic fluctuations with the use of news data. One of the reasons for this is the development of ML techniques. Moreover, increased computing powers allows us to analyze large datasets in the research. Another reason is that the “true” news can reflect changes in agents' expectations before we see the reaction of economic variables. Thus, such news has predictive power and therefore, if it can be extracted, it can be used to forecast at some extent future fluctuations.

Considering the literature in more details we can point out several papers. For instance, Ardia, Bluteau, & Boudt (2019) in their paper revealed that inclusion of the sentiment index based on the text-analysis in the basic forecasting model with standard macroeconomic and financial variables can increase predictive power at long-period horizon. See also Algaba, Ardia, Bluteau, Borms, & Boudt (2020) for the extension and related work. In Audrino, Sigrist, & Ballinari (2020) the authors by using a novel data set and predictive regression model have shown that sentiment and attention measures increase the predictive power of the future stock market volatility.

One of the recent studies that investigate the relevance of incorporating news indexes into the forecasting the macroeconomic variables is the paper written by Larsen & Thorsrud (2019). The authors extracted 459 745 articles from Norwegian business newspaper and by using novel topic modelling approach investigate whether the textual data can help in predicting future economic fluctuations of the key macroeconomic variables. The results suggest that many news topics increase the predictive power for economic variables.

Despite the fact that there are a lot of studies in the framework of macroeconomic forecasting with the use of news data for different European countries, USA and Japan, only few academic papers consider Russian case: see Yakovleva (2018), Seleznev & Mamedli (2020), Ulyankin (2020).

In Yakovleva (2018) the author considers only one macroeconomic variable – PMI (Purchasing Managers’ Index). SVM (supervised machine learning method) and LDA (Latent Dirichlet Analysis) are used in order to proceed with the topic modeling and sentiment index constructing. LASSO and Ridge regressions are used in order to verify the forecasting power of the sentiment index. Seleznev & Mamedli (2020) paper is still on the development stage. Ulyankin (2020) paper mainly focuses on the different types of news
indexes and can be considered as technical one. As the main model to verify forecasting power ARIMA was used.

As we can see, the predicting of the economic fluctuations with the use of news data is still insufficiently considered and analyzed in the context of Russia. Our work is not unique but consider more macroeconomic variables and use more complicated and accurate methods of topic modeling and forecast. We introduce SVAR approach for investigating aforementioned aim because it allows to obtain results that have more precise economic justification for particular economy in comparison with unrestricted approaches that was used in Ulyankin (2020).

The scientific problem of our study is to investigate whether the predictive power of forecasts for macroeconomic variables can be improved with the use of news data in the context of Russia. The novelty of the study is the development of Russian specific approach to model the topics and implementation of this methodology to Russia. The specific approach consists in the consideration of the russian news and building the model by using this news.

We used web-parsing of new articles only from the web-source in order to form the raw text database. We have chosen the newspaper that concern economic, finance and development issues and related topics – “Kommersant”. We extracted all articles, that were published in the period of 2010-2020.

We apply NLU (Natural Language Understanding, subsection of NLP) algorithms and techniques for topic modeling (the process of learning, recognizing and extracting the topics). Especially, we implement LDA (Latent Dirichlet Allocation) and deep-learning LDA2VEC, since they showed their effectiveness in the published papers related to the mentioned framework. We should point out, that deep-learning method (LDA2VEC) wasn’t used in the Russian academic papers in the related field previously. Then the news and sentiment indexes are constructed with the use of modeled topics. After that the predictive power of such indexes are calculated and verified. The end point of our research is SVAR analysis of the set of macroeconomics variables (GDP ($Y$), Investments ($I$), Consumption ($C$), Inflation ($\pi$), Unemployment ($u$), Total Factor Productivity (TFP), Purchasing Managers’ Index (PMI) and Business Confidence Index (BCI)) supplemented by inclusion of news and sentiment indexes in order to evaluated the improvement in predictive power.
We show that the inclusion of news indexes and sentiment indexes, based on the LDA approach and LDA2VEC, in the baseline forecast model (SVAR) can improve the quality of the predictions and increase the predictive power. The results can be used in conducting monetary policy. Taking into account the current information space and sentiments of the public can allow to choose the future policy and instruments correctly and more precisely.

References


