DRAFT PAPER

Macroeconomic Forecasting with the Use of News Data

Author:
Dmitry A. Mikhaylov
The Russian Presidential Academy of National Economy and Public Administration
Center for Mathematical Modeling of Economic Processes

Moscow
2023

Table of Content

Introduction

During the last decade a lot of academic papers consider the possibility of predicting the economic fluctuations and macroeconomic variables volatility with the use of news data (Larsen & Thorsrud, 2019), (Kalamara, Turrell, Redl, Kapetanios, & Kapadia, 2020), (Audrino, Sigrist, & Ballinari, 2020), (Algaba, Ardia, Bluteau, Borms, & Boudt, 2020). The reason for this is the development of new machine learning techniques and enhancement of the existed methods. In addition, the recent technological progress in computing power of computers makes it possible to process the large datasets and work with Big Data. All mentioned above give us an opportunity to analyze news corpuses and implement the results in the framework of economic empirical studies.

A lot of people in each country all around the world deals with many sources of information during the day: newspapers, web-sites, radio, academic papers, TV programs etc. Such an information can be highly valuable for economic decision making. Economic agents form theirs' expectations taking into account signals, that they received from the news (Beaudry & Portier, 2014). In (Larsen & Thorsrud, 2019) the authors propose that some signals are noise, but some can bring fundamental information about the future events. The news, which signal fundamental information, can be called as "true" news. We propose that shifts in agents' expectations in line with "true" news may cause fluctuations in economic variables. This mechanism stands in line with Keynesian idea of "animal spirit", which tells us that expectations of investors and consumers may affect the economic situation.

All mentioned above means that news can reflect the changes in agents' expectations before the real (and sometimes nominal) economic variables will react on the specific event described in the news. Thus, "true" news should bring some forecasting power. In this case, we can partly predict the future fluctuations of economic variables with the help of news data, especially during the shocks and crisis, when the standard model and relationships may be invalid.

In economic literature the concept of FIRE (Full Information Rational Expectations) is often applied as one of the key assumptions of how economic agents form their expectations and response to shocks. However, several studies show that FIRE may not hold (Coibion & Gorodnichenko, 2015), (Bordalo, Gennaioli, Ma, & Shleifer, 2020). In this case, predicting model based on such FIRE concept may lead to biased estimations and forecasts. Thus, in attempt to predict future fluctuations researchers widely use machine learning techniques. To be more precise, they apply NLU (Natural Language Understanding, subsection of NLP) algorithms and techniques. NLU allows machines (computers) to understand the human language. In order to make it possible for the computer to understand human language texts, analysis of the topics of texts should be conducted. The process of learning, recognizing and extracting the topics is called topic modelling. The most applied techniques of topic modeling are: LSA (Latent Semantic Analysis), PLSA (Probabilistic Latent Semantic Analysis), LDA (Latent Dirichlet Allocation) and deep-learning LDA2VEC.

Several studies in the related framework have revealed that including news indexes, attention variables and sentiment measures constructed with the help of the topic modeling techniques mentioned above improve future forecasts and raise the predictive power. For example, (Audrino, Sigrist, & Ballinari, 2020) showed that including sentiment and attention indexes in the analysis improved predictive power of the forecasts for the future stock market volatility. (Goshima, Ishijima, Shintani, & Yamamoto, 2019) have found that their news indicator increases the accuracy of forecasting the inflation in Japan for long-term horizon (more than 3 months). (Shapiro, Sudhof, & Wilson, 2020) have found out that positive sentiment shocks affect some macroeconomic variable and allows to increase predictive power. Especially, we can point out the (Larsen & Thorsrud, 2019) paper. The authors have created their own custom aggregated news index and have shown that a lot of news topics have a predictive power for several macroeconomic 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 a few academic papers consider Russian case. Firstly, we can point out the (Yakovleva, 2018) paper. The author tries to verify, whether the forecasting power can be increased with the help of sentiment index implemented into the machine learning models and factor analysis model. In addition, on XXI April International Academic Conference on Economic and Social Development M. Mamedli and S. Seleznev have presented their abstract of the paper "What can we learn from the news?" (Seleznev & Mamedli, 2020). The authors aim to verify whether the forecast of some Russian macroeconomic variables can be improved by using sentiment indexes, which will be constructed with the use of topic modeling techniques. Recently, one more academic paper was published in the related field analyzing Russian dataset. (Ulyankin, 2020) develops a bunch of news and sentiment indexes and verifies their predictive power in the framework of ARIMA model. Nevertheless, we can point out, that the specified topic (**Macroeconomic Forecasting with the Use of News Data)** is still insufficiently considered and investigated in the context of Russia.

The **scientific problem** of our study is the investigation of whether predictive power of the forecast of macroeconomic variables can be improved with the use of news data in the context of Russia. We apply NLU algorithms and techniques for topic modeling. Especially, we implement LDA (Latent Dirichlet Allocation) since this approach has shown its effectiveness in the published papers related to the mentioned framework. Then the frequency news and sentiment news indexes are constructed with the use of modeled topics. The end point of our research is the forecast analysis of the set of macroeconomics variables [CPI **($\boldsymbol{\pi}$)**, Business Confidence Index **(BCI)**, Consumer

Confidence Index **(CCI)**, Export **(EX)**, Import **(IM)**, Net Export **(NX)**] supplemented by inclusion of frequency and sentiment news indexes in order to evaluated the improvement in predictive power. We build ARIMAX model as the baseline forecasting model.

Thus, we can highlight the main **objectives** of our study:

1. Collect the news data from one Russian news sources.
2. Process the collected news data: filter out the words from stop-word list, reduce words to their lemmas (lemmatization) and some other filtering procedures.
3. Conduct the topic modelling applying several techniques (LDA, K-Means)
4. Build the frequency news indexes and sentiment news indexes.
5. Develop the forecast model and verify the improvement in predictive power.

The **novelty** of the study is the development of Russian specific approach to topic modelling, building the indexes and implementing it in the forecasting model on Russian datasets. As soon as macroeconomic aggregates are often revised in Russia the prediction in such conditions seems to be a challenge. Thus news-based forecasting can allow to improve predictive power and increase forecasting performance (Seleznev & Mamedli, 2020).

Results. We have shown that the inclusion of frequency news indexes and sentiment news indexes, based on the LDA approach in the forecast models can improve the quality of the predictions and increase the predictive power for some variables.

Practical use. The obtained results can be used in conducting monetary policy. Taking into account the current information space and sentiments of the public can allow Central Banks to choose the future policy and instruments correctly and more precisely.

Literature review

Studies in Finance

By 2021 a lot of studies in the framework of macroeconomic and financial forecasts with the use of news data were published in the foreign literature. Several papers considering forecast in finance make an attempt to investigate what is the predictive power of the sentiment and attention variables constructed with the use of news-media platform and whether the forecasts for financial market returns can be improved by news-based approach. Some studies suggest that increasing investor's attention forecasts higher stock prices in the short-run period and lower in the long-run period (Da, Engelberg, & Gao, 2011), (Joseph, Wintoki, & Zhang, 2011). Other papers have revealed that sentiment indexes constructed with the help of data collected from Twitter or some other social media internet sources can be used to predict stock volatility (Bollen, Mao, & Zeng, 2011), (Alsing & Bahceci, 2015), (Sun, Najand, & Shen, 2016). For example, (Alsing & Bahceci, 2015) use comments and posts from Twitter in order to construct sentiment index and predict share price fluctuations of Walmart, Netflix and Microsoft. The results suggest that the forecast accuracy achieve 80 percent level in attempt to predict Walmart stock price fluctuations.

In recent studies in the finance framework, it was shown that forecasts can be improved by inclusion of variables which reflect news and sentiments, for example (Caporin & Poli, 2017). Especially, we can point out the (Audrino, Sigrist, & Ballinari, 2020) paper. The authors extend the basic heterogenous autoregressive model by inputting the sentiment and attention indexes as additional predictors. In order to classify the sentiment of messages on Twitter the authors use Deep-LSA technique. Thus, 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. Moreover, not only newspapers and other conventional source of information can be used in the framework of the related analysis. (Antweiler & Frank, 2004) revealed that sentiment in messages, which are posted regularly on "Yahoo! Finance" can also increase the predictive power of stock market volatility.

Topic Modelling

Firstly, we should point out that the authors quite often implement an easy procedure of topic modelling bypassing the complicated one (as LDA, PLSA, LDA2VEC, etc.). (Ulyankin, 2020) analyzes and compares different economic activities indexes with respect to their predictive power. However, in author's research no topic modelling was conducted. The author considers all the news related to such topics as "economics", "business" and "politics" sorted by the web-sites categories.

(Ardia, Bluteau, & Boudt, 2019) use the corpus of several US newspapers available at LexisNexis, supplemented by its Smart Indexing technology, that allows easily to sort and identify the topic of the specific article.

(Tobback, Naudts, Daelemans, Junque de Fortuny, & Martens, 2018) in their paper consider two methods (modality annotation and SVM) of text mining in order to create the EPU index (Economic Policy Uncertainty index). The original one assumes, that the news is related to policy uncertainty if it contains the specific keywords. SVM method (Support Vector Machine classification) is an example of supervised machine learning text mining technique. The authors manually labeled 500 articles that contained the word "economy" by the number "1" (if the specific article was related to some policy uncertainty) and "-1", otherwise. Then the model was trained on this sample and the whole news dataset was labeled by appropriate index. The authors propose and verify, that SVM method performs better than the original, and it allows to improve the predictive power of the forecast.

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. In order to proceed with the topic modeling, they use the LDA approach, introduced by (Blei, Ng, & Jordan, 2003). The model was constructed with the use of 7500 Gibbs simulations and 80 unique topics were formed. Considering 17 high probability words in each topic it was revealed that at least one word intersects with another topic through all the 80 topics.

One another studies in the related framework conducted by a Russian researcher is (Yakovleva, 2018). The author uses the study of (Thorsrud, 2016) as a basis for her research. She seeks to investigate whether the use of high-frequency indicator based on news can increase the predictive power of forecasts for future economic activity. Around 60 000 news items from single Russian news source from 2014 to 2018 were collected. In order to proceed with topic modelling the author implemented probabilistic topic model – LDA (Latent Dirichlet Allocation) method. LDA method allows to cope with the limitations and drawbacks of algebraic topic models, such as LSA (Latent Semantic Analysis) and VSM (Vector Space Model). As the main hyperparameters the default set was used. The final number of topics was chosen with the help of coherence value. Thus, the author created 50 relevant topics by implementing LDA on the extracted news dataset.

At the current moment one more study in this framework is conducted by another Russian researchers from Bank of Russia. M. Mamedli and S. Seleznev takes an attempt to implement widespread practice in the foreign researches in the Russian framework. Theirs news data set consist of over 260 000 of news articles of the Russian newspaper “Vedomosti”. They also use the vintage values of the key macroeconomic variables including all the historical revision: GDP, IP, construction, real wage growth, unemployment, real sales and investments. The authors use 3 modelling topic techniques: LDA, LSI (Latent Semantic Indexing) and DOC2VEC. For each topic model the authors define the optimal number of topics by coherence analysis.

News Indexes

A lot of studies consider macroeconomic forecasting in the framework of news-based approach, constructing its own news (frequency and sentiment) indexes. For instance, (Ardia, Bluteau, & Boudt, 2019) in their paper seeking to verify predictive power of the custom sentiment index based on news. The aim is to investigate the value added by the sentiment index for forecasting the log-growth of the US industrial production (k.o. economic growth of the US). The authors apply the methodology that classify the tones of the articles related to the economic growth published in the major US newspapers as “positive”, “negative” and “neutral”. They 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. Some other authors seeking to investigate whether the forecasting of some aggregated indexes (CPI, PMI and etc.) can be enhanced by the inclusion of news and sentiment indexes constructed with the use of machine learning techniques. For example, (Nyman, Ormerod, Smith, & Tuckett, 2014) considering the news articles which contain two words “anxiety” and “excitement” evaluate the frequency of occurrence of such words and calculate the related measure. Then, this indicator which reflects average frequency of appearance was included in the regression. As the result the predictive power for the forecast of Consumer Price Index in Michigan was improved.

(Yakovleva, 2018) in her paper constructs the own sentiment index. In order to proceed with tone labeling the author uses the supervised machine learning technique - SVM method (Support Vector Machine). Around 3438 news articles are manually classified by the author as “1”, if tone of the article is positive, and “-1” for negative tone. The training set contains 2600 news articles. On that set the classifier is trained in order to “learn” how to label the articles by tones. On the test sample the trained model correctly predicted the tone of the article in 64% cases. As the result of SVM method the researcher gets probability distribution of tones. All the articles with probability less than 60% probability of getting specific tone were exclude from the news dataset. Finally, the tone is multiplied by the probability for each of the article and the day average measure is calculated.

(Larsen & Thorsrud, 2019) in their paper apply another approach. In order to create the sentiment news index, they count the number of positive and negative words in the news corpus. The tone of the words was defined with the help of **Harvard IV-4 Psychological Dictionary**, which contains the word list and relative tone. The authors include in their dictionary of descriptors 40 positive and 39 negative words. Then the positive and negative words were counted through one day in each article and divided by the total number of words in the news articles during the day. Finally, the sentiment index was calculated by subtracting the negative day ratio from the positive day ratio.

One of the fundamental papers related to the different methods of constructing indexes is (Ulyankin, 2020). He analyzes several “manual”, search, news and sentiment indexes and compares them with respect to their predictive power. Especially, we should point out the methodology of news index constructing. The author calculates frequency index and sentiment index for the news corpus.

In order to create the frequency index, the method represented in (Столбов, 2011) is used. Firstly, the list of crisis descriptors was formed. Then, the author counts the number of articles, that contain each of the descriptor during the day. After that, the appropriate weight is assigned to each number with the help of correlation coefficient. Finally, all the weighted numbers are summed up and the index is constructed.

In order to tone the news articles and construct the sentiment index, the author implements the Russian language tonal dictionary, created by the project (“Карта слова”). All the words in this dictionary are colored by the scalar value of the emotional-evaluative charge from the continuous range [-1; 1]: “-1” means the maximal negative tone of the word, “1” means the maximal positive tone of the word, and 0 means neutral tone or the absence of the specific word in the dictionary. The author changes all the words in each article on the relative tone (continuous number) from the dictionary “Карта слова”. The tone of each article is calculated as the sum of all the tone numbers divided by the number of words in the specific article. Then the data was aggregated to monthly basis. For additional information and extension see (Голощапова & Андреев, 2017).

Time-series Analysis and Forecasting

Along with the machine learning techniques, different techniques of time series analysis and forecasting are widely used to investigate whether predictive power of the forecast of macroeconomic variables can be improved with the use of news data or not. The starting point of the discussion is the empirical evidence that news shocks account for a major part of fluctuations in GDP and may be a nature of the business cycles (Beaudry & Portier, 2006). However, the nature and understanding of news shocks used for the analysis and forecasting vary over the time.

Pioneering studies assume stock prices to be a proper variable for capturing any changes in agents' expectations and use them to represent them (see for instance, (Beaudry & Portier, 2006), (Barsky & Sims, 2011)). Both (Beaudry & Portier, 2006) and (Barsky & Sims, 2011) use a vector autoregressive approach and identify news shocks with the innovation in stock prices orthogonalized with respect to total factor productivity to investigate whether agents' expectations matter for the fluctuations in macroeconomic variables or not. In this case, news shocks should be interpreted as expected changes in future total factor productivity observed in advance and reflected in stock prices.

(Beaudry & Portier, 2006) find that permanent changes in total factor productivity are preceded by booms in stock markets, indicating the predictive power of the variable representing agents' expectations. Additionally, authors state that expected developments of the economy are more likely to account for fluctuations in macroeconomic variables compared to the unexpected ones.

(Barsky & Sims, 2011) obtain that news shocks cause changes in consumption, output, hours of work and investment, highlighting the usefulness of incorporating news shocks in the models for better both explanatory and predictive power. However, authors point out the need of seeking deeper structural explanations for the nature of news shocks. Further (Barsky, R. B.; Sims, E. R., 2012) introduced consumer confidence index instead of stock prices to represent agents' expectations and, thus, to obtain news shocks. They find that the consumer confidence index has high predictive power for macroeconomic variables such as consumption and output.

(Forni, Gambetti, Lippi, & Sala, 2017) also represents agents' expectations with the stock prices and use vector autoregressive approach to investigate whether agents' expectations matter for the fluctuations in macroeconomic variables or not. Authors find that bulk of disturbances in dynamics of GDP, consumption, and investment are caused by agents' expectations of oncoming developments of the economy, pointing them out to be a major source of business cycle fluctuations. However, they complement the critiques of the vector autoregressive approach proposed by (Blanchard, L'Huillier, & Lorenzoni, 2013) and (Barsky & Sims, 2011), pointing out the failure of these models to distinguish between news shocks and noise shocks.

Limitations of the use of several proxies for news shocks along with the occurrence of machine learning techniques that allow to process large datasets rise the research interest to the direct use of news (see for instance (Larsen & Thorsrud, 2019), (Kalamara, Turrell, Redl, Kapetanios, & Kapadia, 2020), (Goshima, Ishijima, Shintani, & Yamamoto, 2019)) and internet search data (see for instance, (McLaren & Shanbhogue, 2011), (Choi & Varian, 2012), (D'Amuri & Marcucci, 2017)) aggregating them into indexes to investigate their explanatory and predictive power for macroeconomic variables.

Mentioned papers (McLaren & Shanbhogue, 2011), (Choi & Varian, 2012), (D’Amuri & Marcucci, 2017) introduce autoregressive models with exogenous variables for forecasting different macroeconomic variables using data from Google Trend. (McLaren & Shanbhogue, 2011) and (Choi & Varian, 2012) find that use of internet search data is beneficial for forecasting unemployment. However, (Choi & Varian, 2012) claims that use of internet search data is helpful for forecasting macroeconomic variables only for the nearest future. Additionally, (McLaren & Shanbhogue, 2011) notice several limitations of the data from Google Trend, including lack of information on the actual volume of searches and basing data on a subsample. (D’Amuri & Marcucci, 2017) also point out the enhancement of predictive power of the model for unemployment, when data from Google Trend is added to the model, specifically in turning point of the economic developments.

Overall, using data from Google Trend solves the problem caused by use of several proxies for news shocks, however, limits the horizon of accurate forecasting. Thus, the research interest switches to the direct use of news aggregated into indexes for investigation their explanatory and predictive power for macroeconomic variables. Similarly mentioned papers (see for instance (Larsen & Thorsrud, 2019), (Kalamara, Turrell, Redl, Kapetanios, & Kapadia, 2020), (Goshima, Ishijima, Shintani, & Yamamoto, 2019)) introduce autoregressive models with exogenous variables for evaluating the effect of adding several news indexes into models for their predictive power.

(Larsen & Thorsrud, 2019) introduces autoregressive models with exogenous variables to evaluate the predictive power of different news topics for several macroeconomic variables, such as output, investments, consumptions, total factor productivity and asset prices. Authors conclude that many news topics can increase the predictive power of the model. Then, topics with the highest predictive power are used for news index construction and structural analysis is conducted, verifying that news shocks account for fluctuations in macroeconomic variables. (Yakovleva, 2018) and (Feuerriegel & Gordon, 2019) follow quite the same procedure, but seek to incorporate several topics in one model simultaneously. Thereby, the problem of limited numbers of available observations for the variables of interest and, thus, limited number of topics that might be included in the model is highlighted.

(Goshima, Ishijima, Shintani, & Yamamoto, 2019) find that including news index in the model for unemployment improves its predictive power, specifically for the long-term horizon. (Kalamara, Turrell, Redl, Kapetanios, & Kapadia, 2020) point out the enhancement of predictive power of the models for unemployment, output and inflation, when news index is added to the models, specifically in recession periods.

(Ulyankin, 2020) introduces an autoregressive integrated moving average model and finds that news indexes based on data from the Internet have predictive power and granger causes the new indexes based on surveys.

Machine Learning Forecasting Models

Machine learning techniques are also widely used to forecast macroeconomic variables. They enable to include a large number of regressors into the model. It is especially useful for the forecasting macroeconomic variables with use of news topics. For example, (Yakovleva, 2018) obtains 50 news topics to be simultaneously included into the model as regressors to forecast PMI and along with (Feuerriegel & Gordon, 2019) state the usefulness of machine learning techniques to deal with a situation when number of news topics significantly exceeds the number of observations and common linear regressions cannot be used. However, (Yakovleva, 2018) tests LASSO and Ridge regressions and ends up with the only 24 news topics instead of 50 initially chosen due to the overfitting of the models.

(Feuerriegel & Gordon, 2019) uses several regressions, such as LASSO, Ridge, ENET and RF to forecast GDP, unemployment rate and inflation. Differently to (Yakovleva, 2018), authors add time lags of the variable of interest to the model to forecast the variable of interest. (Kalamara, Turrell, Redl, Kapetanios, & Kapadia, 2020) follows (Feuerriegel & Gordon, 2019) and tests LASSO, Ridge, ENET and RF regressions to forecast GDP, unemployment rate, inflation, and consumption. Differently to (Feuerriegel & Gordon, 2019) authors include only one time lag of the variable of interest and a news topic to the model to forecast the variable of interest.

All aforementioned papers ((Yakovleva, 2018), (Feuerriegel & Gordon, 2019), (Kalamara, Turrell, Redl, Kapetanios, & Kapadia, 2020)) state that inclusion news topics into the model to forecasting macroeconomic variables improves the accuracy of the forecast.

Data

News Data Collection

The starting point of the study is scraping the news data. The quality and the source of news data are crucial in receiving the final results in our research. We used web-parsing of news articles only from the web-source in order to form the raw text database. Following the procedure reflected in the (Yakovleva, 2018), (Larsen & Thorsrud, 2019), we managed to use single news source. We should point out, that the consideration of only one news web-source might have several drawbacks. Firstly, the specific chosen newspaper may be biased in the sense of covering news (political issues, personal opinions of editors, etc.). Secondly, each newspaper has its own target audience. So that, the analysis of single one newspaper led to capturing only one specific audience and missing the others. Thirdly, some of the newspapers may not to publish the news about specific events (especially, economic and political events) due to several reasons. In this case, the coverage of topics included in the research may not to be comprehensive and extensive. However, there is a problem of scraping the news from the web-sites related to the specific web architecture. We have concentrated on only one single newspaper with the easiest web architecture of the code. The future extension of research in the way of considering several news sources should be done.

As mentioned in (Yakovleva, 2018) the good news web-source should have three main key features: (1) relativeness to the economic topics (news), long time-series availability, (3) easy and convenient data parsing. Since we are interested in macroeconomic forecasting, we have chosen the newspaper that concerns economic, finance and development issues and related topics – **“Коммерсантъ”**. The web-site of this newspaper allows to get access to news of the long time period. Also, the news from **“Коммерсантъ”** web-site can be easily extracted (parsed).

We extracted all articles, that were published in the **period from the beginning of 2010 to the end of 2020**. The raw dataset can be found in **Appendix A1. News Data**. The link to the code for parsing the news is also presented in the **Appendix A2. Parsing code.**

News Data Processing

The next step is data processing. In most of the cases we replicated methods and techniques, that were implemented in related literature and studies (Yakovleva, 2018), (Larsen & Thorsrud, 2019), (Audrino, Sigrist, & Ballinari, 2020).

Firstly, we deleted all the null observations, which emerged during the code execution. Secondly, we applied lemmatization technique. Lemmatization was conducted with the use of “pymystem3” package (Yandex’s “Mystem” adapted for Python). There was an option to use "pymorphy2" package for lemmatization, but "pymystem3" package is better in managing contextual homonymy. In plenty of papers stemming is used as the procedure of data processing (Thorsrud, 2016), (Audrino, Sigrist, & Ballinari, 2020). Nevertheless, we prefer lemmatization to stemming as soon as the last may convert a word to a form that doesn't exist in Russian language (for example: the word "иду" may be converted to "ид"). Thirdly, we filtered out punctuations, spaces and other special symbols from our dataset.

Finally, the words from the stop-word list also should be removed. Stop-word list typically consists of words, that are not meaningful and do not contain any information related to the topics (Larsen & Thorsrud, 2019). For this aim we took stop-word list from "nlkt" package and add to that list some extra simple meaningless words and several other words which were assigned to the majority of the created by LDA procedure topics at first attempt: ***['наш', 'гг', '"Ъ"', "господин", "год", "сообщать", "заявлять", "также", "новый", "человек", "россия", "—", 'какой-то', 'просто', 'это', 'весь', 'свой', 'мочь', 'очень', 'самый', ' ', '\n', 'ъ', 'рф', 'тыс', 'ул', 'время', 'который', 'однако', 'the', 'владимир', 's', 'становиться']***. These words did not contribute to adequate and effective distinction between different topics. However, such words may appear frequently in the text, so that taking the large weight in the specific article. Thus, the topic, that was created in such a way may be biased. The processed news dataset can be found in **Appendix B1. Processed News Data**. The link to the code for processing the news is also presented in the **Appendix B2. Processing Code.**

We represent two random articles that have been filtered and processed:

Table #1

Examples of the original and processed articles

Original Article	Processed Article
Российские и британские акционеры ТНК-ВР подписали акционерное соглашение, завершающее процесс урегулирования корпоративного конфликта. В ближайшие недели будет объявлено о назначении нового главы компании. Как говорится в сообщении ВР от 9 января, пересмотренное соглашение нацелено на улучшение баланса интересов между владельцами компании – ВР и консорциумом AAR. Согласно измененному договору, вместо равного представительства сторон в совете	["российский", "британский", "акционер", "подписывать", "акционерный", "соглашение", "завершать", "процесс", "урегулирование", "корпоративный", "конфликт", "близкий", "неделя", "объявлять", "назначение", "глава", "говориться", "сообщение", "вр", "январь", "пересматривать", "соглашение", "нацеливать", "улучшение", "баланс", "интерес", "владелец", "компания", "вр", "консорциум", "aar", "согласно", "изменять", "договор", "вместо", "равный",

директоров компании будет назначаться по четыре представителя от BP и ARR плюс три независимых директора. Предусматривается, что ряд вопросов, включая одобрение крупных сделок и изменение финансовой структуры группы, будут требовать единогласной поддержки советом. BP будет по-прежнему выдвигать кандидатуру на пост главного управляющего компании, а AAR назначать председателя совета.	“представительство”, “сторона”, “совет”, “директор”, “компания”, “назначаться”, “четыре”, “представитель”, “bp”, “arr”, “плюс”, “независимый”, “директор”, “предусматриваться”, “ряд”, “вопрос”, “включая”, “одобрение”, “крупный”, “сделка”, “изменение”, “финансовый”, “структура”, “группа”, “требовать”, “единогласный”, “поддержка”, “совет”, “вр”, “выдвигать”, “кандидатура”, “пост”, “главный”, “управляющий”, “компания”, “aar”, “назначать”, “председатель”, “совет”]
Российское правительство утвердило правила маркировки федеральными специальными марками всей алкогольной продукции в стране, в том числе ввозимой из-за рубежа. До этого специальными марками помечалась продукция, произведенная внутри страны, а ввезенная — акцизами. Цена спецмарок составит 1,89 тыс. руб. за 1 тыс. штук без учета налога на добавленную стоимость. Выдачу и контроль за использованием таких марок будет осуществлять Росалкогольрегулирование. При совершении таможенных процедур контролировать маркировку будут таможенные органы. При этом таможня сможет выдавать до 1 апреля 2021 года акцизные марки на импортный алкоголь, если заявление на это было подано до 1 января 2021 года. Ввозить алкоголь с акцизными марками можно будет до 1 января 2022 года.	[“российский”, “правительство”, “утвердить”, “правило”, “маркировка”, “федеральный”, “специальный”, “марка”, “алкогольный”, “продукция”, “страна”, “число”, “ввозить”, “рубеж”, “специальный”, “марка”, “помечаться”, “продукция”, “производить”, “внутри”, “страна”, “ввозить”, “акциз”, “цена”, “спецмарка”, “составлять”, “руб”, “штука”, “учет”, “налог”, “добавлять”, “контроль”, “использование”, “марка”, “осуществлять”, “росалкогольрегулирование”, “совершение”, “таможенный”, “процедура”, “контролировать”, “маркировка”, “таможенный”, “таможня”, “смочь”, “выдавать”, “апрель”, “акцизный”, “марка”, “импортный”, “алкоголь”, “заявление”, “подавать”, “январь”, “ввозить”, “алкоголь”, “акцизный”, “марка”, “январь”]

Source: *author’s calculations*

Numerical Data Description and Collection

Several macroeconomic variables are chosen to verify whether the inclusion of news and sentiment indexes in the forecasting models for the Russian economy improves the predictive power. Data on these variables was collected from different sources, primarily from OECD Database and IMF database (for more details, see **Appendix C2. Numerical Data Description Table**) and covers the period **from the beginning of 2010 to the end of 2020**. The raw numerical dataset can be found in **Appendix C1. Numerical (Macro) Data**.

The first group of macroeconomic variables used to attain aforementioned aims of the research consists of BCI and CCI. (Yakovleva, 2018) and (Larsen & Thorsrud, 2019) highlight the importance of forecasting indexes reflecting the expectations of decision-makers for macroeconomic policy conduction. (Yakovleva, 2018) finds that such index as PMI is highly correlated with GDP and can be used as an accurate measure of business cycle dynamics instead of GDP that is commonly collected quarterly or yearly and published with the lag, limiting the range of models that can be used for the forecast and structural analysis. (Ulyankin, 2020) points out that indexes reflecting the expectations of decision-makers itself has high predictive power.

The next macroeconomic variable is inflation following (Kalamara, Turrell, Redl, Kapetanios, & Kapadia, 2020) and (Goshima, Ishijima, Shintani, & Yamamoto, 2019). The dynamic of inflation is highly affected by agents' expectations. Therefore, adding news and sentiment indexes in the models for forecasting inflation in Russia is expected to cause the enhancement of the forecast quality.

Finally, we also consider such macroeconomic variables as Export **(EX)**, Import **(IM)** and Net Export **(NX)** due to availability of such data on monthly basis. It also allows to us to verify whether the inclusion of domestic new indexes can improve the predictions of macroeconomic variables that are highly connected with foreign sector and other countries.

Methodology

Topic Modelling

After filtering and processing the data topic modeling procedure should be conducted. As we mentioned in the **Introductory** part of our work, in order to understand texts, analysis of the topics of texts should be conducted. The process of learning, recognizing and extracting the topics is called topic modelling. The most applied techniques of topic modeling are: LSA (Latent Semantic Analysis), PLSA (Probabilistic Latent Semantic Analysis), LDA (Latent Dirichlet Allocation) and deep-learning LDA2VEC.

Our method of topic modeling was LDA approach (as in (Yakovleva, 2018), (Larsen & Thorsrud, 2019), (Seleznev & Mamedli, 2020)). LDA allows to extract interpretable topics from the original news dataset. Each of the topic is represented by a set of words that is highly associated with that topic. We suggest that LDA approach allows us to get more robust results.

Informally, LDA considers the co-appearance of separated terms in texts, so that terms which likely to appear together would be placed in the same set of words (topic). It is important to mention two aspects: firstly, LDA does not label topics by itself – names for sets are developed by authors. Secondly, it may happen that one particular word is distributed to several topics if it is related with different themes (for example, “судья” may appear either in “sport” or “law” topic). But, as we mentioned before, if certain word is presented in all topics, we may consider it as a preposition or conjunction, which should be filtered out.

Formally, hierarchical Bayesian model of LDA may be illustrated with joint probability distribution function (Blei, Ng, & Jordan, 2003):

$$p(\theta, z, w|\alpha, \beta) = p(\theta|\alpha)\prod_{n=1}^{N} p(z_n|\theta)p(w_n|z_n, \beta)$$

- θ – a topic mixture
- z – a set of topics
- w – a set of words

Hyperparameters

- α – density of document-term
- β – density of words

Implementation of the LDA method has some restrictions. For example, LDA can perform worse than comparable techniques if the corpus consists of short articles (Davison, 2010). In our case the average length of one particular article is **approximately 166 words**. So that, the LDA technique seems to be appropriate in our research.

News Indexes

After the topic modelling procedure, the news indexes are calculated. We are going to create two indexes: **frequency news index** and **sentiment news index.**

In order to build the **Frequency News Index,** we combine the approaches used in (Nyman, Ormerod, Smith, & Tuckett, 2014), (Larsen & Thorsrud, 2019) and (Ulyankin, 2020). In the first paper (Nyman, Ormerod, Smith, & Tuckett, 2014) the authors considering the news articles which contain two words "anxiety" and "excitement" evaluate the frequency of occurrence of such words and calculate the related measure. Construction the frequency news index based on two descriptors may be biased and doesn't capture all the picture. So that, we are going to use the extended lists of descriptors in our analysis. We form three lists of descriptors: "basic", "optimal" and "full". Each list has a different number of words included and different sense. In the "basic" list we include the minimal number of words (**40 words**), that catches the economic situations, dynamics and fluctuations. The "optimal" list contains of **148 words** and reflects the economic topics fully. The "full" list includes **227 words** that are related to economics, law, politics, culture and art.

In the second paper (Larsen & Thorsrud, 2019) the authors split all the descriptors into two list: **"positive"** and **"negative"** words. We replicate the same procedure in our analysis. All three lists with "positive" and "negative" words division can be found in **Appendix D1. The list of Descriptors for Frequency News Index**.

In order to calculate the frequency news index, we extend the methodology, represented in (Larsen & Thorsrud, 2019) and (Ulyankin, 2020). Firstly, we count the number of times positive and negative words from the descriptors list appears in each of the article for all the timeline. As the result we get two metrics: positive and negative word frequency for each article. Then we sum up these two numbers for each of the article from the same topic (which was formed with the help of topic modelling procedure) during the day separately, and divide it by the total number of words in these corresponding articles. Thus, we obtain the positive and negative ratio for each of the topic during the day. Finally, we subtract the negative ratio from the positive ratio and get the daily frequency news index. In order to get the monthly and quarterly index we sum up all the daily indexes and divide by the appropriate (~30 or ~90) number of days.

In order to build the **Sentiment News Index,** we partially replicate the approach used in (Ulyankin, 2020). As in the mentioned paper, we use the tonal dictionary of Russian language "Карта слов" in order to tone each word in each article. Words in this dictionary are colored by the scalar value of the emotional-evaluative charge from the continuous range [-1; 1]: "-1" means the maximal negative tone of the word, "1" means the maximal positive tone of the word, and 0 means neutral tone or the absence of the specific word in the dictionary. Firstly, we change all the words in each article in our corpus by the appropriate emotional-evaluative value from the dictionary “Карта слов”. Then we sum all the assigned numbers up and divide it by the number of words in the article. So that we get the emotional value of the article. After that, the emotional value of each article during the day related to the specific topic are summed up and divided by the number of articles in that day. Following such a procedure we are getting the **daily topic specific sentiment index**. In order to get the monthly and quarterly index we sum up all the daily indexes and divide it by the appropriate (~30 or ~90) number of days.

Time-series and Machine Learning models

We introduce autoregressive integrated moving average model $\mathbf{ARIMA(p,d,q)}$ and autoregressive integrated moving average model with the exogenous variable $\mathbf{ARIMAX\ (p,d,q)}$ to forecast several variables of interest including and excluding news index. In our research we consider several machine learning models: **Linear**, **Ridge**, **Lasso**, **Elastic Net**, **Random Forest**, **XGBOOST**. The procedure of the analysis is the following:

1. Data sample is split into training and test samples. The test sample includes **48%** of observations.
2. Target variables and news indexes contain seasonal variation, some of them also have trend. The first difference is used to gain stationarity.
3. Using the training sample, the following autoregressive model is estimated for the target variable:

$$\mathbf{X'_t = \sum_{i=1}^{p} \phi_i X'_{t-i} + \varepsilon_t\ ,} \qquad \mathbf{(1)}$$

where $\boldsymbol{X'_t}$ is the first difference of the target variable in period $\boldsymbol{t}$, $\boldsymbol{\phi_1, \phi_2, \dots, \phi_p}$ are coefficients for the lags of the target variable, $\boldsymbol{\varepsilon_t}$ is error term in period $\boldsymbol{t}$, $\boldsymbol{p}$ is a maximum number of lags of the target variable. Several values of $\boldsymbol{p}$ are used: **1**, **3** and **6**.

4. Using the training sample, the following autoregressive model with exogenous variables (news indexes) is estimated for the target variable:

$$\mathbf{X'_t = \sum_{i=1}^{p} \phi_i X'_{t-i} + \sum_{j=1}^{n} \boldsymbol{\varphi_{j,t} \omega'_{j,t}} + \varepsilon_t\ ,} \qquad \mathbf{(2)}$$

where $\boldsymbol{X'_t}$ is the first difference of the target variable in period $\boldsymbol{t}$, $\boldsymbol{\phi_1, \phi_2, \dots, \phi_p}$ are coefficients for the lags of the target variable, $\boldsymbol{\varepsilon_t}$ is error term in period $\boldsymbol{t}$, $\boldsymbol{p}$ is a maximum number of lags of the target variable, $\boldsymbol{\varphi_1, \varphi_2, \dots, \varphi_n}$ are coefficient for the lags of exogenous variables, $\boldsymbol{\omega'_1, \omega'_2, \dots, \omega'_n}$; $\boldsymbol{\omega_t}$ are lags of exogenous variables, n is a number of news indexes. Several values of $\boldsymbol{p}$ are used: **1**, **3** and **6**. The current values of news indexes are used as exogenous variables to understand, whether they are able to explain the current state. This approach is nowcasting.

The lagged values of news indexes are used as exogenous variables to understand, whether they are able to predict the future state. This approach is forecasting, and the estimating autoregressive model is the following:

$$\mathbf{X'_t} = \sum_{i=1}^{p} \boldsymbol{\phi}_i \mathbf{X'_{t-i}} + \sum_{j=1}^{n} \boldsymbol{\varphi}_{j,t} \boldsymbol{\omega'}_{j,t-1} + \boldsymbol{\varepsilon}_t \,, \qquad \textbf{(3)}$$

where $\boldsymbol{X'_t}$ is the first difference of the target variable in period $\boldsymbol{t}$, $\boldsymbol{\phi_1}, \boldsymbol{\phi_2}, \dots, \boldsymbol{\phi_p}$ are coefficients for the time lags of the target variable, $\boldsymbol{\varepsilon_t}$ is error term in period $\boldsymbol{t}$, $\boldsymbol{p}$ is a maximum number of time lags of the target variable, $\boldsymbol{\varphi_1}, \boldsymbol{\varphi_2}, \dots, \boldsymbol{\varphi_n}$ are coefficient for exogenous variables, $\boldsymbol{\omega'_1}, \boldsymbol{\omega'_2}, \dots, \boldsymbol{\omega'_n}$; $\boldsymbol{\omega_t}$ are exogenous variables, n is a number of news indexes. Several values of $\boldsymbol{p}$ are used: **1**, **3** and **6**.

Besides, machine learning methods (***lasso, ridge, random forest, XG boost, elastic net***) are used to estimate the autoregressive models represented by the **equation 2** and **equation 3**. The hyperparameters are chosen to minimize RMSE.

5. Then forecasts/nowcasts and forecast/nowcast errors are calculated using the test sample and fixed coefficients and parameters based on the results for the training sample.

Besides, retraining is introduced, then after each forecast/nowcast this observation is added to the training sample, deleted from the test sample and the model (and hyper parameters for machine learning methods) is estimated again until test set is empty.

Two metrics such as **RMSE** and **MAE** are used to evaluate the accuracy of forecast/nowcast and to investigate, whether incorporating news topics into the model improves the accuracy of forecasting / nowcasting. In our research we are interested in the following comparison and questions:

(1) ML - Retraining matter. *Whether the retraining procedure in the forecast/nowcast machine learning models improve the errors of predictions?*

We subtract from the appropriate metrics of the specific model **with-out retraining** the same metrics but for the model **with retraining**. If the difference is positive – it means that the errors were improved by the retraining procedure in case of the machine learning models.

(2) TS with retraining - News Indexes matter. *Whether news indexes in the forecast/nowcast time series models with retraining improves the errors of predictions?*

We subtract from the appropriate metrics of the specific model **with retraining** and without news indexes the same metrics but for the model **with news indexes**. If the difference is positive – it means that the errors were improved with the help of news indexes in case of time-series models with retraining.

(3) TS with-out retraining - News Indexes matter. *Whether news indexes in the forecast/nowcast time series models with-out retraining improves the errors of predictions?*

We subtract from the appropriate metrics of the specific model **with-out retraining** and without news indexes the same metrics but for the model **with news indexes**. If the difference is positive – it means that the errors were improved with the help of news indexes in case of time-series models with-out retraining.

(4) TS - Retraining matter. *Whether the retraining procedure in the forecast/nowcast time-series models improve the errors of predictions?*

We subtract from the appropriate metrics of the specific model **with-out retraining** the same metrics but for the model **with retraining**. If the difference is positive – it means that the errors were improved by the retraining procedure in case of the time-series models.

(5) ML versus TS – with news indexes with-out retraining case. *Whether machine learning models with news indexes and with-out retraining are better than time-series models with news indexes and with-out retraining?*

We subtract from the appropriate metrics of the ML model **with-out retraining** the same metrics but for the TS model **with-out retraining**. If the difference is positive – it means that the errors were improved with the help of machine learning procedure (with news indexes and with-out retraining).

(6) ML versus TS – with news indexes with retraining case. *Whether machine learning models with news indexes and with retraining are better than time-series models with news indexes and with retraining?*

We subtract from the appropriate metrics of the ML model **with retraining** the same metrics but for the TS model **with retraining**. If the difference is positive – it means that the errors were improved with the help of machine learning procedure (with news indexes and with retraining).

Estimations (results)

Topic Modelling

In practice, to process topic modeling via LDA we need a **corpus of text**, a **dictionary of lemmas** and desirable **number of topics**.

1. The **corpus of text** was already formed during the data processing procedure;
2. To create **dictionary,** we selected all lemmas which were presented in more than **5% of articles** but less than in **50% of articles**. So that, we excluded the rarest and the most frequent words. Our aim was to include **100,000 words** in dictionary of lemmas;
3. To choose the **optimal number of topics** we implement the coherence value. Topic may be defined as coherent if most relevant terms for the topic frequently co-appear in corpus articles and this co-appearance is not accidental. As we can see below, 38 topics model showed the highest coherence value: **around 0,53.**

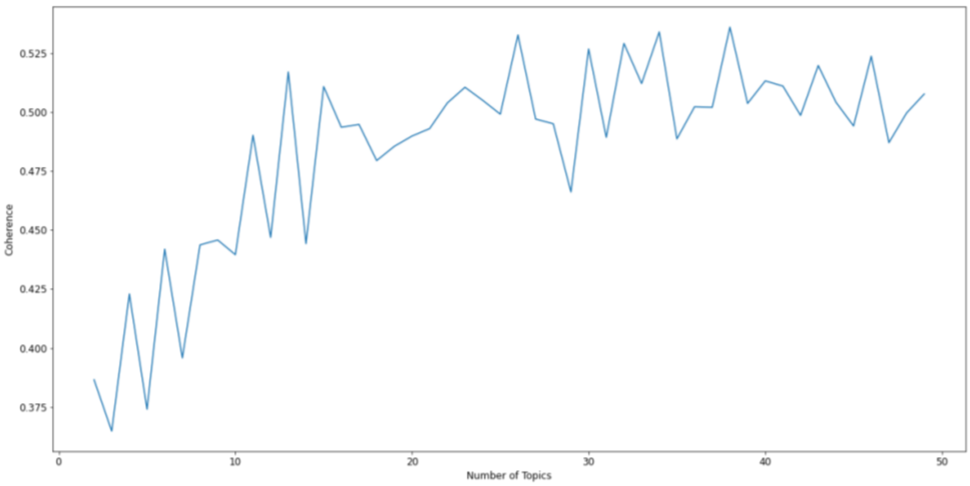


Figure 1. Coherence metrics for LDA models with 2-48 topics

Source: *author's calculations*

4. As the value for **hyperparameters** α and β we use the default values for Gensim.

As a result, we got **38 topics**. The link to the code for topic modelling procedure is presented in the **Appendix E2. Topic Modelling Procedure Code**. News dataset with assigned topics can be found in **Appendix E1. Topic Modelling Dataset**

According to (Lau, Newman, Karimi, & Baldwin, 2010) ten words with the highest weights contain around 30% of all information about the specific topic. Thus, we list such ten words related with each topic in the **Appendix E3. LDA Topic Modeling Table**. We also labeled topics in more “human” way where it was possible. As we can see from the table, the topics can be quite easily identified and labeled by the human language name in most of the cases.

Table #2

Examples of Topics Related to “Economics”, “Finance”, “Banks” (LDA approach)

Computer Language name of the topic	**Human Language** name of the topic	**Key Words** (top 10 words)
Topic #1	**“Нефть”**	0.013*"рынок" + 0.011*"цена" + 0.011*"рост" + 0.009*"нефть" + 0.007*"российский" + 0.007*"страна" + 0.006*"рубль" + 0.006*"сша" + 0.006*"уровень" + 0.005*"экономика"
Topic #7	**“Фондовый рынок”**	0.017*"компания" + 0.010*"рынок" + 0.006*"фонд" + 0.005*"руб" + 0.005*"цена" + 0.004*"млн" + 0.004*"ставка" + 0.004*"банк" + 0.004*"российский" + 0.003*"директор"
Topic #9	**“Центральный Банк”**	0.034*"банк" + 0.016*"млрд" + 0.009*"кредит" + 0.008*"руб" + 0.007*"цб" + 0.006*"ставка" + 0.006*"млн" + 0.006*"кредитный" + 0.006*"рынок" + 0.005*"составлять"
Topic #11	**“США-Россия”**	0.014*"сша" + 0.009*"президент" + 0.008*"страна" + 0.007*"российский" + 0.005*"американский" + 0.005*"отношение" + 0.005*"сторона" + 0.005*"трамп" + 0.004*"мид" + 0.004*"власть"
Topic #20	**“Транспортировка грузов”**	0.009*"самолет" + 0.004*"первый" + 0.004*"сказать" + 0.004*"находиться" + 0.003*"борт" + 0.003*"экипаж" + 0.003*"происходить" + 0.003*"говорить" + 0.003*"корабль" + 0.003*"сша"
Topic #21	**“Недвижимость”**	0.010*"цена" + 0.010*"квартира" + 0.009*"дом" + 0.008*"жилье" + 0.007*"строительство" + 0.007*"недвижимость" + 0.007*"рынок" + 0.007*"компания" + 0.006*"dj" + 0.006*"кв"
Topic #22	**“Россия-Украина”**	0.014*"президент" + 0.008*"выборы" + 0.008*"глава" + 0.007*"партия" + 0.006*"украина" + 0.005*"вопрос" + 0.005*"страна" + 0.004*"решение" + 0.004*"депутат" + 0.004*"принимать"
Topic #29	**“Международный спорт”**	0.008*"первый" + 0.007*"мир" + 0.007*"турнир" + 0.006*"место" + 0.006*"российский" + 0.006*"второй" + 0.005*"олимпийский" + 0.005*"чемпионат" + 0.005*"сборная" + 0.005*"выигрывать"
Topic #34	**“Производство и Предприятия”**	0.012*"завод" + 0.009*"предприятие" + 0.008*"производство" + 0.006*"млн" + 0.005*"рынок" + 0.005*"компания" + 0.005*"продукция" + 0.004*"область" + 0.004*"млрд" + 0.004*"проект"
Topic #35	**“Украина-Газ”**	0.017*"газ" + 0.012*"газпром" + 0.009*"украина" + 0.007*"поставка" + 0.007*"российский" + 0.007*"ес" + 0.006*"страна" + 0.006*"компания" + 0.005*"млрд" + 0.005*"соглашение"

Source: *author’s calculations*

In **Table#2** we represent 10 topics that seems to be most valuable and meaningful in terms of nowcasting/forecasting power. These 10 topics will be used in order to construct 10 news indexes. Each index will illustrate one of these 10 topics. Thus, we get 10 news index variables in order to verify forecasting/nowcasting errors.

We also create and plot the **Intertopic Distance Maps**, that visualize the topics in a two-dimensional space. The blue (red) circle space illustrates the number of words that is related to the topic in framework of our news corpus. The bar chart at the right of the map lists the 30 most relevant words related to the chosen topic and their frequency through the topic and news corpus.

Here in the text, we present one of the most explicit economic topics chosen through Intertopic Distance Map – **Topic #1 "Нефть"**. The names of topics in the **Table #2** and the names of topics on the **Intertopic Distance Map** are different due to the specific execution procedure of the code. Nevertheless, it is still can be identified and interpreted. As we can see from the map, most of the topics don't "intersect" and don't "overlay" each other. It means, that most of the topics contain unique words and related to the specific theme of the news article.

In order to illustrate the modeled topics, we present topic as the **Cloud of Words**.

Figure 3. Cloud of Words for Topic #37 ("Нефть")

Source: *author's calculations*

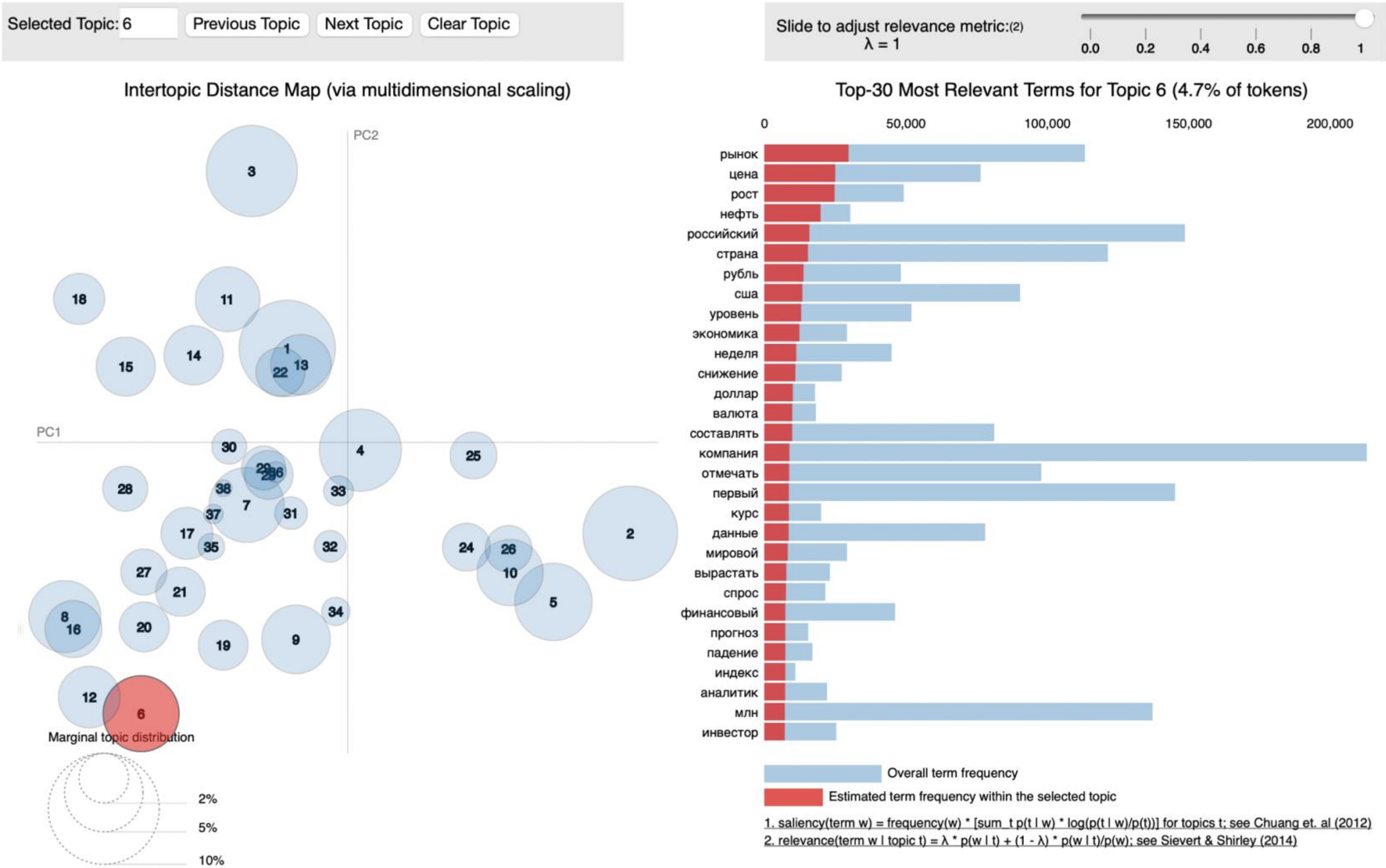


Figure 2. Intertopic Distance Map (Topic #1 "Нефть")

Source: *author's calculations*

News Indexes

Following the procedure described in the **Methodology** we have created the Frequency News Index and Sentiment News Index on monthly and quarterly basis for both topic classification (LDA and K-Means Clustering). The link to the code for Frequency News Index calculating is presented in the **Appendix F2. Frequency News Index Calculation Code**. The link to the code for Sentiment News Index calculating is presented in the **Appendix F5. Sentiment News Index Calculation Code**.

Due to the fact, that in our analysis we consider long time period (16 years) we can't represent the table with all the calculated indexes (even on quarterly basis) here in the text. So that, we apply the links to both datasets: see **Appendix F1. Frequency News Index Dataset** and **Appendix F4. Sentiment News Index Dataset**, respectively.

Because of dealing with two topic specifications and considering index for each of the topic separately the one single time-series graph seems to be unreadable and hard to understand. So that, we plot several pallets of all the graphs for each of the index, for each of the topic separately on the quarterly basis. See **Appendix F3. Frequency News Index Graphs** and **Appendix F6. Sentiment News Index Graphs**, respectively.

Forecasting and nowcasting: comparison and results

Due to high dimensional tables with estimations, we are not going to present all the tables with the results by each macroeconomics variable but provide the summary description. All the tables can be provided on request.

(1) **ML - Retraining matter.** *Whether the retraining procedure in the forecast/nowcast machine learning models improve the errors of predictions?*

The positive improvement in errors (lower errors) was obtained in case of all the macro variables for all the model specification except linear model, elastic net and XGBOOST for 1 lag.

The result means that retraining procedure matters for the Machine Learning models and allows to get lower errors of forecasting/nowcasting.

(2) **TS with retraining - News Indexes matter.** *Whether news indexes in the forecast/nowcast time series models with retraining improves the errors of predictions?*

The negative improvement in errors (higher errors) was obtained in case of all the macro variables for all the model specification.

The result means that the inclusion of news indexes in the retrained time-series model increases the forecast/nowcast errors. So that, news indexes don't allow to improve forecast/nowcast.

(3) **TS with-out retraining - News Indexes matter.** *Whether news indexes in the forecast/nowcast time series models with-out retraining improves the errors of predictions?*

The negative improvement in errors (higher errors) was obtained in case of all the macro variables for all the model specification.

The result means that the inclusion of news indexes in the non-retrained time-series model increases the forecast/nowcast errors. So that, news indexes don't allow to improve forecast/nowcast.

(4) **TS - Retraining matter.** *Whether the retraining procedure in the forecast/nowcast time-series models improve the errors of predictions?*

For **BCI** and **CCI** the retraining procedure brings higher errors than non-retraining models for all the specifications with indexes. That means, that there is no need in retraining in case of news index inclusion. Moreover, the same result is obtained for time-series models with-out new indexes (3 and 6 lags models) except 1 lag model. So that the standard AR(1) process performs better after retraining procedure.

For **CPI, IM** and **NX** the retraining procedure with news indexes brings lower errors that non-retraining one in case of forecasting but not in nowcasting. However, for the sentiment index the retraining procedure improves the nowcasting, too.

For **EX** the retraining procedure with news indexes brings lower errors that non-retraining one in case of forecasting and nowcasting.

The result means that for most of the considered macroeconomic variables the retraining procedure allows to improve the forecasting, but not the nowcasting.

(5) **ML versus TS – with news indexes with-out retraining case.** *Whether machine learning models with news indexes and with-out retraining are better than time-series models with news indexes and with-out retraining?*

Almost for all the cases at least one ML model with indexes outperform the time-series model with indexes with the same number of lags. In most of the cases RIDGE and LASSO models were the best one.

(6) **ML versus TS – with news indexes with retraining case.** *Whether machine learning models with news indexes and with retraining are better than time-series models with news indexes and with retraining?*

Almost for all the cases ML model with indexes outperform the time-series model with indexes with the same number of lags. The result is quite obvious due to the result from (4) – retraining procedure increases the errors for time-series models.

Finally, we can compare all the models (96 models for each of the variable). See **Appendix H** for more details.

(1) Considering **BCI** variable **(Appendix H1)**, we can note that the best forecasting and nowcasting model specification is simple AR(3) process, except the case with Sentiment index. In the latter case Lasso model with 6 lags with sentiment index outperform all AR(3) process.
(2) Considering **CCI** variable **(Appendix H2)**, we can note that the best forecasting and nowcasting model specification is Elastic Net with 6 lags with-out retraining procedure for Frequency index of all the lists specification, but not for the Sentiment Index.
(3) Considering **CPI** variable **(Appendix H3)**, we can note that the best forecasting and nowcasting model specification is simple AR(1) process.
(4) Considering **EX** variable **(Appendix H4)**, we can note that the best forecasting and nowcasting models for any news index are several specifications of machine learning models (Lasso, XGBOOST, Random Forest) and no AR process at the top 10 models in the list.
(5) Considering **IM** variable **(Appendix H5)**, we can note that the best forecasting and nowcasting models for any news index (except Optimal Frequency Index) are several specifications of machine learning models (Lasso, XGBOOST, Random Forest, Elastic Net) and no AR process at the top 10 models in the list (except ARXF(1) for the Frequency Index Optimal)
(6) Considering **NX** variable **(Appendix H6)**, we can note that the best forecasting and nowcasting models for any news index (except Sentiment Index) are several specifications of machine learning models (Lasso, XGBOOST, Random Forest, Elastic Net) and no AR process at the top 10 models in the list (except ARXN(3) and ARXN(1) for the Sentiment Index)

Thus, we can state that for some of the chosen macroeconomic variables (CCI, EX, IM, NX) machine learning models with news indexes sufficiently improves the forecasting/nowcasting power.

Further Work

In our work we consider only the one Russian news source. Such an approach may lead to bias results. So that, the further extension of the analysis may be the inclusion and consideration of several news sources, such as “РБК”, “Ведомости”, “Интерфакс”, etc. This allows to catch the different style of writing the news, different political views and different audiences. It can be scientifically valuable to analyze the model where several news resources are included with relative weights. Thus, the effect of different news sources on forecast can be evaluated and verified.

Moreover, we have used only one basic topic modelling procedure: LDA. There exists one more advanced method – LDA2VEC, which performs better in topic modelling procedure. So, the implementation of such method and comparison with the LDA results can be the relevant extension of the current paper.

In addition, in our paper we didn’t consider the structural relationship between the variables and the structure of the economy. Thus, it can be quite important to replicate the (Larsen & Thorsrud, 2019) approach related to full-sample SVAR and out-of sample SVAR forecasting in the framework of Russian news corpuses.

Finally, in our analysis we consider only a limited range of macroeconomic variables. So that the further work can be done in the widening the range of chosen macroeconomic variables (for example GDP, Investment, Consumption etc.). We also didn’t consider such a variable as Total Factor Productivity **(TFP)** due to the non-trivial computation process and lack of relevant calculated data in Russian datasets. Moreover, there is no monthly or quarterly data for TFP variable. So that, the only yearly analysis is allowed. In this case the dataset of news should be quite large (around 40 years or more), that seems quite impossible in case of Russia history.

Conclusion

In recent years the idea of predicting and explaining the economic fluctuations with the use of news data has become more popular in research papers. In our research we have made an attempt to investigate whether the predictive power of forecasts for macroeconomic variables can be improved with the use of news data in Russia. The **novelty** of the study is the development Russian specific approach to model the topics and implement it on Russia case. As soon as macroeconomic aggregates are often revised in Russia the prediction in such conditions seems to be a challenge.

We have chosen the newspaper that concern economic, finance and development issues and related topics – "Коммерсантъ" as the source of news articles. All the articles published in the period of 2010-2020 were extracted. Then the appropriate data processing procedures were conducted. We have replicated the method commonly used in the related articles: LDA (Yakovleva, 2018), (Larsen & Thorsrud, 2019), (Seleznev & Mamedli, 2020) in order to proceed with topic modelling. As the result 38 topics were obtained and most of them was labeled in human language, meaning the good and nice readability and interpretability. Finally, only 10 meaningful topics were chosen for further analysis.

Frequency and Sentiment Indexes were built for these 10 chosen topics and used for the forecasting several macroeconomic variables. We have run 96 model specifications for each of the macroeconomic variable. Our analysis revealed that for some of the chosen macroeconomic variables (CCI, EX, IM, NX) machine learning models with news indexes sufficiently improves the forecasting/nowcasting power relative to classical time-series models, even time-series with the same news indexes.

References

Algaba, A., Ardia, D., Bluteau, K., Borms, S., & Boudt, K. (2020). Econometrics Meets Sentiment: An Overview of Methodology and Aplications. *Journal of Economic Surveys, 34*, pp. 512-547.

Alsing, O., & Bahceci, O. (2015). Stock Market Prediction using Social Media Analysis. *KTH Royal Institute of Technology*.

Antweiler, W., & Frank, M. Z. (2004). Is all that talk just noise? the information content of internet stock message boards. *The Journal of Finance, 59*, pp. 1259–1294.

Ardia, D., Bluteau, K., & Boudt, K. (2019). Questioning the news about economic growth: Sparse forecasting using thousands of news-based sentiment values. *International Journal of Forecasting*, pp. 1370-1386.

Audrino, F., Sigrist, F., & Ballinari, D. (2020). The Impact of Sentiment and Attention MEasures on Stock Market Volatility. *International Journal of Forecasting, 36*, pp. 334-357.

Barsky, R. B., & Sims, E. R. (2011). News shocks and business cycles. *Journal of monetary Economics, 58(3)*, pp. 273-289.

Barsky, R. B.; Sims, E. R. (2012). Information, animal spirits, and the meaning of innovations in consumer confidence. *American Economic Review, 102(4)*, pp. 1343-1377.

Beaudry, P., & Portier, F. (2006). Stock prices, news, and economic fluctuations. *American Economic Review, 96(4)*, pp. 1293-1307.

Beaudry, P., & Portier, F. (2014). News-driven Business Cycles: Insights and Challneges. *Journal of Economic Literature, 52*, pp. 993-10074.

Blanchard, O. J., L'Huillier, J. P., & Lorenzoni, G. (2013). News, noise, and fluctuations: An empirical exploration. *American Economic Review, 103(7)*, pp. 3045-3070.

Blei, D., Ng, A., & Jordan, M. (2003). Latent Dirichlet Allocation. *Journal of Machinne Learning Research*, pp. 993-1022.

Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computer Science, 2(1)*, pp. 1-8.

Bordalo, P., Gennaioli, N., Ma, Y., & Shleifer, A. (2020). Overreeaction in Macroeconomic Expectations. *American Economic Review, 110*, pp. 2748-2782.

Caporin, M., & Poli, F. (2017). Building news measures from textual data and an application to volatility forecasting. *Econometrics, 5(3)*.

Choi, H., & Varian, H. (2012). Predicting the present with Google Trends. *Economic record, 88*, pp. 2-9.

Coibion, O., & Gorodnichenko, Y. (2015). Information Rigidity and the Expectations Formation Process: A Simple Framework and New Facts. *American Economic Review, 105,* pp. 2644-2678.

Da, Z., Engelberg, J., & Gao, P. (2011). In search of attention. *The Journal of Finance*, *66,* pp. 1461–1499.

Davison, L. H. (2010). Empirical Study of Topic Modeling in Twitter. pp. 80-88.

D'Amuri, F., & Marcucci, J. (2017). The predictive power of Google searches in forecasting US unemployment. *International Journal of Forecasting*, *33(4),* pp. 801-816.

Feuerriegel, S., & Gordon, J. (2019). News-based forecasts of macroeconomic indicators: A semantic path model for interpretable predictions. *European Journal of Operational Research, 272(1)*, pp. 162-175.

Forni, M., Gambetti, L., Lippi, M., & Sala, L. (2017). Noisy news in business cycles. *American Economic Journal: Macroeconomics, 9(4)*, pp. 122-152.

Goshima, K., Ishijima, H., Shintani, M., & Yamamoto, H. (2019). *Forecasting Japanese Inflation with a News-based Leading Indicator of Economic Activities.* Center for Advanced Research in Finance.

Joseph, K., Wintoki, M. B., & Zhang, Z. (2011). Forecasting abnormal stock returns and trading volume using investor sentiment: Evidence from online search. *International Journal of Forecasting, 27*, pp. 1116-1127.

Kalamara, E., Turrell, A., Redl, C., Kapetanios, G., & Kapadia, S. (2020). *Making Text Count: Economic Forecasting Using Newspaper Text.* Bank of England.

Larsen, V. H., & Thorsrud, L. A. (2019). The Value of News for Economic Developments. *Journal of Econometrics, 210*, pp. 203-2018.

Lau, J., Newman, D., Karimi, S., & Baldwin, T. (2010). 23rd International Conference on Computational Linguistics. *Best Topic Word Selection for Topic Labelling.* Stroudsburg.

McLaren, N., & Shanbhogue, R. (2011). Using internet search data as economic indicators. *Bank of England Quarterly Bulletin*.

Nyman, R., Ormerod, P., Smith, R., & Tuckett, D. (2014). ECB Workshop on Big Data for Forecasting and statistics. *Big Data and Economic Forecasting: A Top-Down Approach Using Directed Algorithmic Text Analysis.*

Seleznev, S., & Mamedli, M. (2020). XXI April International Academic Conference on Economic and Social Development. *What can we learn from the news?*

Shapiro, A. H., Sudhof, M., & Wilson, D. (2020). *Measuring News Sentiment.* Federal Reserve Bank of San Francisco.

Sun, L., Najand, M., & Shen, J. (2016). Stock return predictability and investor sentiment: A high-frequency perspective. *Journal of Banking & Finance, 73*, pp. 147–164.

Thorsrud, A. (2016). *Words are the New Numbers: A Newsy Coincident Index of Business Cycles.* Norges Bank Research.

Tobback, E., Naudts, H., Daelemans, W., Junque de Fortuny, E., & Martens, D. (2018). Belgian Economic Policy Uncertainty Index: Improvement through Text Mining. *International Journal of Forecasting, 34*, pp. 355-365.

Ulyankin, F. (2020). Forecasting Russian Macroeconomic Indicators Based on Information from News and Search Queries. *Russian Journal of Money and Finance, 79(4)*, pp. 75-97.

Yakovleva, K. (2018). Text Mining-based Economic Activity Estimation. *Ruussian Journal of Money and Finance*, *77*, pp. 26-41.

Голощапова, И. О., & Андреев, М. Л. (2017). Оценка инфляционных ожиданий российского населения методами машинного обучения. *Вопросы экономики, 6*, стр. 71-93.

Столбов, М. (2011). Статистика поиска в Google как индикатор финансовой конъюнктуры. *Вопросы экономики*, стр. 79-93.

Appendix A

Appendix A1. News Data

Link to the raw news datasets for coding procedure:

https://drive.google.com/drive/folders/1BvKUTpL4MuEK9HhecLiVgs2-YC1RPxFw?usp=sharing

Link to the raw news dataset for viewing:

https://drive.google.com/drive/folders/1BvGctQ-YXNs2TnEeqJOc4qLiIysHqABw?usp=sharing

Appendix A2. Parsing Code

Link to the code for parsing (.ipynb):

https://drive.google.com/drive/folders/1BzlMX0ZwyRkdQ0axFwYu5fFbMVft9yNi?usp=sharing

Appendix B

Appendix B1. Processed News Data

Link to the processed news dataset for coding procedure:

https://drive.google.com/drive/folders/1OrGizDkowR2uZGr65AM35JeMpN_WCooO?usp=sharing

Link to the processed news dataset for viewing:

https://drive.google.com/drive/folders/1cDyoZsn2Un0VVwxXgkW056CmQ4NqURGv?usp=sharing

Appendix B2. Processing Code

Link to the code for news data processing (.ipynb):

https://drive.google.com/drive/folders/1C2T-7s6ktBdjmD6Yn70MdAePUemj9aBL?usp=sharing

Appendix C

Appendix C1. Numerical (Macro) Data

Link to the raw news datasets for viewing:

https://drive.google.com/drive/folders/1udSfEO1FBvwxOsqJ31QrWSxDe9RkWhog?usp=sharing

Appendix C2. Numerical Data Description Table

Appendix table #1

Numerical Data Description

Variable	Frequency	Features	Data source	Definition
CCI	monthly	points; seasonally adjusted by the original source	OECD Database	This consumer confidence indicator provides an indication of future developments of households' consumption and saving, based upon answers regarding their expected financial situation, their sentiment about the general economic situation, unemployment and capability of savings.
BCI	monthly	points; seasonally adjusted by the original source	OECD Database	Business Confidence Index (BCI) is based on a monthly opinion survey and covers opinion on developments in production, orders, and stocks of finished goods in the industry sector.
Inflation	monthly	percent	OECD Database	Inflation is measured by Consumer Price Index, including inflation from energy sources.
Export	monthly	in real term; in rubles; seasonally adjusted by the original source	IMF Database	
Import	monthly	in real term; in rubles; seasonally adjusted by the original source	IMF Database	
Net Export	monthly	in real term; in rubles; seasonally adjusted by the original source	IMF Database	

Source: *IMF Database, OECD Database*

Appendix D

Appendix D1. The List of Descriptors for Frequency News Index

Appendix table #2

List of Descriptors for Frequency News Index [basic]

Positive	Negative
"устойчивый", "устойчивость", "стабильный", "стабильность", "уверенность", "уверенный", "благоприятный", "улучшение", "рост", "подъем", "подъём", "положительный", "прогресс", "развитие", "позитивный", "спокойствие", "выигрывать", "расти", "подниматься"	"устойчивый", "устойчивость", "стабильный", "стабильность", "уверенность", "уверенный", "благоприятный", "улучшение", "рост", "подъем", "подъём", "положительный", "прогресс", "развитие", "позитивный", "спокойствие", "выигрывать", "расти", "подниматься"

Source: *author's calculations*

Appendix table #3

List of Descriptors for Frequency News Index [optimal]

Positive	Negative
"безопасность", "безопасный", "бережливость", "бережливый", "благоприятный", "богатство", "богатый", "возобновлять", "восполнять", "выздоравливать", "выздоровление", "договариваться", "договориться", "доступный", "защита", "компенсация", "компенсировать", "крепкий", "льготы", "надёжность", "надежный", "надёжный", "невиновный", "облегчать", "облегчение", "облегчить", "обновление", "обновлять", "ожидать", "оздоровление", "оптимистичный", "подниматься", "подъем", "подъём", "позитивный", "положительный", "помогать", "помощь", "пополнение", "порядок", "пособие", "прогресс", "прогрессивный", "продвигать", "продвижение", "прочный", "развитие", "расти", "рост", "сохранение", "сохранять", "спокойствие", "стабильность", "стабильный", "страхование", "страховка",	"банкрот", "банкротство", "бедность", "бедный", "безденежный", "безденежье", "безнадежный", "безнадёжный", "безработный", "беспокойство", "беспомощность", "беспомощный", "беспорядки", "волнение", "волноваться", "вторгаться", "вторжение", "вымогательство", "вымогать", "девальвация", "дезинформация", "дезинформировать", "дезорганизация", "дезорганизовать", "депрессивный", "депрессия", "дефолт", "дисбаланс", "дно", "долг", "должник", "клевета", "коррупция", "кризис", "кризисный", "криминал", "криминальный", "небезопасность", "небезопасный", "негативный", "недоступный", "неустойчивость", "неустойчивый", "обанкротиться", "обеспокоенность", "обеспокоить", "обеспокоиться", "ограничение", "опасность", "опасный", "ослабление", "отрицательный", "отставка",

"строительство", "строить", "уверенность", "уверенный", "улучшение", "устойчивость", "устойчивый", "энергичный"	"падать", "падение", "паника", "пессимистический", "пессимистичный", "понижение", "потери", "прекращение", "разрушать", "разрушение", "разрушительный", "рецессия", "санкции", "сдерживание", "сдерживать", "скандал", "скандальный", "слабый", "сокращение", "спад", "тревога", "тревожный", "увольнение", "утекать", "утечка", "утрачивать", "ухудшение", "ущерб", "фальсификация", "фальсифицировать", "штрафы"

Source: *author's calculations*

Appendix table #4

List of Descriptors for Frequency News Index [full]

Positive	Negative
"баланс", "безопасность", "безопасный", "бережливость", "бережливый", "благоприятный", "богатство", "богатый", "бодрить", "бодрость", "большой", "веселить", "весело", "веселье", "влюбляться", "возобновлять", "восполнять", "восторженный", "выздоравливать", "выздоровление", "выигрывать", "высокий", "герой", "героический", "добро", "доброта", "добрый", "договариваться", "договориться", "доступный", "друг", "дружба", "забота", "заботиться", "защита", "компенсация", "компенсировать", "красивый", "красота", "крепкий", "льготы", "любовь", "мир", "надёжность", "надежный", "надёжный", "невиновный", "облегчать", "облегчение", "облегчить", "обновление", "обновлять", "одобрение", "ожидать", "оздоровление", "оптимистичный", "память", "победа", "побеждать", "подниматься", "подъем", "подъём", "позитивный", "положительный", "помогать", "помощь", "пополнение", "порядок", "пособие", "прогресс", "прогрессивный", "продвигать", "продвижение", "прочный", "развитие", "расти", "рост", "семейный", "семья", "сильный", "сохранение", "сохранять", "спокойствие", "стабильность",	"агрессивный", "агрессия", "банкрот", "банкротство", "бедность", "бедный", "безденежный", "безденежье", "безнадежный", "безнадёжный", "безнаказанность", "безнаказанный", "безработный", "беспокойство", "беспомощность", "беспомощный", "беспорядки", "бракованный", "взламывать", "взлом", "взрыв", "возбуждать", "война", "волнение", "волноваться", "вранье", "врать", "вторгаться", "вторжение", "вымогательство", "вымогать", "гибнуть", "гнев", "девальвация", "дезинформация", "дезинформировать", "дезорганизация", "дезорганизовать", "депрессивный", "депрессия", "дефолт", "дисбаланс", "дно", "долг", "должник", "жадность", "жадный", "загрязнение", "загрязнять", "злиться", "злой", "злость", "избивать", "избить", "клевета", "коррупция", "кризис", "кризисный", "криминал", "криминальный", "маленький", "небезопасность", "небезопасный", "негативный", "недоступный", "ненависть", "неустойчивость", "неустойчивый", "низкий", "обанкротиться", "обеспокоенность", "обеспокоить", "обеспокоиться", "огорчаться", "огорчение", "огорчить", "ограничение",

"стабильный", "страхование", "страховка", "строительство", "строить", "уверенность", "уверенный", "улучшение", "устойчивость", "устойчивый", "энергичный"	"опасность", "опасный", "оскорбить", "оскорбление", "оскорблять", "ослабление", "отрицательный", "отставка", "падать", "падение", "паника", "пессимистический", "пессимистичный", "плохой", "погибнуть", "подделка", "подделывать", "понижение", "поражение", "потери", "прекращение", "преступление", "прогреметь", "проигрывать", "проигрыш", "разрушать", "разрушение", "разрушительный", "рецессия", "самоубийство", "санкции", "сдерживание", "сдерживать", "скандал", "скандальный", "слабый", "смерть", "сокращение", "спад", "суицид", "тревога", "тревожный", "убивать", "убийство", "увольнение", "умереть", "умирать", "утекать", "утечка", "утрачивать", "ухудшение", "ущемлять", "ущерб", "фальсификация", "фальсифицировать", "штрафы"

Appendix E

Appendix E1. Topic Modeling Dataset

Link to the dataset for coding procedure:

https://drive.google.com/drive/folders/1ORxPb9MGklS2MqvyycFKC1CS4JYKf4Ly?usp=sharing

Link to the dataset to view:

https://drive.google.com/drive/folders/1xehrjtuN8V6H2o74zmH8O4KScVMg6hFJ?usp=sharing

Appendix E2. Topic Modeling Procedure Code

Link to the code for topic modelling (.ipynb):

https://drive.google.com/drive/folders/1C39JYHbVR8Pw3Yy8cUcQLKCeTdrvw0ag?usp=sharing

Appendix E3. LDA Topic Modeling Table

Appendix Table #5

Generated topics and related words (LDA approach)

Computer Language name of the topic	**Human Language** name of the topic	**Key Words** (top 10 words)
Topic #0	**“Арбитражное судопроизводство”**	0.034*"суд" + 0.027*"дело" + 0.022*"область" + 0.018*"арбитражный" + 0.018*"ооо" + 0.018*"решение" + 0.013*"отношение" + 0.013*"инн" + 0.011*"конкурсный" + 0.010*"огрн"
Topic #1	**“Нефть”**	0.013*"рынок" + 0.011*"цена" + 0.011*"рост" + 0.009*"нефть" + 0.007*"российский" + 0.007*"страна" + 0.006*"рубль" + 0.006*"сша" + 0.006*"уровень" + 0.005*"экономика"
Topic #2	**“Сети и Технологии”**	0.010*"компания" + 0.005*"система" + 0.005*"рынок" + 0.004*"сеть" + 0.004*"проект" + 0.004*"работать" + 0.004*"пользователь" + 0.004*"технология" + 0.003*"оператор" + 0.003*"говорить"
Topic #3	**“Судебное дело”**	0.020*"суд" + 0.011*"дело" + 0.006*"право" + 0.005*"решение" + 0.005*"адвокат" + 0.004*"иск" + 0.004*"судебный" + 0.004*"судья" + 0.004*"уголовный" + 0.004*"нарушение"
Topic #4	**“Образование”**	0.008*"млн" + 0.006*"вуз" + 0.004*"первый" + 0.004*"торги" + 0.004*"образование" + 0.003*"работа" + 0.003*"школа" + 0.003*"компания" + 0.003*"продажа" + 0.003*"руб"
Topic #5	**“Банки и кредиты #1”**	0.022*"банк" + 0.006*"акция" + 0.005*"сбербанк" + 0.005*"кредит" + 0.005*"компания" + 0.004*"млн" + 0.004*"сделка" + 0.003*"заемщик" + 0.003*"млрд" + 0.003*"актив"
Topic #6	**“Федеральное регулирование”**	0.007*"закон" + 0.007*"руб" + 0.005*"проверка" + 0.005*"федеральный" + 0.005*"орган" + 0.004*"организация" + 0.004*"нарушение" + 0.004*"дело" + 0.004*"документ" + 0.004*"правительство"
Topic #7	**“Фондовый рынок #1”**	0.017*"компания" + 0.010*"рынок" + 0.006*"фонд" + 0.005*"руб" + 0.005*"цена" + 0.004*"млн" + 0.004*"ставка" + 0.004*"банк" + 0.004*"российский" + 0.003*"директор"
Topic #8		0.015*"руб" + 0.006*"дело" + 0.005*"область" + 0.004*"компания" + 0.004*"соус" + 0.004*"решение" + 0.003*"суд" + 0.003*"ооо" + 0.003*"фонд" + 0.003*"банк"
Topic #9	**“Центральный Банк”**	0.034*"банк" + 0.016*"млрд" + 0.009*"кредит" + 0.008*"руб" + 0.007*"цб" + 0.006*"ставка" + 0.006*"млн" + 0.006*"кредитный" + 0.006*"рынок" + 0.005*"составлять"
Topic #10	**“Фондовый рынок #2”**	0.017*"компания" + 0.010*"млрд" + 0.007*"руб" + 0.007*"рынок" + 0.005*"млн" + 0.005*"автомобиль" + 0.005*"банк" + 0.004*"объем" + 0.004*"акция" + 0.004*"долг"

Topic #11	**“США-Россия”**	0.014*"сша" + 0.009*"президент" + 0.008*"страна" + 0.007*"российский" + 0.005*"американский" + 0.005*"отношение" + 0.005*"сторона" + 0.005*"трамп" + 0.004*"мид" + 0.004*"власть"
Topic #12	**“Оборона и Армия”**	0.013*"военный" + 0.005*"страна" + 0.005*"сила" + 0.004*"российский" + 0.004*"армия" + 0.004*"войско" + 0.004*"президент" + 0.004*"оборона" + 0.004*"власть" + 0.004*"город"
Topic #13		0.007*"компания" + 0.006*"область" + 0.005*"руб" + 0.005*"млн" + 0.005*"уголь" + 0.004*"ооо" + 0.004*"дело" + 0.004*"суд" + 0.003*"группа" + 0.003*"первый"
Topic #14	**“Происшествия Москвы”**	0.022*"улица" + 0.010*"площадь" + 0.009*"москва" + 0.008*"город" + 0.007*"пожар" + 0.006*"центр" + 0.006*"здание" + 0.005*"митинг" + 0.005*"власть" + 0.004*"дом"
Topic #15	**“Уголовные нарушения”**	0.014*"дело" + 0.007*"уголовный" + 0.006*"сотрудник" + 0.006*"суд" + 0.006*"следствие" + 0.005*"задерживать" + 0.005*"следственный" + 0.004*"мвд" + 0.004*"находиться" + 0.004*"ст"
Topic #16	**“Авиаперевозки”**	0.011*"аэропорт" + 0.008*"рейс" + 0.007*"авиакомпания" + 0.006*"пассажир" + 0.004*"транспорт" + 0.004*"самолет" + 0.004*"российский" + 0.004*"компания" + 0.003*"билет" + 0.003*"перевозчик"
Topic #17	**“Банковское дело”**	0.007*"говорить" + 0.006*"вопрос" + 0.006*"сказать" + 0.005*"банк" + 0.005*"путин" + 0.004*"должный" + 0.004*"страна" + 0.004*"проблема" + 0.004*"деньги" + 0.004*"считать"
Topic #18	**“Мода и одежда”**	0.008*"коллекция" + 0.007*"одежда" + 0.004*"компания" + 0.004*"костюм" + 0.004*"платье" + 0.003*"марка" + 0.003*"сезон" + 0.003*"мода" + 0.003*"обувь" + 0.003*"продажа"
Topic #19	**“Досуг”**	0.004*"день" + 0.003*"большой" + 0.002*"говорить" + 0.002*"работа" + 0.002*"первый" + 0.002*"ресторан" + 0.002*"каждый" + 0.002*"дом" + 0.002*"место" + 0.002*"сделать"
Topic #20	**“Транспортировка грузов”**	0.009*"самолет" + 0.004*"первый" + 0.004*"сказать" + 0.004*"находиться" + 0.003*"борт" + 0.003*"экипаж" + 0.003*"происходить" + 0.003*"говорить" + 0.003*"корабль" + 0.003*"сша"
Topic #21	**“Недвижимость”**	0.010*"цена" + 0.010*"квартира" + 0.009*"дом" + 0.008*"жилье" + 0.007*"строительство" + 0.007*"недвижимость" + 0.007*"рынок" + 0.007*"компания" + 0.006*"dj" + 0.006*"кв"
Topic #22	**“Россия-Украина”**	0.014*"президент" + 0.008*"выборы" + 0.008*"глава" + 0.007*"партия" + 0.006*"украина" + 0.005*"вопрос" + 0.005*"страна" + 0.004*"решение" + 0.004*"депутат" + 0.004*"принимать"
Topic #23	**“Госкоспании”**	0.012*"акция" + 0.011*"директор" + 0.011*"компания" + 0.009*"совет" + 0.007*"навальный" + 0.007*"оао" + 0.006*"область" + 0.005*"губернатор" + 0.005*"акционер" + 0.005*"правительство"

Topic #24	**“Выборы”**	0.009*"глава" + 0.008*"депутат" + 0.007*"выборы" + 0.006*"губернатор" + 0.006*"мэр" + 0.005*"единый" + 0.005*"партия" + 0.005*"дело" + 0.005*"область" + 0.004*"александр"
Topic #25	**“Кино и театр”**	0.012*"театр" + 0.010*"режиссер" + 0.009*"франция" + 0.009*"сша" + 0.008*"роль" + 0.006*"кино" + 0.006*"звезда" + 0.006*"формула" + 0.006*"октябрь" + 0.005*"германия"
Topic #26	**“Спорт”**	0.015*"матч" + 0.011*"команда" + 0.009*"клуб" + 0.008*"чемпионат" + 0.007*"сборная" + 0.007*"первый" + 0.006*"счет" + 0.005*"игра" + 0.005*"лига" + 0.005*"второй"
Topic #27	**“Пандемия”**	0.009*"москва" + 0.005*"пандемия" + 0.005*"вакцина" + 0.005*"работа" + 0.004*"случай" + 0.004*"строительство" + 0.003*"город" + 0.003*"российский" + 0.003*"страна" + 0.003*"число"
Topic #28	**“Фильмы”**	0.006*"фильм" + 0.004*"первый" + 0.004*"жизнь" + 0.003*"история" + 0.003*"хороший" + 0.003*"главный" + 0.003*"говорить" + 0.002*"мир" + 0.002*"большой" + 0.002*"герой"
Topic #29	**“Международный спорт”**	0.008*"первый" + 0.007*"мир" + 0.007*"турнир" + 0.006*"место" + 0.006*"российский" + 0.006*"второй" + 0.005*"олимпийский" + 0.005*"чемпионат" + 0.005*"сборная" + 0.005*"выигрывать"
Topic #30	**“Медицина”**	0.013*"ребенок" + 0.007*"пациент" + 0.007*"врач" + 0.007*"медицинский" + 0.006*"компания" + 0.006*"дело" + 0.005*"страховой" + 0.004*"область" + 0.004*"больница" + 0.004*"руб"
Topic #31	**“Проекты и бюджеты фирм”**	0.015*"млрд" + 0.009*"руб" + 0.008*"бюджет" + 0.008*"компания" + 0.006*"составлять" + 0.005*"млн" + 0.005*"проект" + 0.004*"доход" + 0.004*"расход" + 0.004*"рынок"
Topic #32	**“Управление компаний”**	0.010*"ооо" + 0.010*"дело" + 0.007*"отношение" + 0.007*"проект" + 0.007*"компания" + 0.006*"управляющий" + 0.006*"ас" + 0.006*"вводить" + 0.006*"нп" + 0.005*"определение"
Topic #33	**“Сделки компаний”**	0.024*"компания" + 0.017*"млн" + 0.008*"млрд" + 0.005*"проект" + 0.005*"составлять" + 0.005*"руб" + 0.005*"сделка" + 0.005*"крупный" + 0.004*"сеть" + 0.004*"рынок"
Topic #34	**“Производство и Предприятия”**	0.012*"завод" + 0.009*"предприятие" + 0.008*"производство" + 0.006*"млн" + 0.005*"рынок" + 0.005*"компания" + 0.005*"продукция" + 0.004*"область" + 0.004*"млрд" + 0.004*"проект"
Topic #35	**“Украина-Газ”**	0.017*"газ" + 0.012*"газпром" + 0.009*"украина" + 0.007*"поставка" + 0.007*"российский" + 0.007*"ес" + 0.006*"страна" + 0.006*"компания" + 0.005*"млрд" + 0.005*"соглашение"
Topic #36	**“Торги и Аукционы”**	0.030*"торги" + 0.017*"имущество" + 0.015*"продажа" + 0.015*"аукцион" + 0.013*"открытый" + 0.013*"форма" + 0.012*"организатор" + 0.011*"проведение" + 0.011*"управляющий" + 0.010*"конкурсный"
Topic #37		0.005*"судно" + 0.004*"российский" + 0.004*"компания" + 0.004*"facebook" + 0.004*"реклама" + 0.003*"получать" + 0.003*"сайт" + 0.003*"млн" + 0.003*"власть" + 0.003*"первый"

Source: *author’s calculations*

Appendix F

Appendix F1. Frequency News Index Dataset

Link to the dataset for coding procedure:

https://drive.google.com/drive/folders/1LYYmPAQka4nv956D6GieY2j7VmU7ypjm?usp=sharing

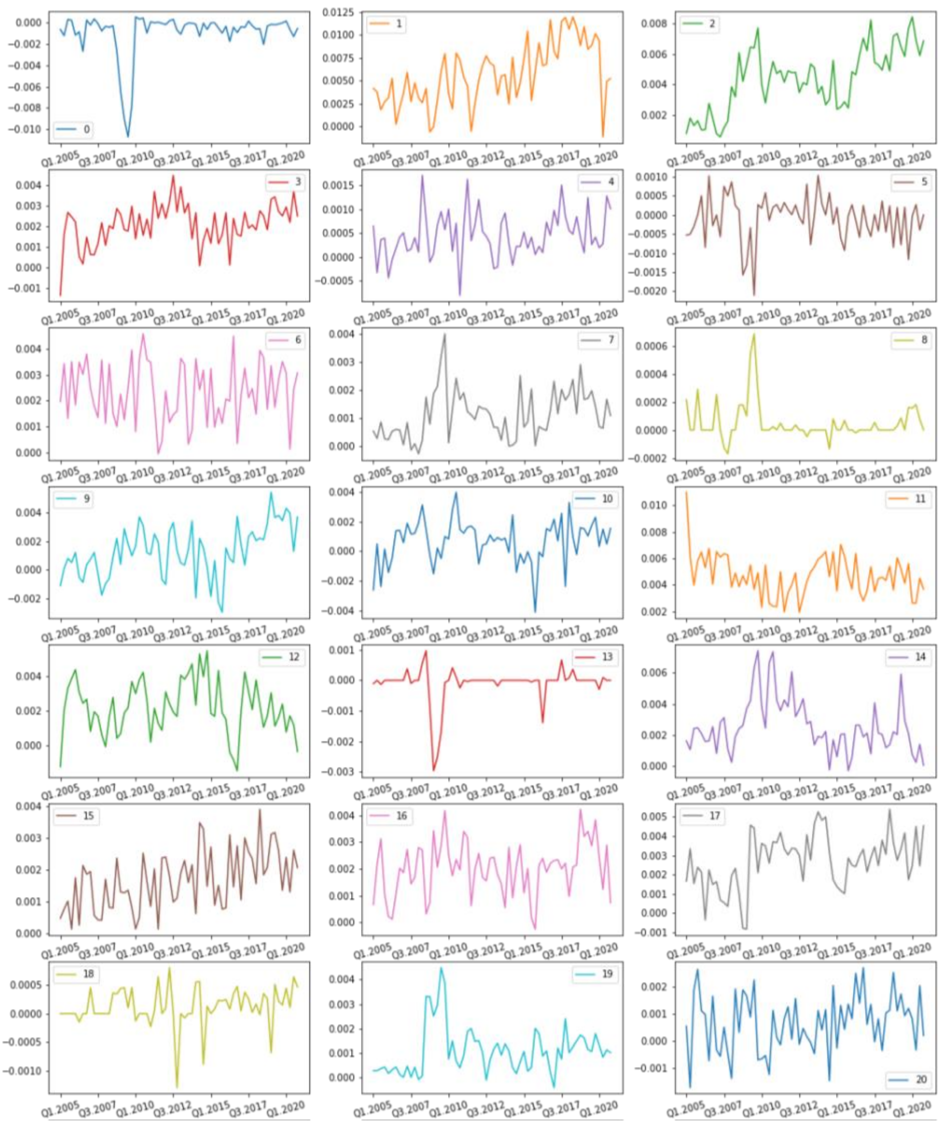
Appendix F2. Frequency News Index Calculation Code

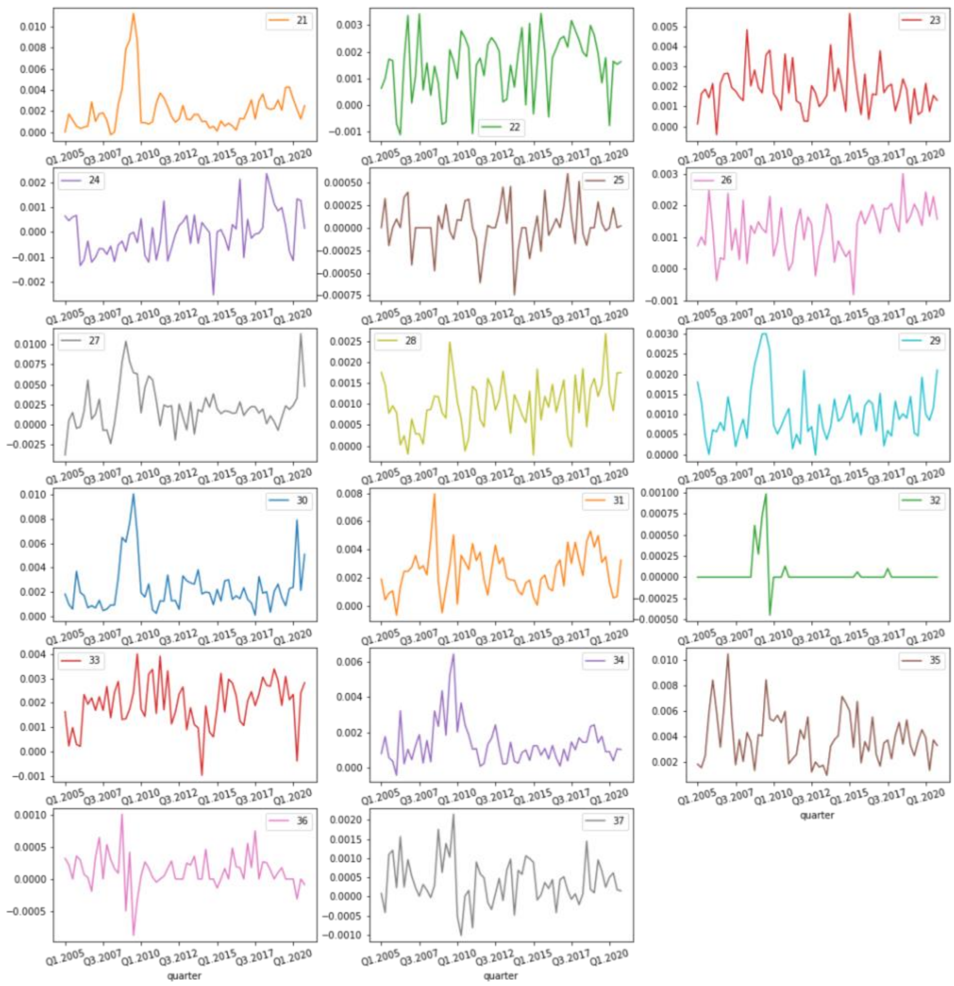
Link to the code for frequency news index (.ipynb):

https://drive.google.com/drive/folders/1Sa-c0PJNKmpLQNEeiyeAsoihCAeQVAn9?usp=sharing

Appendix F3. Frequency News Index Graphs

Appendix Figure #4





Appendix Figure 4. Frequency News Index per LDA topic (quarterly)

Source: *author's calculations*

Appendix F4. Sentiment News Index Dataset

Link to the dataset for coding procedure:

https://drive.google.com/drive/folders/1Tyjr3e058wkF3Aqs1lpP7dEHu-Loystp?usp=sharing

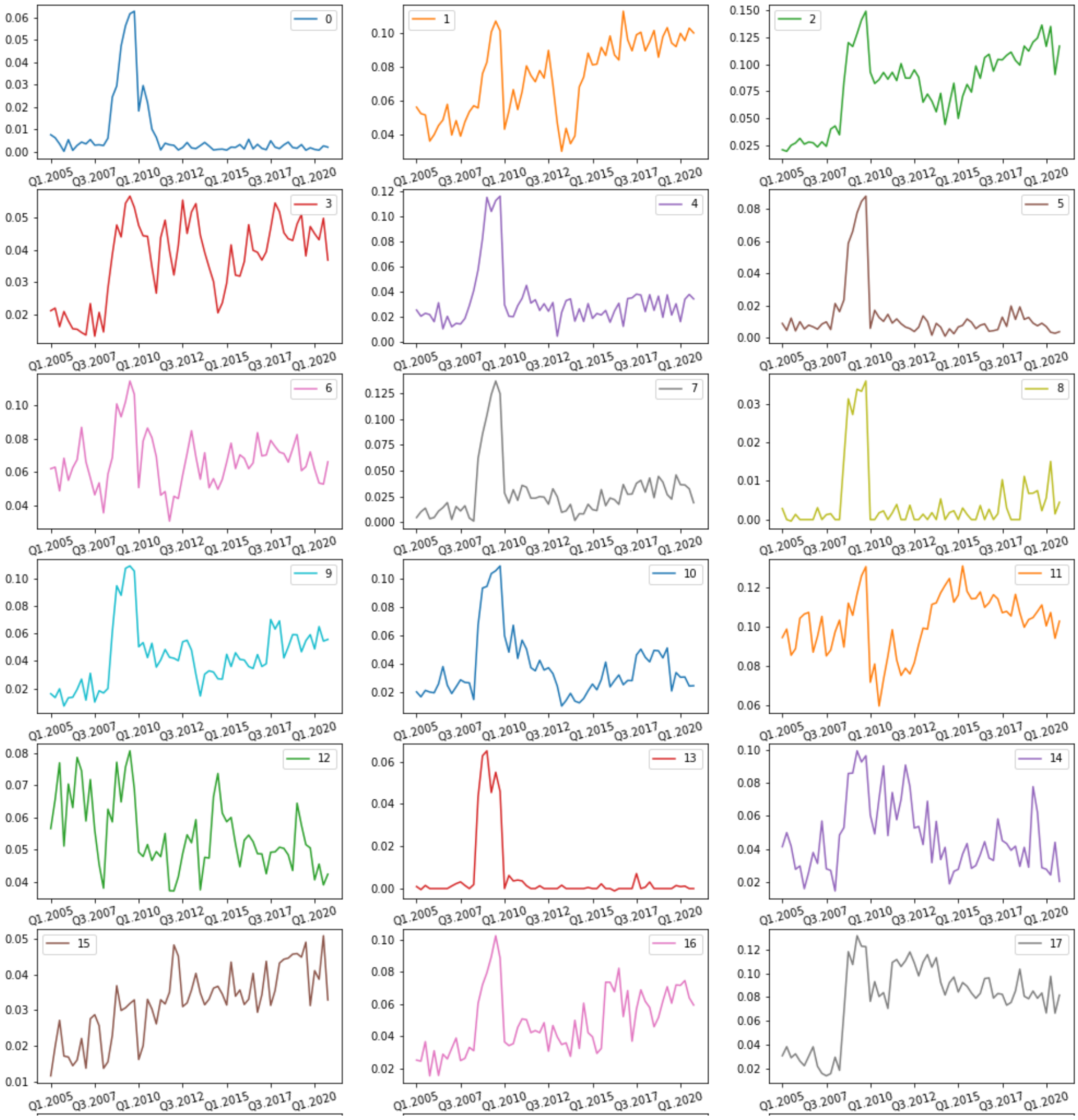
Appendix F5. Sentiment News Index Calculation Code

Link to the code for sentiment news index (.ipynb):

https://drive.google.com/drive/folders/1YmGBK0GILP6LWBwUIF8H6aermtY5PFY6?usp=sharing

Appendix F6. Sentiment News Index Graphs

Appendix Figure #5



Appendix Figure 6. Sentiment News Index per LDA topic (quarterly)

Source: *author's calculations*

Appendix G

Appendix G1. Time-series Forecasting Code

Link to the code for forecasting models (.ipynb):

https://drive.google.com/drive/folders/1j_aQmbD5z0O9VLS6P7VGt0Ep19qhXOyi?usp=sharing

Appendix H

Appendix H1. BCI models

Frequency Index [Basic]

	MAE	RMSE	model
1	0,1358991	0,1907882	ARF(3) without
2	0,1358991	0,1907882	ARN(3) without
3	0,1429468	0,1928899	ARF(6) without
4	0,1429468	0,1928899	ARN(6) without
5	0,1384428	0,1956255	ARF(6) with
6	0,1384428	0,1956255	ARN(6) with
7	0,1375944	0,1969688	ARF(3) with
8	0,1375944	0,1969688	ARN(3) with
9	0,1431262	0,1981164	ML(F)_linear_(6)_with()
10	0,1567965	0,2020769	ML(N)_ridge_(6)_with(out)
11	0,1457088	0,2029704	ML(F)_linear_(3)_with()
12	0,1472923	0,2031520	ML(F)_lasso_(3)_with(out)
13	0,1567763	0,2032474	ML(N)_lasso_(6)_with(out)
14	0,1574202	0,2039059	ML(N)_lasso_(3)_with(out)
15	0,1560607	0,2042421	ML(N)_elastic_(6)_with(out)
16	0,1528926	0,2042817	ML(F)_ridge_(6)_with(out)
17	0,1502540	0,2043391	ML(N)_linear_(6)_with()
18	0,1592661	0,2044745	ML(N)_ridge_(3)_with(out)
19	0,1559610	0,2046917	ML(F)_lasso_(6)_with(out)
20	0,1528427	0,2049887	ML(F)_ridge_(3)_with(out)
21	0,1541997	0,2057477	ML(F)_elastic_(6)_with(out)
22	0,1561486	0,2064443	ML(F)_linear_(6)_with(out)
23	0,1518341	0,2065729	ML(F)_elastic_(3)_with(out)
24	0,1559172	0,2066057	ML(F)_lasso_(3)_with()
25	0,1530413	0,2066644	ML(F)_linear_(3)_with(out)
26	0,1571688	0,2072725	ML(N)_linear_(6)_with(out)
27	0,1563565	0,2073388	ML(N)_elastic_(6)_with()
28	0,1563582	0,2073496	ML(F)_elastic_(6)_with()
29	0,1589698	0,2074371	ARXF(6) without
30	0,1595479	0,2075873	ML(N)_elastic_(3)_with(out)
31	0,1560374	0,2085505	ML(N)_linear_(3)_with()
32	0,1611228	0,2090288	ML(N)_linear_(3)_with(out)
33	0,1568831	0,2122769	ARXF(3) without
34	0,1604894	0,2125055	ARXN(3) without
35	0,1634752	0,2126530	ML(N)_elastic_(3)_with()
36	0,1634772	0,2126603	ML(F)_elastic_(3)_with()
37	0,1614385	0,2136133	ARXN(6) without
38	0,1658985	0,2176349	ML(F)_ridge_(6)_with()
39	0,1750230	0,2278948	ML(F)_ridge_(3)_with()
40	0,1757873	0,2306222	ARF(1) without
41	0,1757873	0,2306222	ARN(1) without
42	0,1755297	0,2341281	ARF(1) with
43	0,1755297	0,2341281	ARN(1) with
44	0,1780718	0,2357349	ML(N)_elastic_(1)_with()
45	0,1780789	0,2357459	ML(F)_elastic_(1)_with()
46	0,1814973	0,2377095	ML(F)_lasso_(1)_with(out)
47	0,1816392	0,2385112	ML(N)_lasso_(1)_with(out)
48	0,1774229	0,2388781	ML(N)_boost_(3)_with()

Frequency Index [Full]

	MAE	RMSE	model
1	0,1358991	0,1907882	ARF(3) without
2	0,1358991	0,1907882	ARN(3) without
3	0,1429468	0,1928899	ARF(6) without
4	0,1429468	0,1928899	ARN(6) without
5	0,1384428	0,1956255	ARF(6) with
6	0,1384428	0,1956255	ARN(6) with
7	0,1450200	0,1957659	ML(F)_ridge_(6)_with(out)
8	0,1375944	0,1969688	ARF(3) with
9	0,1375944	0,1969688	ARN(3) with
10	0,1431196	0,1984722	ML(F)_elastic_(6)_with(out)
11	0,1478696	0,1986555	ML(F)_ridge_(3)_with(out)
12	0,1434300	0,1988189	ML(F)_linear_(6)_with(out)
13	0,1536882	0,1989667	ML(N)_ridge_(6)_with(out)
14	0,1498693	0,1997978	ML(F)_lasso_(6)_with(out)
15	0,1448311	0,2000267	ML(N)_linear_(6)_with()
16	0,1505413	0,2004833	ML(F)_lasso_(3)_with(out)
17	0,1541127	0,2007417	ML(N)_lasso_(6)_with(out)
18	0,1542256	0,2007720	ML(N)_lasso_(3)_with(out)
19	0,1423036	0,2010093	ML(F)_elastic_(3)_with(out)
20	0,1557776	0,2014184	ML(N)_ridge_(3)_with(out)
21	0,1435687	0,2015389	ML(F)_linear_(3)_with(out)
22	0,1450150	0,2015437	ARXF(3) without
23	0,1475364	0,2039818	ML(N)_linear_(3)_with()
24	0,1434968	0,2059094	ML(F)_linear_(6)_with()
25	0,1541916	0,2063332	ML(N)_linear_(3)_with(out)
26	0,1590764	0,2063595	ML(N)_elastic_(6)_with(out)
27	0,1563554	0,2071479	ML(N)_elastic_(3)_with(out)
28	0,1588596	0,2072890	ML(N)_linear_(6)_with(out)
29	0,1563682	0,2073456	ML(N)_elastic_(6)_with()
30	0,1572706	0,2085365	ML(F)_elastic_(6)_with()
31	0,1604840	0,2090453	ARXN(3) without
32	0,1576406	0,2106104	ARXF(6) without
33	0,1509564	0,2117593	ML(F)_linear_(3)_with()
34	0,1634868	0,2126629	ML(N)_elastic_(3)_with()
35	0,1634805	0,2126794	ML(F)_elastic_(3)_with()
36	0,1724442	0,2211995	ARXN(6) without
37	0,1757873	0,2306222	ARF(1) without
38	0,1757873	0,2306222	ARN(1) without
39	0,1755297	0,2341281	ARF(1) with
40	0,1755297	0,2341281	ARN(1) with
41	0,1802893	0,2352196	ML(N)_lasso_(1)_with(out)
42	0,1780818	0,2357487	ML(N)_elastic_(1)_with()
43	0,1780777	0,2357695	ML(F)_elastic_(1)_with()
44	0,1824441	0,2376499	ML(N)_ridge_(1)_with(out)
45	0,1830228	0,2387129	ML(F)_ridge_(6)_with()
46	0,1830641	0,2387763	ML(F)_ridge_(3)_with()
47	0,1831995	0,2389542	ML(N)_ridge_(3)_with()
48	0,1832234	0,2389726	ML(N)_ridge_(6)_with()

Frequency Index [Optimal]

	MAE	RMSE	model
1	0,1358991	0,1907882	ARF(3) without
2	0,1358991	0,1907882	ARN(3) without
3	0,1429468	0,1928899	ARF(6) without
4	0,1429468	0,1928899	ARN(6) without
5	0,1425454	0,1954835	ML(F)_ridge_(6)_with(out)
6	0,1384428	0,1956255	ARF(6) with
7	0,1384428	0,1956255	ARN(6) with
8	0,1445397	0,1959052	ML(F)_lasso_(6)_with(out)
9	0,1375944	0,1969688	ARF(3) with
10	0,1375944	0,1969688	ARN(3) with
11	0,1468555	0,1990487	ML(F)_ridge_(3)_with(out)
12	0,1494506	0,1991747	ML(N)_ridge_(6)_with(out)
13	0,1436928	0,1995813	ML(F)_lasso_(3)_with(out)
14	0,1505710	0,2006405	ML(N)_ridge_(3)_with(out)
15	0,1528154	0,2022689	ML(N)_lasso_(3)_with(out)
16	0,1467788	0,2023486	ML(F)_elastic_(6)_with(out)
17	0,1422967	0,2030001	ML(F)_linear_(6)_with()
18	0,1547154	0,2039169	ML(N)_lasso_(6)_with(out)
19	0,1482526	0,2056108	ML(F)_elastic_(3)_with(out)
20	0,1531122	0,2061683	ML(F)_linear_(6)_with(out)
21	0,1563580	0,2073412	ML(N)_elastic_(6)_with()
22	0,1563572	0,2073510	ML(F)_elastic_(6)_with()
23	0,1460725	0,2075294	ML(F)_linear_(3)_with()
24	0,1535437	0,2075930	ARXF(3) without
25	0,1491237	0,2087355	ML(N)_linear_(6)_with()
26	0,1546560	0,2093498	ML(F)_linear_(3)_with(out)
27	0,1496223	0,2103470	ML(N)_linear_(3)_with()
28	0,1605349	0,2110334	ML(F)_lasso_(3)_with()
29	0,1555522	0,2116323	ML(N)_elastic_(3)_with(out)
30	0,1634753	0,2126627	ML(F)_elastic_(3)_with()
31	0,1645255	0,2138368	ML(N)_elastic_(3)_with()
32	0,1599743	0,2146211	ML(N)_linear_(3)_with(out)
33	0,1638794	0,2156078	ARXF(6) without
34	0,1629979	0,2167640	ARXN(3) without
35	0,1624915	0,2184674	ML(N)_elastic_(6)_with(out)
36	0,1718382	0,2267544	ML(N)_linear_(6)_with(out)
37	0,1757873	0,2306222	ARF(1) without
38	0,1757873	0,2306222	ARN(1) without
39	0,1768057	0,2311377	ARXN(6) without
40	0,1755297	0,2341281	ARF(1) with
41	0,1755297	0,2341281	ARN(1) with
42	0,1798745	0,2343418	ML(F)_ridge_(6)_with()
43	0,1788913	0,2346064	ML(N)_lasso_(1)_with(out)
44	0,1780733	0,2357409	ML(N)_elastic_(1)_with()
45	0,1780758	0,2357488	ML(F)_elastic_(1)_with()
46	0,1802380	0,2357524	ML(N)_ridge_(1)_with(out)
47	0,1822696	0,2385615	ML(N)_forest_(6)_with(out)
48	0,1828937	0,2385995	ML(F)_ridge_(3)_with()

Sentiment Index

	MAE	RMSE	model
1	0,1418067	0,1901424	ML(F)_lasso_(6)_with(out)
2	0,1358991	0,1907882	ARF(3) without
3	0,1358991	0,1907882	ARN(3) without
4	0,1430887	0,1909768	ML(F)_elastic_(6)_with(out)
5	0,1452797	0,1926394	ML(F)_linear_(6)_with(out)
6	0,1424198	0,1927435	ML(F)_lasso_(3)_with(out)
7	0,1429468	0,1928899	ARF(6) without
8	0,1429468	0,1928899	ARN(6) without
9	0,1422239	0,1933105	ML(F)_elastic_(3)_with(out)
10	0,1409212	0,1933295	ML(F)_linear_(3)_with(out)
11	0,1455230	0,1938199	ML(F)_ridge_(6)_with(out)
12	0,1455377	0,1940108	ML(F)_ridge_(3)_with(out)
13	0,1384428	0,1956255	ARF(6) with
14	0,1384428	0,1956255	ARN(6) with
15	0,1375944	0,1969688	ARF(3) with
16	0,1375944	0,1969688	ARN(3) with
17	0,1481228	0,1977458	ML(F)_linear_(6)_with()
18	0,1478164	0,1987595	ARXF(3) without
19	0,1475295	0,1989769	ML(F)_linear_(3)_with()
20	0,1512451	0,2017072	ARXF(6) without
21	0,1517348	0,2028496	ML(F)_lasso_(3)_with()
22	0,1565597	0,2075334	ML(F)_elastic_(6)_with()
23	0,1571570	0,2079804	ML(N)_elastic_(6)_with()
24	0,1587945	0,2097839	ML(F)_lasso_(6)_with()
25	0,1622092	0,2106411	ML(N)_lasso_(6)_with(out)
26	0,1629928	0,2113679	ML(N)_lasso_(3)_with(out)
27	0,1587754	0,2122840	ML(N)_linear_(3)_with()
28	0,1637458	0,2128718	ML(F)_elastic_(3)_with()
29	0,1641782	0,2131954	ML(N)_elastic_(3)_with()
30	0,1620229	0,2134104	ML(N)_linear_(6)_with()
31	0,1601784	0,2152387	ARXN(3) without
32	0,1591221	0,2159986	ML(N)_elastic_(3)_with(out)
33	0,1581793	0,2165425	ML(N)_linear_(3)_with(out)
34	0,1687441	0,2200596	ML(F)_ridge_(3)_with()
35	0,1744795	0,2236018	ML(N)_ridge_(3)_with(out)
36	0,1744795	0,2236018	ML(N)_ridge_(6)_with(out)
37	0,1749440	0,2280705	ML(N)_elastic_(6)_with(out)
38	0,1766348	0,2304002	ML(F)_ridge_(6)_with()
39	0,1757873	0,2306222	ARF(1) without
40	0,1757873	0,2306222	ARN(1) without
41	0,1785721	0,2326792	ML(N)_linear_(6)_with(out)
42	0,1755297	0,2341281	ARF(1) with
43	0,1755297	0,2341281	ARN(1) with
44	0,1782347	0,2362061	ML(F)_elastic_(1)_with()
45	0,1726120	0,2364019	ML(F)_forest_(3)_with(out)
46	0,1788072	0,2364231	ML(N)_elastic_(1)_with()
47	0,1738826	0,2369653	ML(F)_forest_(6)_with(out)
48	0,1838550	0,2398486	ML(N)_lasso_(1)_with(out)

49	0,1831995	0,2389542	ML(N)_ridge_(3)_with()
50	0,1832234	0,2389726	ML(N)_ridge_(6)_with()
51	0,1837665	0,2399180	ML(N)_ridge_(1)_with(out)
52	0,1942969	0,2403889	ML(N)_forest_(3)_with(out)
53	0,1832651	0,2403941	ML(F)_forest_(3)_with(out)
54	0,1826599	0,2405999	ML(F)_forest_(6)_with(out)
55	0,1842257	0,2407766	ML(F)_ridge_(1)_with(out)
56	0,1864601	0,2423230	ML(N)_linear_(1)_with()
57	0,1941390	0,2426898	ML(N)_forest_(6)_with(out)
58	0,1909537	0,2438128	ML(N)_elastic_(1)_with(out)
59	0,1917572	0,2444805	ML(N)_linear_(1)_with(out)
60	0,1932928	0,2446881	ARXN(1) without
61	0,1820910	0,2450235	ML(N)_boost_(6)_with()
62	0,1954035	0,2474319	ML(F)_elastic_(1)_with(out)
63	0,1926761	0,2479802	ML(F)_forest_(1)_with(out)
64	0,1947595	0,2480904	ML(N)_boost_(6)_with(out)
65	0,1913808	0,2481137	ML(F)_linear_(1)_with()
66	0,1973888	0,2483175	ARXF(1) without
67	0,1973397	0,2484040	ML(F)_linear_(1)_with(out)
68	0,1975055	0,2492913	ML(N)_forest_(1)_with(out)
69	0,1946772	0,2497915	ML(N)_boost_(3)_with(out)
70	0,1902985	0,2508408	ML(N)_lasso_(3)_with()
71	0,1902985	0,2508408	ML(N)_lasso_(6)_with()
72	0,1851563	0,2509893	ML(N)_forest_(6)_with()
73	0,1873953	0,2544911	ML(N)_forest_(3)_with()
74	0,1838712	0,2565794	ML(F)_boost_(3)_with(out)
75	0,1859296	0,2569857	ML(F)_boost_(6)_with(out)
76	0,1941816	0,2585818	ML(F)_lasso_(6)_with()
77	0,1966359	0,2590148	ML(N)_forest_(1)_with()
78	0,1981981	0,2603697	ML(N)_boost_(1)_with(out)
79	0,1968789	0,2611662	ML(F)_boost_(1)_with(out)
80	0,1948443	0,2631978	ML(N)_boost_(1)_with()
81	0,1856116	0,2638665	ML(F)_forest_(6)_with()
82	0,2000876	0,2641976	ARXN(1) with
83	0,1867109	0,2645676	ML(F)_boost_(6)_with()
84	0,1894740	0,2646379	ML(F)_forest_(3)_with()
85	0,1903580	0,2650222	ML(F)_boost_(3)_with()
86	0,1976421	0,2722514	ML(F)_forest_(1)_with()
87	0,2056071	0,2726305	ML(F)_boost_(1)_with()
88	0,2277514	0,2888737	ARXF(1) with
89	0,2261020	0,3029175	ML(F)_lasso_(1)_with()
90	0,2301251	0,3094289	ML(N)_lasso_(1)_with()
91	0,2304631	0,3103277	ML(N)_ridge_(1)_with()
92	0,2304631	0,3103277	ML(F)_ridge_(1)_with()
93	0,2587863	0,3203445	ARXN(3) with
94	0,2698372	0,3291733	ARXN(6) with
95	0,3043080	0,3598209	ARXF(6) with
96	0,3120741	0,3646011	ARXF(3) with

49	0,1830299	0,2401153	ML(F)_lasso_(1)_with(out)
50	0,1839086	0,2417826	ML(F)_forest_(6)_with(out)
51	0,1853961	0,2418381	ML(F)_forest_(3)_with(out)
52	0,1837698	0,2426249	ML(N)_forest_(3)_with(out)
53	0,1874641	0,2428892	ML(F)_forest_(1)_with(out)
54	0,1865723	0,2438873	ML(F)_ridge_(1)_with(out)
55	0,1848698	0,2439999	ML(N)_forest_(6)_with(out)
56	0,1858974	0,2462552	ML(F)_elastic_(1)_with(out)
57	0,1855871	0,2464739	ARXF(1) without
58	0,1862291	0,2467444	ML(F)_linear_(1)_with(out)
59	0,1891269	0,2491821	ML(N)_forest_(1)_with(out)
60	0,1902985	0,2508408	ML(N)_lasso_(3)_with()
61	0,1909066	0,2521112	ML(F)_lasso_(3)_with()
62	0,1909066	0,2521112	ML(N)_lasso_(6)_with()
63	0,1909066	0,2521112	ML(F)_lasso_(6)_with()
64	0,1767455	0,2522797	ML(N)_boost_(6)_with(out)
65	0,1849516	0,2523452	ML(F)_boost_(3)_with(out)
66	0,1920047	0,2524123	ML(N)_linear_(1)_with()
67	0,1992134	0,2531863	ML(N)_elastic_(1)_with(out)
68	0,1829170	0,2531947	ML(F)_linear_(1)_with()
69	0,1999750	0,2539573	ML(N)_linear_(1)_with(out)
70	0,1912872	0,2546290	ML(N)_boost_(1)_with(out)
71	0,1853792	0,2548755	ML(F)_boost_(6)_with(out)
72	0,2045444	0,2556025	ARXN(1) without
73	0,1833433	0,2560428	ML(F)_forest_(1)_with()
74	0,1923355	0,2560505	ML(N)_lasso_(1)_with()
75	0,1803546	0,2570605	ML(F)_boost_(6)_with()
76	0,1798543	0,2571159	ML(F)_boost_(3)_with()
77	0,1808314	0,2588706	ML(F)_forest_(6)_with()
78	0,1842422	0,2590541	ML(N)_forest_(6)_with()
79	0,1955161	0,2608935	ML(F)_boost_(1)_with(out)
80	0,1909741	0,2619704	ML(N)_forest_(3)_with()
81	0,1814728	0,2622632	ML(N)_boost_(3)_with(out)
82	0,1865161	0,2629416	ML(F)_forest_(3)_with()
83	0,1998677	0,2643842	ARXN(1) with
84	0,1976859	0,2700128	ML(F)_boost_(1)_with()
85	0,2111618	0,2700757	ML(N)_boost_(1)_with()
86	0,1981349	0,2715240	ML(N)_boost_(6)_with()
87	0,2009363	0,2736913	ML(N)_boost_(3)_with()
88	0,2101650	0,2744418	ML(N)_forest_(1)_with()
89	0,2320053	0,2924374	ARXF(1) with
90	0,2200861	0,2960982	ML(N)_ridge_(1)_with()
91	0,2484410	0,3131631	ARXN(3) with
92	0,2327066	0,3141482	ML(F)_lasso_(1)_with()
93	0,2361682	0,3191008	ML(F)_ridge_(1)_with()
94	0,2623945	0,3239346	ARXN(6) with
95	0,2728076	0,3296988	ARXF(3) with
96	0,2801333	0,3356103	ARXF(6) with

49	0,1830856	0,2387927	ML(N)_ridge_(6)_with()
50	0,1831995	0,2389542	ML(N)_ridge_(3)_with()
51	0,1807422	0,2393310	ML(F)_forest_(3)_with(out)
52	0,1826639	0,2397040	ML(F)_lasso_(1)_with(out)
53	0,1801048	0,2402201	ML(N)_forest_(3)_with(out)
54	0,1822423	0,2407577	ML(F)_forest_(1)_with(out)
55	0,1853061	0,2418729	ML(F)_lasso_(6)_with()
56	0,1849132	0,2420731	ML(F)_forest_(6)_with(out)
57	0,1760291	0,2432364	ML(F)_boost_(3)_with(out)
58	0,1837035	0,2432417	ML(F)_elastic_(1)_with(out)
59	0,1813105	0,2434601	ML(N)_elastic_(1)_with(out)
60	0,1855602	0,2438835	ARXF(1) without
61	0,1865723	0,2438873	ML(F)_ridge_(1)_with(out)
62	0,1852074	0,2441097	ML(F)_linear_(1)_with(out)
63	0,1822585	0,2443978	ML(N)_linear_(1)_with(out)
64	0,1843195	0,2460391	ARXN(1) without
65	0,1820941	0,2460700	ML(F)_linear_(1)_with()
66	0,1768146	0,2466522	ML(F)_boost_(6)_with(out)
67	0,1874823	0,2488130	ML(N)_forest_(1)_with(out)
68	0,1906388	0,2497650	ML(F)_boost_(1)_with(out)
69	0,1823357	0,2502534	ML(N)_forest_(6)_with()
70	0,1846022	0,2518182	ML(N)_linear_(1)_with()
71	0,1876960	0,2520576	ML(N)_boost_(3)_with(out)
72	0,1783251	0,2520963	ML(F)_boost_(6)_with()
73	0,1884256	0,2525360	ML(N)_boost_(6)_with(out)
74	0,1823945	0,2527514	ML(N)_boost_(6)_with()
75	0,1858039	0,2534714	ML(N)_forest_(3)_with()
76	0,1834146	0,2548537	ML(N)_boost_(3)_with()
77	0,1929706	0,2564090	ML(N)_lasso_(3)_with()
78	0,1863188	0,2599727	ML(F)_forest_(6)_with()
79	0,1957683	0,2609743	ML(N)_lasso_(6)_with()
80	0,1891878	0,2631209	ML(F)_forest_(3)_with()
81	0,2001782	0,2651675	ML(N)_forest_(1)_with()
82	0,1978343	0,2661385	ML(F)_boost_(1)_with()
83	0,2049635	0,2663082	ML(N)_boost_(1)_with(out)
84	0,1878282	0,2663154	ML(F)_forest_(1)_with()
85	0,1866055	0,2674542	ML(F)_boost_(3)_with()
86	0,2031773	0,2685306	ML(N)_boost_(1)_with()
87	0,2161342	0,2780109	ARXN(1) with
88	0,2167941	0,2803242	ARXF(1) with
89	0,2143747	0,2874132	ML(N)_lasso_(1)_with()
90	0,2200861	0,2960982	ML(N)_ridge_(1)_with()
91	0,2320282	0,3128605	ML(F)_lasso_(1)_with()
92	0,2519200	0,3164583	ARXN(3) with
93	0,2584356	0,3172413	ARXF(3) with
94	0,2620607	0,3183478	ARXF(6) with
95	0,2361682	0,3191008	ML(F)_ridge_(1)_with()
96	0,2675914	0,3284109	ARXN(6) with

49	0,1832500	0,2400766	ML(F)_lasso_(1)_with(out)
50	0,1804959	0,2409232	ML(F)_elastic_(1)_with(out)
51	0,1801763	0,2409684	ML(F)_linear_(1)_with(out)
52	0,1870935	0,2431186	ML(N)_ridge_(1)_with(out)
53	0,1838318	0,2440409	ARXF(1) without
54	0,1868228	0,2440418	ML(F)_ridge_(1)_with(out)
55	0,1754316	0,2444211	ML(F)_boost_(3)_with()
56	0,1860826	0,2448501	ML(N)_forest_(6)_with(out)
57	0,1773922	0,2449699	ML(N)_boost_(3)_with(out)
58	0,1746498	0,2454528	ML(F)_boost_(3)_with(out)
59	0,1839163	0,2469553	ML(F)_forest_(1)_with(out)
60	0,1787177	0,2470575	ML(F)_boost_(6)_with(out)
61	0,1783144	0,2473718	ML(N)_boost_(6)_with(out)
62	0,1930945	0,2484412	ARXN(6) without
63	0,1846898	0,2496446	ML(F)_linear_(1)_with()
64	0,1897874	0,2513410	ML(N)_forest_(3)_with(out)
65	0,1922272	0,2533156	ML(N)_forest_(1)_with(out)
66	0,1830292	0,2539520	ML(F)_boost_(1)_with(out)
67	0,1917305	0,2548092	ML(N)_linear_(1)_with()
68	0,1892078	0,2554404	ML(F)_boost_(1)_with()
69	0,1913211	0,2554901	ML(N)_boost_(1)_with(out)
70	0,1849717	0,2560306	ML(N)_boost_(3)_with()
71	0,1821436	0,2561902	ML(N)_forest_(6)_with()
72	0,1972565	0,2579283	ARXN(1) without
73	0,1953264	0,2579687	ML(N)_elastic_(1)_with(out)
74	0,1906478	0,2580180	ML(N)_boost_(1)_with()
75	0,1958483	0,2590312	ML(N)_linear_(1)_with(out)
76	0,1957683	0,2609743	ML(N)_lasso_(6)_with()
77	0,1812116	0,2611472	ML(F)_boost_(6)_with()
78	0,1792262	0,2613748	ML(F)_forest_(6)_with()
79	0,1974847	0,2635766	ML(N)_lasso_(3)_with()
80	0,1800915	0,2639191	ML(F)_forest_(3)_with()
81	0,1874005	0,2652692	ML(N)_boost_(6)_with()
82	0,2030904	0,2662301	ARXF(1) with
83	0,1924926	0,2694652	ML(N)_forest_(1)_with()
84	0,2057203	0,2713167	ARXN(1) with
85	0,1965552	0,2738369	ML(N)_forest_(3)_with()
86	0,1857606	0,2739016	ML(F)_forest_(1)_with()
87	0,2200861	0,2960982	ML(N)_ridge_(3)_with()
88	0,2200861	0,2960982	ML(N)_ridge_(6)_with()
89	0,2261020	0,3029175	ML(N)_lasso_(1)_with()
90	0,2304631	0,3103277	ML(N)_ridge_(1)_with()
91	0,2320282	0,3128605	ML(F)_lasso_(1)_with()
92	0,2569688	0,3171967	ARXF(6) with
93	0,2361682	0,3191008	ML(F)_ridge_(1)_with()
94	0,2610912	0,3210169	ARXF(3) with
95	0,2816428	0,3428850	ARXN(3) with
96	0,3016515	0,3618299	ARXN(6) with

Appendix H2.CCI models

Frequency Index [Basic]

	MAE	RMSE	model
1	0,0692149	0,1006574	ML(F)_elastic_(6)_with(out)
2	0,0698531	0,1012909	ML(F)_linear_(6)_with(out)
3	0,0698753	0,1013117	ML(F)_lasso_(6)_with(out)
4	0,0609347	0,1021340	ML(F)_elastic_(6)_with()
5	0,0610956	0,1022166	ML(N)_elastic_(6)_with()
6	0,0613125	0,1037349	ARN(6) with
7	0,0613125	0,1037349	ARF(6) with
8	0,0723561	0,1042370	ARXF(6) without
9	0,0676006	0,1045960	ARN(6) without
10	0,0676006	0,1045960	ARF(6) without
11	0,0727375	0,1059782	ML(N)_lasso_(6)_with(out)
12	0,0727612	0,1060770	ML(N)_linear_(6)_with(out)
13	0,0732262	0,1062153	ML(N)_elastic_(6)_with(out)
14	0,0710071	0,1063050	ML(F)_lasso_(6)_with()
15	0,0711801	0,1064247	ML(F)_linear_(6)_with()
16	0,0676693	0,1066081	ML(N)_lasso_(6)_with()
17	0,0677251	0,1067181	ML(N)_linear_(6)_with()
18	0,0776812	0,1127331	ARXN(6) without
19	0,0712860	0,1198517	ARN(3) with
20	0,0712860	0,1198517	ARF(3) with
21	0,0757638	0,1200731	ARN(3) without
22	0,0757638	0,1200731	ARF(3) without
23	0,0787040	0,1211266	ML(N)_linear_(3)_with()
24	0,0787040	0,1211266	ML(N)_elastic_(3)_with()
25	0,0787040	0,1211266	ML(N)_lasso_(3)_with()
26	0,0767349	0,1227407	ML(F)_lasso_(3)_with(out)
27	0,0767349	0,1227407	ML(F)_elastic_(3)_with(out)
28	0,0767349	0,1227407	ML(F)_linear_(3)_with(out)
29	0,0749053	0,1237051	ML(F)_linear_(3)_with()
30	0,0749053	0,1237051	ML(F)_elastic_(3)_with()
31	0,0749053	0,1237051	ML(F)_lasso_(3)_with()
32	0,0856974	0,1241003	ML(N)_linear_(3)_with(out)
33	0,0856974	0,1241003	ML(N)_elastic_(3)_with(out)
34	0,0856974	0,1241003	ML(N)_lasso_(3)_with(out)
35	0,0769466	0,1244815	ARXF(3) without
36	0,0886093	0,1269259	ARXN(3) without
37	0,0849466	0,1355441	ML(N)_ridge_(6)_with()
38	0,1005973	0,1430010	ML(N)_ridge_(6)_with(out)
39	0,0916675	0,1453832	ML(F)_ridge_(6)_with()
40	0,0975559	0,1475385	ML(F)_ridge_(6)_with(out)
41	0,0977134	0,1533111	ML(F)_ridge_(3)_with()
42	0,0978535	0,1534096	ML(N)_ridge_(3)_with()
43	0,1025441	0,1575136	ML(F)_ridge_(3)_with(out)
44	0,1079624	0,1590070	ML(N)_ridge_(3)_with(out)
45	0,1155155	0,1790597	ML(N)_boost_(3)_with(out)
46	0,1286210	0,1892804	ML(F)_boost_(3)_with()
47	0,1267246	0,1906703	ML(N)_boost_(3)_with()
48	0,1300381	0,1971419	ML(N)_boost_(6)_with(out)
49	0,1320455	0,2021981	ML(N)_boost_(6)_with()
50	0,1356957	0,2080955	ML(F)_boost_(6)_with()
51	0,1326881	0,2097027	ML(F)_boost_(3)_with(out)

Frequency Index [Full]

	MAE	RMSE	model
1	0,0622175	0,1029906	ML(N)_elastic_(6)_with()
2	0,0613125	0,1037349	ARF(6) with
3	0,0613125	0,1037349	ARN(6) with
4	0,0676006	0,1045960	ARF(6) without
5	0,0676006	0,1045960	ARN(6) without
6	0,0654805	0,1049537	ML(F)_elastic_(6)_with()
7	0,0733196	0,1106140	ML(F)_lasso_(6)_with()
8	0,0694958	0,1109626	ML(N)_linear_(6)_with()
9	0,0734755	0,1109639	ML(F)_linear_(6)_with()
10	0,0709725	0,1112341	ML(N)_lasso_(6)_with()
11	0,0769595	0,1144661	ML(N)_elastic_(6)_with(out)
12	0,0766986	0,1151599	ML(N)_lasso_(6)_with(out)
13	0,0807476	0,1163361	ML(F)_lasso_(6)_with(out)
14	0,0812174	0,1170295	ML(F)_elastic_(6)_with(out)
15	0,0813538	0,1172094	ML(F)_linear_(6)_with(out)
16	0,0786508	0,1184912	ARXN(6) without
17	0,0712860	0,1198517	ARF(3) with
18	0,0712860	0,1198517	ARN(3) with
19	0,0844082	0,1199402	ARXF(6) without
20	0,0757638	0,1200731	ARF(3) without
21	0,0757638	0,1200731	ARN(3) without
22	0,0821154	0,1223579	ML(N)_linear_(6)_with(out)
23	0,0805853	0,1242001	ML(N)_linear_(3)_with()
24	0,0805853	0,1242001	ML(N)_elastic_(3)_with()
25	0,0805853	0,1242001	ML(N)_lasso_(3)_with()
26	0,0798692	0,1272780	ML(F)_linear_(3)_with()
27	0,0798692	0,1272780	ML(F)_elastic_(3)_with()
28	0,0798692	0,1272780	ML(F)_lasso_(3)_with()
29	0,0865218	0,1296880	ML(N)_lasso_(3)_with(out)
30	0,0865218	0,1296880	ML(N)_linear_(3)_with(out)
31	0,0865218	0,1296880	ML(N)_elastic_(3)_with(out)
32	0,0861874	0,1311503	ARXN(3) without
33	0,0858548	0,1316899	ML(F)_lasso_(3)_with(out)
34	0,0858548	0,1316899	ML(F)_elastic_(3)_with(out)
35	0,0858548	0,1316899	ML(F)_linear_(3)_with(out)
36	0,0861990	0,1321033	ARXF(3) without
37	0,0849718	0,1355858	ML(N)_ridge_(6)_with()
38	0,0977740	0,1437483	ML(N)_ridge_(6)_with(out)
39	0,0915563	0,1451308	ML(F)_ridge_(6)_with()
40	0,0977134	0,1533111	ML(N)_ridge_(3)_with()
41	0,0977134	0,1533111	ML(F)_ridge_(3)_with()
42	0,1044811	0,1555137	ML(F)_ridge_(6)_with(out)
43	0,1021172	0,1560620	ML(N)_ridge_(3)_with(out)
44	0,1106441	0,1662711	ML(F)_ridge_(3)_with(out)
45	0,1215195	0,1782635	ML(N)_boost_(3)_with(out)
46	0,1308584	0,1895177	ML(N)_boost_(6)_with(out)
47	0,1373638	0,1895561	ML(F)_boost_(6)_with(out)
48	0,1303234	0,1912130	ML(F)_boost_(3)_with(out)
49	0,1344057	0,2003482	ML(F)_boost_(3)_with()
50	0,1303196	0,2016589	ML(F)_boost_(6)_with()
51	0,1371534	0,2107320	ML(F)_boost_(1)_with()

Frequency Index [Optimal]

	MAE	RMSE	model
1	0,0615879	0,1021729	ML(N)_elastic_(6)_with()
2	0,0611102	0,1022590	ML(F)_elastic_(6)_with()
3	0,0613125	0,1037349	ARF(6) with
4	0,0613125	0,1037349	ARN(6) with
5	0,0676006	0,1045960	ARF(6) without
6	0,0676006	0,1045960	ARN(6) without
7	0,0725763	0,1061663	ML(N)_lasso_(6)_with(out)
8	0,0725764	0,1063253	ML(N)_elastic_(6)_with(out)
9	0,0725698	0,1064635	ML(N)_linear_(6)_with(out)
10	0,0700910	0,1079702	ML(N)_lasso_(6)_with()
11	0,0704953	0,1084127	ML(N)_linear_(6)_with()
12	0,0718939	0,1086481	ML(F)_lasso_(6)_with()
13	0,0721672	0,1090911	ML(F)_linear_(6)_with()
14	0,0770181	0,1097839	ML(F)_elastic_(6)_with(out)
15	0,0772526	0,1101724	ML(F)_lasso_(6)_with(out)
16	0,0755670	0,1106276	ARXN(6) without
17	0,0777892	0,1109828	ML(F)_linear_(6)_with(out)
18	0,0806598	0,1130209	ARXF(6) without
19	0,0803376	0,1171319	ML(N)_elastic_(3)_with(out)
20	0,0800515	0,1171925	ML(N)_lasso_(3)_with(out)
21	0,0800515	0,1171925	ML(N)_linear_(3)_with(out)
22	0,0780091	0,1183755	ML(N)_linear_(3)_with()
23	0,0780091	0,1183755	ML(N)_lasso_(3)_with()
24	0,0803077	0,1185814	ARXN(3) without
25	0,0726413	0,1193046	ML(N)_elastic_(3)_with()
26	0,0725760	0,1197472	ML(F)_elastic_(3)_with()
27	0,0712860	0,1198517	ARF(3) with
28	0,0712860	0,1198517	ARN(3) with
29	0,0757638	0,1200731	ARF(3) without
30	0,0757638	0,1200731	ARN(3) without
31	0,0780679	0,1258706	ML(F)_linear_(3)_with()
32	0,0780679	0,1258706	ML(F)_lasso_(3)_with()
33	0,0805154	0,1265854	ML(F)_lasso_(3)_with(out)
34	0,0805154	0,1265854	ML(F)_linear_(3)_with(out)
35	0,0810897	0,1265879	ML(F)_elastic_(3)_with(out)
36	0,0817777	0,1275672	ARXF(3) without
37	0,0849649	0,1356357	ML(F)_ridge_(6)_with()
38	0,0957028	0,1422805	ML(N)_ridge_(6)_with(out)
39	0,0902660	0,1431240	ML(N)_ridge_(6)_with()
40	0,0986944	0,1460818	ML(F)_ridge_(6)_with(out)
41	0,1009061	0,1524261	ML(N)_ridge_(3)_with(out)
42	0,0977134	0,1533111	ML(N)_ridge_(3)_with()
43	0,0983398	0,1537752	ML(F)_ridge_(3)_with()
44	0,1085627	0,1618931	ML(F)_ridge_(3)_with(out)
45	0,1155037	0,1800699	ML(N)_boost_(3)_with(out)
46	0,1187729	0,1851163	ML(F)_boost_(3)_with(out)
47	0,1272385	0,1870829	ML(F)_boost_(6)_with(out)
48	0,1302511	0,1970497	ML(N)_boost_(6)_with(out)
49	0,1464104	0,2131071	ML(F)_boost_(1)_with(out)
50	0,1316566	0,2171206	ML(F)_forest_(3)_with()
51	0,1343880	0,2231864	ML(N)_forest_(3)_with(out)

Sentiment Index

	MAE	RMSE	model
1	0,0613125	0,1037349	ARF(6) with
2	0,0613125	0,1037349	ARN(6) with
3	0,0676006	0,1045960	ARF(6) without
4	0,0676006	0,1045960	ARN(6) without
5	0,0665907	0,1054222	ML(F)_elastic_(6)_with()
6	0,0677910	0,1062144	ML(F)_linear_(6)_with()
7	0,0719480	0,1065115	ML(N)_lasso_(6)_with(out)
8	0,0719410	0,1065931	ML(N)_linear_(6)_with(out)
9	0,0719410	0,1065931	ML(N)_elastic_(6)_with(out)
10	0,0689751	0,1069002	ML(F)_lasso_(6)_with()
11	0,0678772	0,1089263	ML(N)_lasso_(6)_with()
12	0,0678147	0,1090162	ML(N)_linear_(6)_with()
13	0,0678147	0,1090162	ML(N)_elastic_(6)_with()
14	0,0768131	0,1100809	ML(F)_lasso_(6)_with(out)
15	0,0774658	0,1104620	ML(F)_elastic_(6)_with(out)
16	0,0756434	0,1108848	ARXN(6) without
17	0,0797537	0,1139560	ARXF(6) without
18	0,0772103	0,1164551	ML(F)_linear_(6)_with(out)
19	0,0712860	0,1198517	ARF(3) with
20	0,0712860	0,1198517	ARN(3) with
21	0,0757638	0,1200731	ARF(3) without
22	0,0757638	0,1200731	ARN(3) without
23	0,0798734	0,1242592	ML(F)_linear_(3)_with()
24	0,0798734	0,1242592	ML(F)_elastic_(3)_with()
25	0,0798734	0,1242592	ML(F)_lasso_(3)_with()
26	0,0766902	0,1243212	ML(N)_elastic_(3)_with()
27	0,0793064	0,1267324	ML(N)_linear_(3)_with()
28	0,0855710	0,1281417	ML(N)_linear_(3)_with(out)
29	0,0846117	0,1282401	ARXN(3) without
30	0,0860184	0,1283909	ML(N)_elastic_(3)_with(out)
31	0,0898341	0,1329240	ML(N)_lasso_(3)_with(out)
32	0,0915114	0,1331817	ML(F)_lasso_(3)_with(out)
33	0,0915114	0,1331817	ML(F)_elastic_(3)_with(out)
34	0,0915114	0,1331817	ML(F)_linear_(3)_with(out)
35	0,0928907	0,1351500	ARXF(3) without
36	0,0880930	0,1367952	ML(N)_lasso_(3)_with()
37	0,0856345	0,1370491	ML(N)_ridge_(6)_with()
38	0,0858573	0,1371011	ML(F)_ridge_(6)_with()
39	0,0977134	0,1533111	ML(F)_ridge_(3)_with()
40	0,0977774	0,1533592	ML(N)_ridge_(3)_with()
41	0,1045735	0,1534589	ML(F)_ridge_(6)_with(out)
42	0,1031036	0,1544648	ML(N)_ridge_(6)_with(out)
43	0,1124717	0,1677666	ML(F)_ridge_(3)_with(out)
44	0,1099867	0,1679454	ML(N)_ridge_(3)_with(out)
45	0,1256099	0,1870522	ML(F)_boost_(3)_with()
46	0,1256384	0,1871892	ML(F)_boost_(3)_with(out)
47	0,1272363	0,1884027	ML(F)_boost_(6)_with()
48	0,1348819	0,1908419	ML(F)_boost_(6)_with(out)
49	0,1259944	0,1917389	ML(N)_boost_(3)_with(out)
50	0,1300256	0,1951441	ML(N)_boost_(3)_with()
51	0,1339064	0,2041074	ML(N)_boost_(6)_with()

52	0,1279993	0,2162508	ML(F)_forest_(3)_with()
53	0,1444582	0,2230377	ML(N)_linear_(1)_with(out)
54	0,1444582	0,2230377	ML(N)_elastic_(1)_with(out)
55	0,1436789	0,2232174	ML(N)_ridge_(1)_with(out)
56	0,1432794	0,2235322	ML(N)_lasso_(1)_with(out)
57	0,1451580	0,2236520	ARXN(1) without
58	0,1394261	0,2239526	ML(N)_forest_(3)_with()
59	0,1413467	0,2253015	ML(F)_elastic_(1)_with(out)
60	0,1427195	0,2262700	ML(F)_linear_(1)_with(out)
61	0,1369402	0,2264904	ML(F)_ridge_(1)_with(out)
62	0,1370811	0,2266787	ML(F)_lasso_(1)_with(out)
63	0,1431912	0,2269438	ARXF(1) without
64	0,1440835	0,2272358	ML(N)_linear_(1)_with()
65	0,1440835	0,2272358	ML(N)_elastic_(1)_with()
66	0,1357878	0,2283875	ML(N)_forest_(3)_with(out)
67	0,1297771	0,2292097	ML(F)_forest_(6)_with()
68	0,1360860	0,2300889	ML(F)_ridge_(1)_with()
69	0,1358767	0,2303859	ARN(1) without
70	0,1358767	0,2303859	ARF(1) without
71	0,1359770	0,2303944	ML(F)_elastic_(1)_with()
72	0,1364909	0,2305594	ML(F)_lasso_(1)_with()
73	0,1348087	0,2314636	ML(N)_forest_(6)_with(out)
74	0,1375650	0,2320640	ML(N)_lasso_(1)_with()
75	0,1430611	0,2324556	ML(F)_linear_(1)_with()
76	0,1378247	0,2324762	ML(N)_ridge_(1)_with()
77	0,1381476	0,2329841	ARN(1) with
78	0,1381476	0,2329841	ARF(1) with
79	0,1328476	0,2337145	ML(F)_forest_(3)_with(out)
80	0,1500010	0,2375372	ML(F)_forest_(1)_with()
81	0,1344292	0,2380665	ML(F)_forest_(6)_with(out)
82	0,1550819	0,2385443	ML(F)_boost_(6)_with(out)
83	0,1446818	0,2451454	ML(N)_forest_(6)_with()
84	0,1563565	0,2494673	ML(F)_forest_(1)_with(out)
85	0,1582642	0,2508043	ML(N)_forest_(1)_with()
86	0,1586906	0,2528870	ML(N)_forest_(1)_with(out)
87	0,1653528	0,2613822	ML(N)_boost_(1)_with(out)
88	0,1780914	0,2769000	ML(N)_boost_(1)_with()
89	0,2306344	0,2822930	ARXF(1) with
90	0,1956478	0,2916382	ML(F)_boost_(1)_with()
91	0,2507721	0,3021195	ARXN(1) with
92	0,2197133	0,3457578	ML(F)_boost_(1)_with(out)
93	0,4349076	0,4882395	ARXN(3) with
94	0,4938061	0,5482906	ARXF(3) with
95	0,4812675	0,5486863	ARXN(6) with
96	0,5363939	0,6074102	ARXF(6) with

52	0,1324113	0,2121669	ML(F)_forest_(3)_with()
53	0,1320492	0,2246703	ML(N)_forest_(3)_with(out)
54	0,1361597	0,2255904	ML(F)_forest_(6)_with()
55	0,1413698	0,2278964	ML(N)_lasso_(1)_with(out)
56	0,1443073	0,2290812	ML(N)_ridge_(1)_with(out)
57	0,1455619	0,2295815	ML(N)_elastic_(1)_with(out)
58	0,1457249	0,2297062	ML(N)_linear_(1)_with(out)
59	0,1467524	0,2297452	ML(F)_boost_(1)_with(out)
60	0,1460075	0,2301246	ARXN(1) without
61	0,1423476	0,2302558	ML(F)_lasso_(1)_with(out)
62	0,1347826	0,2303574	ML(N)_boost_(6)_with()
63	0,1358767	0,2303859	ARF(1) without
64	0,1358767	0,2303859	ARN(1) without
65	0,1362491	0,2308765	ML(F)_forest_(3)_with(out)
66	0,1368640	0,2310523	ML(F)_lasso_(1)_with()
67	0,1332968	0,2316837	ML(N)_forest_(6)_with(out)
68	0,1373073	0,2316907	ML(N)_lasso_(1)_with()
69	0,1448585	0,2320424	ML(F)_ridge_(1)_with(out)
70	0,1451051	0,2323454	ML(F)_elastic_(1)_with(out)
71	0,1378622	0,2323587	ML(F)_forest_(6)_with(out)
72	0,1377218	0,2323847	ML(N)_elastic_(1)_with()
73	0,1378247	0,2324762	ML(N)_ridge_(1)_with()
74	0,1378247	0,2324762	ML(F)_ridge_(1)_with()
75	0,1452112	0,2324950	ML(F)_linear_(1)_with(out)
76	0,1379267	0,2326351	ML(F)_elastic_(1)_with()
77	0,1381476	0,2329841	ARF(1) with
78	0,1381476	0,2329841	ARN(1) with
79	0,1453986	0,2330563	ARXF(1) without
80	0,1460170	0,2345305	ML(N)_linear_(1)_with()
81	0,1397234	0,2360607	ML(N)_forest_(3)_with()
82	0,1394925	0,2369655	ML(N)_boost_(3)_with()
83	0,1484096	0,2384275	ML(N)_boost_(1)_with(out)
84	0,1558868	0,2388601	ML(F)_forest_(1)_with()
85	0,1496912	0,2417975	ML(F)_linear_(1)_with()
86	0,1671132	0,2432847	ML(N)_forest_(1)_with(out)
87	0,1415029	0,2435399	ML(N)_forest_(6)_with()
88	0,1637170	0,2476795	ML(F)_forest_(1)_with(out)
89	0,1659679	0,2597966	ML(N)_boost_(1)_with()
90	0,1688679	0,2609959	ML(N)_forest_(1)_with()
91	0,2154088	0,2715914	ARXF(1) with
92	0,2658796	0,3142388	ARXN(1) with
93	0,4656579	0,5174317	ARXN(3) with
94	0,4915227	0,5455303	ARXF(3) with
95	0,5122442	0,5777158	ARXN(6) with
96	0,5407767	0,6088765	ARXF(6) with

52	0,1308377	0,2237436	ML(F)_forest_(3)_with(out)
53	0,1429217	0,2238920	ML(N)_elastic_(1)_with(out)
54	0,1456318	0,2241672	ML(N)_boost_(1)_with(out)
55	0,1422500	0,2245394	ML(N)_lasso_(1)_with(out)
56	0,1435114	0,2246593	ML(N)_ridge_(1)_with(out)
57	0,1445228	0,2248035	ML(N)_linear_(1)_with(out)
58	0,1454277	0,2252447	ARXN(1) without
59	0,1312976	0,2261243	ML(N)_boost_(6)_with()
60	0,1369520	0,2262145	ML(N)_boost_(3)_with()
61	0,1336195	0,2275898	ML(F)_forest_(6)_with(out)
62	0,1359594	0,2303767	ML(N)_elastic_(1)_with()
63	0,1316988	0,2303779	ML(F)_boost_(6)_with()
64	0,1358767	0,2303859	ARF(1) without
65	0,1358767	0,2303859	ARN(1) without
66	0,1452619	0,2305812	ML(F)_elastic_(1)_with(out)
67	0,1454068	0,2307482	ML(F)_linear_(1)_with(out)
68	0,1451897	0,2308871	ML(F)_ridge_(1)_with(out)
69	0,1451779	0,2309194	ML(F)_lasso_(1)_with(out)
70	0,1454332	0,2310680	ARXF(1) without
71	0,1435781	0,2319967	ML(N)_linear_(1)_with()
72	0,1375650	0,2320640	ML(N)_lasso_(1)_with()
73	0,1378228	0,2324732	ML(F)_lasso_(1)_with()
74	0,1378247	0,2324762	ML(F)_ridge_(1)_with()
75	0,1378267	0,2324791	ML(N)_ridge_(1)_with()
76	0,1378954	0,2325627	ML(F)_elastic_(1)_with()
77	0,1381476	0,2329841	ARF(1) with
78	0,1381476	0,2329841	ARN(1) with
79	0,1369615	0,2335257	ML(F)_boost_(3)_with()
80	0,1366604	0,2338177	ML(N)_forest_(6)_with()
81	0,1498742	0,2353746	ML(N)_forest_(1)_with(out)
82	0,1408962	0,2356643	ML(N)_forest_(6)_with(out)
83	0,1502614	0,2365667	ML(F)_forest_(1)_with()
84	0,1468729	0,2365760	ML(F)_linear_(1)_with()
85	0,1356824	0,2368036	ML(F)_forest_(6)_with()
86	0,1430064	0,2373661	ML(N)_forest_(3)_with()
87	0,1541747	0,2381602	ML(F)_forest_(1)_with(out)
88	0,1574214	0,2539296	ML(N)_forest_(1)_with()
89	0,1679716	0,2643944	ML(F)_boost_(1)_with()
90	0,2093401	0,2666824	ARXF(1) with
91	0,1620153	0,2681077	ML(N)_boost_(1)_with()
92	0,2625896	0,3106363	ARXN(1) with
93	0,4552479	0,5059021	ARXN(3) with
94	0,4979310	0,5508216	ARXF(3) with
95	0,4987729	0,5643657	ARXN(6) with
96	0,5416529	0,6085186	ARXF(6) with

52	0,1456791	0,2052942	ML(F)_boost_(1)_with(out)
53	0,1491253	0,2170123	ML(N)_boost_(6)_with(out)
54	0,1306444	0,2210387	ML(N)_forest_(3)_with(out)
55	0,1483927	0,2238496	ML(F)_boost_(1)_with()
56	0,1388000	0,2244516	ML(F)_forest_(3)_with()
57	0,1341358	0,2252369	ML(F)_forest_(3)_with(out)
58	0,1395400	0,2293677	ML(N)_forest_(6)_with(out)
59	0,1358767	0,2303859	ARF(1) without
60	0,1358767	0,2303859	ARN(1) without
61	0,1370363	0,2306977	ML(F)_forest_(6)_with()
62	0,1373073	0,2316907	ML(N)_lasso_(1)_with()
63	0,1373073	0,2316907	ML(F)_lasso_(1)_with()
64	0,1378247	0,2324762	ML(N)_ridge_(1)_with()
65	0,1378247	0,2324762	ML(F)_ridge_(1)_with()
66	0,1381476	0,2329841	ARF(1) with
67	0,1381476	0,2329841	ARN(1) with
68	0,1381906	0,2341854	ML(N)_elastic_(1)_with()
69	0,1381591	0,2355511	ML(F)_elastic_(1)_with()
70	0,1374109	0,2355902	ML(F)_forest_(6)_with(out)
71	0,1456986	0,2361079	ML(N)_forest_(3)_with()
72	0,1389438	0,2367465	ML(F)_lasso_(1)_with(out)
73	0,1420103	0,2387316	ML(F)_ridge_(1)_with(out)
74	0,1435678	0,2398093	ML(F)_elastic_(1)_with(out)
75	0,1437813	0,2399759	ML(F)_linear_(1)_with(out)
76	0,1487516	0,2401411	ML(N)_lasso_(1)_with(out)
77	0,1444536	0,2401441	ML(N)_forest_(6)_with()
78	0,1449531	0,2404474	ARXF(1) without
79	0,1540647	0,2427948	ML(N)_forest_(1)_with(out)
80	0,1529087	0,2437781	ML(N)_ridge_(1)_with(out)
81	0,1489388	0,2444779	ML(N)_linear_(1)_with()
82	0,1551592	0,2453553	ARXN(1) without
83	0,1546687	0,2453923	ML(N)_elastic_(1)_with(out)
84	0,1548862	0,2455910	ML(N)_linear_(1)_with(out)
85	0,1447866	0,2485015	ML(F)_linear_(1)_with()
86	0,1603598	0,2500205	ML(F)_forest_(1)_with()
87	0,1646500	0,2507100	ML(F)_forest_(1)_with(out)
88	0,1591563	0,2543344	ML(N)_forest_(1)_with()
89	0,1897728	0,2816105	ML(N)_boost_(1)_with(out)
90	0,2476949	0,2987175	ARXF(1) with
91	0,2062792	0,3131555	ML(N)_boost_(1)_with()
92	0,3130812	0,3545393	ARXN(1) with
93	0,4667775	0,5188056	ARXF(3) with
94	0,4900521	0,5536940	ARXF(6) with
95	0,5044536	0,5549341	ARXN(3) with
96	0,5537823	0,6175301	ARXN(6) with

Appendix H3.CPI models

Frequency Index [Basic]

	MAE	RMSE	model
1	0,1891502	0,2348348	ARF(1) without
2	0,1891502	0,2348348	ARN(1) without
3	0,1898104	0,2349820	ML(N)_elastic_(1)_with()
4	0,1898137	0,2349906	ML(F)_elastic_(1)_with()
5	0,1898065	0,2362562	ARF(1) with
6	0,1898065	0,2362562	ARN(1) with
7	0,1905710	0,2365035	ML(F)_lasso_(1)_with(out)
8	0,1905710	0,2365035	ML(F)_lasso_(3)_with(out)
9	0,1905710	0,2365035	ML(F)_lasso_(6)_with(out)
10	0,1949122	0,2375483	ML(F)_ridge_(1)_with(out)
11	0,1946828	0,2377414	ML(F)_ridge_(3)_with(out)
12	0,1946367	0,2378780	ML(F)_ridge_(6)_with(out)
13	0,1934927	0,2386463	ML(F)_lasso_(1)_with()
14	0,1934927	0,2386463	ML(F)_lasso_(3)_with()
15	0,1934927	0,2386463	ML(F)_lasso_(6)_with()
16	0,1942709	0,2396504	ML(F)_ridge_(3)_with()
17	0,1942772	0,2396542	ML(F)_ridge_(6)_with()
18	0,1942164	0,2396562	ML(F)_ridge_(1)_with()
19	0,1947682	0,2397629	ML(N)_ridge_(6)_with()
20	0,1944254	0,2402476	ML(N)_lasso_(1)_with(out)
21	0,1944254	0,2402476	ML(N)_lasso_(3)_with(out)
22	0,1944254	0,2402476	ML(N)_lasso_(6)_with(out)
23	0,1952577	0,2409032	ML(N)_ridge_(1)_with()
24	0,1952577	0,2409032	ML(N)_ridge_(3)_with()
25	0,1982516	0,2437943	ML(N)_lasso_(1)_with()
26	0,1982516	0,2437943	ML(N)_lasso_(3)_with()
27	0,1982516	0,2437943	ML(N)_lasso_(6)_with()
28	0,1923162	0,2453611	ML(N)_elastic_(3)_with()
29	0,1923239	0,2453614	ML(F)_elastic_(3)_with()
30	0,1941563	0,2469871	ML(F)_elastic_(6)_with()
31	0,1941435	0,2469883	ML(N)_elastic_(6)_with()
32	0,1934791	0,2490736	ARF(3) without
33	0,1934791	0,2490736	ARN(3) without
34	0,1936887	0,2493171	ARF(3) with
35	0,1936887	0,2493171	ARN(3) with
36	0,2044008	0,2504418	ML(N)_ridge_(1)_with(out)
37	0,2044008	0,2504418	ML(N)_ridge_(3)_with(out)
38	0,2028996	0,2539566	ARF(6) with
39	0,2028996	0,2539566	ARN(6) with
40	0,2031357	0,2546284	ARF(6) without
41	0,2031357	0,2546284	ARN(6) without
42	0,2275582	0,2650836	ML(F)_linear_(1)_with()
43	0,2171327	0,2704343	ML(N)_ridge_(6)_with(out)
44	0,2328373	0,2723139	ML(F)_elastic_(1)_with(out)
45	0,2325846	0,2728794	ARXF(1) without
46	0,2343741	0,2741318	ML(F)_linear_(1)_with(out)
47	0,2221885	0,2760175	ARXF(1) with
48	0,2335214	0,2775314	ML(F)_linear_(6)_with()
49	0,2375579	0,2797878	ML(F)_linear_(3)_with()
50	0,2282585	0,2802854	ARXN(3) with
51	0,2349184	0,2804511	ML(F)_elastic_(6)_with(out)

Frequency Index [Full]

	MAE	RMSE	model
1	0,1891502	0,2348348	ARF(1) without
2	0,1891502	0,2348348	ARN(1) without
3	0,1898179	0,2350060	ML(F)_elastic_(1)_with()
4	0,1898152	0,2350277	ML(N)_elastic_(1)_with()
5	0,1846556	0,2351549	ML(N)_lasso_(6)_with(out)
6	0,1898065	0,2362562	ARF(1) with
7	0,1898065	0,2362562	ARN(1) with
8	0,1916471	0,2363225	ML(N)_lasso_(6)_with()
9	0,1904832	0,2364326	ML(N)_lasso_(1)_with(out)
10	0,1904832	0,2364326	ML(N)_lasso_(3)_with(out)
11	0,1940457	0,2373444	ML(F)_ridge_(1)_with(out)
12	0,1940457	0,2373444	ML(F)_ridge_(3)_with(out)
13	0,1940457	0,2373444	ML(F)_ridge_(6)_with(out)
14	0,1919889	0,2378900	ML(F)_lasso_(1)_with(out)
15	0,1919889	0,2378900	ML(F)_lasso_(3)_with(out)
16	0,1921218	0,2380033	ML(F)_lasso_(6)_with(out)
17	0,1933947	0,2384592	ML(N)_lasso_(1)_with()
18	0,1933947	0,2384592	ML(N)_lasso_(3)_with()
19	0,1938162	0,2389834	ML(N)_ridge_(1)_with()
20	0,1902665	0,2391752	ML(N)_ridge_(6)_with(out)
21	0,1944656	0,2394526	ML(N)_ridge_(3)_with()
22	0,1943392	0,2394755	ML(N)_ridge_(6)_with()
23	0,1952577	0,2409032	ML(F)_ridge_(1)_with()
24	0,1952577	0,2409032	ML(F)_ridge_(3)_with()
25	0,1952577	0,2409032	ML(F)_ridge_(6)_with()
26	0,1966098	0,2418856	ML(F)_lasso_(1)_with()
27	0,1966098	0,2418856	ML(F)_lasso_(3)_with()
28	0,1968435	0,2421448	ML(F)_lasso_(6)_with()
29	0,1940826	0,2453499	ML(N)_ridge_(3)_with(out)
30	0,1923242	0,2453900	ML(N)_elastic_(3)_with()
31	0,1923548	0,2453915	ML(F)_elastic_(3)_with()
32	0,1956556	0,2483367	ML(F)_elastic_(6)_with()
33	0,1956179	0,2483381	ML(N)_elastic_(6)_with()
34	0,1934791	0,2490736	ARF(3) without
35	0,1934791	0,2490736	ARN(3) without
36	0,1936887	0,2493171	ARF(3) with
37	0,1936887	0,2493171	ARN(3) with
38	0,1999064	0,2529657	ML(N)_ridge_(1)_with(out)
39	0,2028996	0,2539566	ARF(6) with
40	0,2028996	0,2539566	ARN(6) with
41	0,2031357	0,2546284	ARF(6) without
42	0,2031357	0,2546284	ARN(6) without
43	0,2124742	0,2604005	ML(N)_linear_(3)_with()
44	0,2247103	0,2703835	ARXF(1) with
45	0,2227971	0,2739006	ML(N)_linear_(6)_with()
46	0,2282671	0,2789778	ML(N)_linear_(1)_with()
47	0,2296948	0,2855695	ML(N)_forest_(1)_with(out)
48	0,2348825	0,2882371	ML(N)_elastic_(6)_with(out)
49	0,2301054	0,2895937	ML(N)_forest_(1)_with()
50	0,2325620	0,2933693	ML(N)_forest_(3)_with()
51	0,2383696	0,2936979	ML(N)_linear_(6)_with(out)

Frequency Index [Optimal]

	MAE	RMSE	model
1	0,1891502	0,2348348	ARF(1) without
2	0,1891502	0,2348348	ARN(1) without
3	0,1898151	0,2349810	ML(N)_elastic_(1)_with()
4	0,1898065	0,2362562	ARF(1) with
5	0,1898065	0,2362562	ARN(1) with
6	0,1909220	0,2363085	ML(F)_elastic_(1)_with()
7	0,1907437	0,2366661	ML(N)_lasso_(1)_with(out)
8	0,1907437	0,2366661	ML(N)_lasso_(3)_with(out)
9	0,1908301	0,2367509	ML(N)_lasso_(6)_with(out)
10	0,1940457	0,2373444	ML(F)_ridge_(1)_with(out)
11	0,1940457	0,2373444	ML(F)_ridge_(3)_with(out)
12	0,1940457	0,2373444	ML(F)_ridge_(6)_with(out)
13	0,1989817	0,2374677	ML(N)_ridge_(6)_with(out)
14	0,1917230	0,2376702	ML(F)_lasso_(1)_with(out)
15	0,1917230	0,2376702	ML(F)_lasso_(3)_with(out)
16	0,1917230	0,2376702	ML(F)_lasso_(6)_with(out)
17	0,1932564	0,2381685	ML(N)_ridge_(1)_with()
18	0,1937569	0,2390347	ML(N)_lasso_(1)_with()
19	0,1937569	0,2390347	ML(N)_lasso_(3)_with()
20	0,1939266	0,2392360	ML(N)_lasso_(6)_with()
21	0,1947462	0,2397475	ML(N)_ridge_(6)_with()
22	0,1952577	0,2409032	ML(F)_ridge_(1)_with()
23	0,1952577	0,2409032	ML(F)_ridge_(3)_with()
24	0,1952577	0,2409032	ML(F)_ridge_(6)_with()
25	0,1961424	0,2413810	ML(F)_lasso_(1)_with()
26	0,1961424	0,2413810	ML(F)_lasso_(3)_with()
27	0,1961424	0,2413810	ML(F)_lasso_(6)_with()
28	0,1925649	0,2439611	ML(N)_ridge_(3)_with()
29	0,1923320	0,2453589	ML(N)_elastic_(3)_with()
30	0,1935077	0,2467769	ML(F)_elastic_(3)_with()
31	0,1941559	0,2469780	ML(N)_elastic_(6)_with()
32	0,1956523	0,2483251	ML(F)_elastic_(6)_with()
33	0,1934791	0,2490736	ARF(3) without
34	0,1934791	0,2490736	ARN(3) without
35	0,1936887	0,2493171	ARF(3) with
36	0,1936887	0,2493171	ARN(3) with
37	0,2028996	0,2539566	ARF(6) with
38	0,2028996	0,2539566	ARN(6) with
39	0,2031357	0,2546284	ARF(6) without
40	0,2031357	0,2546284	ARN(6) without
41	0,2202431	0,2569118	ML(N)_linear_(1)_with()
42	0,2222192	0,2585630	ML(N)_ridge_(1)_with(out)
43	0,2299377	0,2677190	ML(N)_ridge_(3)_with(out)
44	0,2344954	0,2679936	ARXN(1) without
45	0,2321651	0,2699021	ML(N)_linear_(3)_with()
46	0,2331682	0,2735021	ML(N)_linear_(6)_with()
47	0,2410440	0,2763432	ML(N)_elastic_(1)_with(out)
48	0,2418381	0,2772778	ML(N)_linear_(1)_with(out)
49	0,2246311	0,2774719	ML(F)_linear_(1)_with()
50	0,2416609	0,2801585	ARXN(3) without
51	0,2358590	0,2801897	ML(N)_elastic_(6)_with(out)

Sentiment Index

	MAE	RMSE	model
1	0,1891502	0,2348348	ARF(1) without
2	0,1891502	0,2348348	ARN(1) without
3	0,1892868	0,2350908	ML(N)_elastic_(1)_with()
4	0,1903251	0,2358214	ML(F)_elastic_(1)_with()
5	0,1910185	0,2360421	ML(N)_lasso_(6)_with()
6	0,1898065	0,2362562	ARF(1) with
7	0,1898065	0,2362562	ARN(1) with
8	0,1923704	0,2367763	ML(F)_lasso_(6)_with()
9	0,1913913	0,2373576	ML(F)_lasso_(1)_with(out)
10	0,1913913	0,2373576	ML(F)_lasso_(3)_with(out)
11	0,1916286	0,2377188	ML(N)_lasso_(1)_with(out)
12	0,1916286	0,2377188	ML(N)_lasso_(3)_with(out)
13	0,1936244	0,2386878	ML(F)_ridge_(1)_with()
14	0,1940135	0,2393055	ML(N)_ridge_(1)_with()
15	0,1942704	0,2394208	ML(N)_ridge_(3)_with()
16	0,1944579	0,2394451	ML(F)_ridge_(3)_with()
17	0,1943392	0,2394755	ML(N)_ridge_(6)_with()
18	0,1942772	0,2396542	ML(F)_ridge_(6)_with()
19	0,1947402	0,2399775	ML(N)_lasso_(1)_with()
20	0,1947402	0,2399775	ML(N)_lasso_(3)_with()
21	0,1954413	0,2406585	ML(F)_lasso_(1)_with()
22	0,1954413	0,2406585	ML(F)_lasso_(3)_with()
23	0,1921383	0,2451218	ML(N)_elastic_(3)_with()
24	0,1930377	0,2463249	ML(F)_elastic_(3)_with()
25	0,1936647	0,2465369	ML(N)_elastic_(6)_with()
26	0,1965039	0,2490455	ML(F)_elastic_(6)_with()
27	0,1934791	0,2490736	ARF(3) without
28	0,1934791	0,2490736	ARN(3) without
29	0,1936887	0,2493171	ARF(3) with
30	0,1936887	0,2493171	ARN(3) with
31	0,1946043	0,2498188	ML(N)_ridge_(3)_with(out)
32	0,1936258	0,2501179	ML(N)_ridge_(1)_with(out)
33	0,1946511	0,2502605	ML(N)_ridge_(6)_with(out)
34	0,1967407	0,2508911	ML(N)_lasso_(6)_with(out)
35	0,1980991	0,2536174	ML(F)_lasso_(6)_with(out)
36	0,2028996	0,2539566	ARF(6) with
37	0,2028996	0,2539566	ARN(6) with
38	0,2031357	0,2546284	ARF(6) without
39	0,2031357	0,2546284	ARN(6) without
40	0,2037438	0,2596033	ML(F)_ridge_(6)_with(out)
41	0,2244841	0,2727839	ARXF(1) with
42	0,2224885	0,2832059	ML(F)_ridge_(3)_with(out)
43	0,2410774	0,2908876	ARXN(1) with
44	0,2355289	0,2972742	ML(F)_ridge_(1)_with(out)
45	0,2448362	0,3017053	ML(N)_boost_(6)_with()
46	0,2289575	0,3072650	ML(N)_linear_(1)_with()
47	0,2630757	0,3072687	ARXF(3) with
48	0,2623059	0,3123181	ARXF(6) with
49	0,2510818	0,3166992	ML(N)_linear_(3)_with()
50	0,2610597	0,3172771	ML(F)_linear_(1)_with()
51	0,2613861	0,3195628	ML(N)_forest_(3)_with()

52	0,2396195	0,2859070	ML(F)_elastic_(3)_with(out)
53	0,2371379	0,2865720	ML(N)_boost_(3)_with(out)
54	0,2371379	0,2865720	ML(N)_boost_(6)_with(out)
55	0,2388382	0,2867699	ML(F)_linear_(6)_with(out)
56	0,2373254	0,2868414	ML(N)_boost_(3)_with()
57	0,2388904	0,2872208	ML(N)_boost_(6)_with()
58	0,2401706	0,2901222	ML(N)_boost_(1)_with(out)
59	0,2376052	0,2905283	ML(N)_linear_(1)_with()
60	0,2327901	0,2911635	ML(N)_forest_(1)_with()
61	0,2435743	0,2928179	ML(F)_linear_(3)_with(out)
62	0,2412182	0,2931009	ML(N)_boost_(1)_with()
63	0,2246150	0,2934365	ML(N)_linear_(3)_with()
64	0,2408094	0,2936223	ARXN(6) with
65	0,2501862	0,2939116	ARXF(3) with
66	0,2462156	0,2943541	ARXF(3) without
67	0,2301502	0,2975559	ML(N)_linear_(6)_with()
68	0,2322451	0,3050863	ML(N)_elastic_(3)_with(out)
69	0,2286860	0,3052133	ML(N)_elastic_(6)_with(out)
70	0,2401935	0,3071718	ML(N)_forest_(6)_with()
71	0,2338202	0,3084314	ML(N)_linear_(3)_with(out)
72	0,2441778	0,3085272	ML(N)_forest_(3)_with()
73	0,2349224	0,3119672	ML(N)_linear_(6)_with(out)
74	0,2476064	0,3137886	ML(N)_elastic_(1)_with(out)
75	0,2432394	0,3146710	ML(N)_forest_(1)_with(out)
76	0,2558371	0,3151010	ML(F)_forest_(3)_with()
77	0,2650370	0,3161290	ARXF(6) without
78	0,2498022	0,3170774	ML(N)_linear_(1)_with(out)
79	0,2626081	0,3184647	ML(F)_forest_(6)_with()
80	0,2564037	0,3222465	ML(N)_forest_(6)_with(out)
81	0,2664990	0,3225115	ARXF(6) with
82	0,2532216	0,3265557	ML(N)_forest_(3)_with(out)
83	0,2558415	0,3309869	ARXN(3) without
84	0,2654204	0,3327086	ML(F)_forest_(1)_with()
85	0,2599650	0,3343079	ARXN(6) without
86	0,2776195	0,3345903	ML(F)_boost_(6)_with()
87	0,2698229	0,3350231	ML(F)_boost_(1)_with()
88	0,2652842	0,3355257	ARXN(1) without
89	0,2786392	0,3453729	ML(F)_forest_(3)_with(out)
90	0,2885456	0,3653998	ML(F)_boost_(6)_with(out)
91	0,2931072	0,3663387	ML(F)_forest_(1)_with(out)
92	0,2949881	0,3672040	ML(F)_forest_(6)_with(out)
93	0,3153515	0,3723706	ARXN(1) with
94	0,3070178	0,3730202	ML(F)_boost_(3)_with()
95	0,3031030	0,3859534	ML(F)_boost_(1)_with(out)
96	0,3292090	0,4034857	ML(F)_boost_(3)_with(out)

52	0,2394340	0,2995681	ML(N)_forest_(3)_with(out)
53	0,2461782	0,3014338	ML(N)_linear_(3)_with(out)
54	0,2461385	0,3014558	ARXN(3) without
55	0,2402516	0,3038361	ML(N)_forest_(6)_with()
56	0,2479719	0,3043042	ML(N)_elastic_(3)_with(out)
57	0,2558228	0,3087164	ML(F)_linear_(1)_with()
58	0,2468659	0,3102964	ARXN(6) without
59	0,2459015	0,3103465	ML(N)_forest_(6)_with(out)
60	0,2692177	0,3171343	ARXF(6) with
61	0,2691611	0,3185221	ARXF(3) with
62	0,2529333	0,3212987	ML(N)_boost_(6)_with()
63	0,2587788	0,3227602	ML(N)_boost_(6)_with(out)
64	0,2759555	0,3300336	ML(F)_linear_(3)_with()
65	0,2763645	0,3329021	ML(F)_linear_(6)_with()
66	0,2698867	0,3336439	ML(F)_forest_(1)_with()
67	0,2763259	0,3417423	ML(F)_forest_(3)_with()
68	0,2734450	0,3418141	ARXN(1) with
69	0,2729216	0,3427386	ARXN(1) without
70	0,2832266	0,3440156	ML(F)_boost_(3)_with()
71	0,2764822	0,3442965	ML(N)_elastic_(1)_with(out)
72	0,2792467	0,3476989	ML(N)_linear_(1)_with(out)
73	0,2807071	0,3482160	ML(F)_forest_(6)_with()
74	0,2868386	0,3508853	ML(F)_boost_(3)_with(out)
75	0,2909490	0,3582688	ML(F)_forest_(6)_with(out)
76	0,2991165	0,3620358	ML(F)_forest_(3)_with(out)
77	0,2959294	0,3645125	ML(F)_boost_(6)_with()
78	0,3020208	0,3655518	ML(F)_boost_(6)_with(out)
79	0,3032029	0,3683745	ML(F)_forest_(1)_with(out)
80	0,3095629	0,3723467	ML(F)_boost_(1)_with()
81	0,3030735	0,3801481	ML(N)_boost_(3)_with(out)
82	0,3032738	0,3804422	ML(N)_boost_(1)_with(out)
83	0,3076493	0,3846714	ARXN(3) with
84	0,3123563	0,3874441	ML(N)_boost_(3)_with()
85	0,3165394	0,3908594	ML(N)_boost_(1)_with()
86	0,3383158	0,4055037	ML(F)_boost_(1)_with(out)
87	0,3294461	0,4074036	ML(F)_elastic_(1)_with(out)
88	0,3285601	0,4082884	ARXF(1) without
89	0,3380900	0,4194616	ML(F)_linear_(1)_with(out)
90	0,3597161	0,4401415	ML(F)_elastic_(6)_with(out)
91	0,3650208	0,4474122	ML(F)_elastic_(3)_with(out)
92	0,3772878	0,4575286	ML(F)_linear_(6)_with(out)
93	0,3730646	0,4603945	ARXF(3) without
94	0,3788895	0,4642662	ML(F)_linear_(3)_with(out)
95	0,3878374	0,4656798	ARXF(6) without
96	0,4628649	0,5414038	ARXN(6) with

52	0,2407791	0,2802591	ML(N)_elastic_(3)_with(out)
53	0,2345833	0,2830923	ARXF(1) with
54	0,2459140	0,2857787	ML(N)_linear_(3)_with(out)
55	0,2406645	0,2898786	ML(N)_linear_(6)_with(out)
56	0,2394133	0,2933175	ARXN(1) with
57	0,2402222	0,2977817	ARXF(1) without
58	0,2413807	0,3037598	ML(N)_forest_(1)_with()
59	0,2451429	0,3046145	ML(F)_linear_(3)_with()
60	0,2444254	0,3048049	ML(F)_elastic_(1)_with(out)
61	0,2470177	0,3063359	ML(N)_forest_(3)_with()
62	0,2491843	0,3077632	ML(F)_linear_(6)_with()
63	0,2473417	0,3083437	ML(F)_linear_(1)_with(out)
64	0,2502646	0,3122293	ML(N)_forest_(6)_with()
65	0,2611500	0,3143965	ML(F)_boost_(6)_with(out)
66	0,2440464	0,3156660	ML(F)_forest_(1)_with()
67	0,2633391	0,3159648	ML(F)_boost_(1)_with()
68	0,2626064	0,3167992	ML(F)_boost_(1)_with(out)
69	0,2645881	0,3178194	ML(F)_boost_(3)_with(out)
70	0,2659707	0,3193576	ML(F)_boost_(3)_with()
71	0,2491573	0,3194834	ML(N)_forest_(1)_with(out)
72	0,2683823	0,3199088	ARXN(3) with
73	0,2588501	0,3206154	ML(N)_forest_(6)_with(out)
74	0,2669841	0,3228557	ML(F)_boost_(6)_with()
75	0,2624529	0,3252083	ML(N)_boost_(6)_with()
76	0,2625481	0,3273495	ML(N)_forest_(3)_with(out)
77	0,2519687	0,3274656	ML(F)_forest_(6)_with()
78	0,2823465	0,3290697	ARXF(6) with
79	0,2622675	0,3290724	ML(N)_boost_(1)_with()
80	0,2637266	0,3330956	ML(F)_elastic_(3)_with(out)
81	0,2686448	0,3345004	ML(F)_elastic_(6)_with(out)
82	0,2656218	0,3356002	ML(N)_boost_(3)_with()
83	0,2844417	0,3398929	ARXF(3) with
84	0,2693114	0,3407198	ARXF(3) without
85	0,2761555	0,3408077	ARXF(6) without
86	0,2718249	0,3425649	ML(F)_linear_(3)_with(out)
87	0,2647490	0,3433859	ML(F)_forest_(1)_with(out)
88	0,2589242	0,3444622	ML(F)_forest_(3)_with()
89	0,2755809	0,3459336	ML(N)_boost_(6)_with(out)
90	0,2794242	0,3463243	ML(F)_linear_(6)_with(out)
91	0,2785283	0,3491480	ML(F)_forest_(6)_with(out)
92	0,2782144	0,3501902	ML(N)_boost_(1)_with(out)
93	0,2762727	0,3538389	ML(N)_boost_(3)_with(out)
94	0,2974420	0,3627323	ARXN(6) without
95	0,2868456	0,3718469	ML(F)_forest_(3)_with(out)
96	0,3185174	0,3878411	ARXN(6) with

52	0,2733328	0,3208962	ARXN(6) with
53	0,2609207	0,3210336	ML(F)_linear_(6)_with()
54	0,2645874	0,3217811	ML(N)_boost_(6)_with(out)
55	0,2569747	0,3223477	ML(N)_linear_(6)_with()
56	0,2529525	0,3225567	ML(N)_forest_(1)_with()
57	0,2684184	0,3271604	ML(F)_boost_(1)_with()
58	0,2689519	0,3287018	ML(F)_boost_(3)_with()
59	0,2708704	0,3324736	ML(F)_boost_(6)_with()
60	0,2803283	0,3346442	ARXN(3) with
61	0,2696901	0,3351827	ML(N)_forest_(6)_with(out)
62	0,2663684	0,3383229	ML(F)_linear_(3)_with()
63	0,2695740	0,3388829	ML(N)_forest_(3)_with(out)
64	0,2701244	0,3401183	ML(N)_forest_(6)_with()
65	0,2820575	0,3444097	ML(F)_forest_(3)_with()
66	0,2633596	0,3477660	ML(F)_forest_(1)_with()
67	0,2728862	0,3492413	ARXN(1) without
68	0,2754464	0,3539413	ML(N)_elastic_(1)_with(out)
69	0,2756728	0,3555095	ML(F)_forest_(6)_with()
70	0,2900292	0,3556434	ML(N)_elastic_(3)_with(out)
71	0,2772883	0,3561376	ML(F)_forest_(6)_with(out)
72	0,2936957	0,3568659	ML(F)_boost_(1)_with(out)
73	0,2845899	0,3591103	ML(N)_forest_(1)_with(out)
74	0,2837740	0,3596649	ML(N)_boost_(3)_with()
75	0,2940157	0,3601436	ARXN(3) without
76	0,2944716	0,3609151	ML(F)_forest_(3)_with(out)
77	0,2818341	0,3612390	ML(N)_linear_(1)_with(out)
78	0,2899540	0,3621831	ML(F)_elastic_(6)_with(out)
79	0,3006718	0,3633838	ML(F)_boost_(6)_with(out)
80	0,3003709	0,3656940	ML(N)_linear_(3)_with(out)
81	0,2968357	0,3668013	ML(N)_elastic_(6)_with(out)
82	0,3023665	0,3688885	ML(F)_boost_(3)_with(out)
83	0,2990455	0,3722968	ML(F)_linear_(6)_with(out)
84	0,3044995	0,3761899	ARXF(6) without
85	0,3140906	0,3787279	ARXF(1) without
86	0,2965501	0,3798982	ML(F)_forest_(1)_with(out)
87	0,3133006	0,3801252	ML(F)_elastic_(1)_with(out)
88	0,3087215	0,3815843	ML(N)_linear_(6)_with(out)
89	0,3169500	0,3841860	ML(F)_linear_(1)_with(out)
90	0,3023298	0,3859924	ML(N)_boost_(1)_with()
91	0,2895280	0,3896717	ML(N)_boost_(1)_with(out)
92	0,3236142	0,3903171	ARXN(6) without
93	0,3018779	0,3910608	ML(N)_boost_(3)_with(out)
94	0,3115958	0,3938580	ML(F)_elastic_(3)_with(out)
95	0,3153818	0,3976573	ARXF(3) without
96	0,3211444	0,4060471	ML(F)_linear_(3)_with(out)

Appendix H4. EX models

Frequency Index [Basic]

	MAE	RMSE	model
1	1,132427	1,409377	ML(F)_boost_(3)_with()
2	1,127618	1,419056	ML(F)_boost_(6)_with()
3	1,107894	1,420990	ML(F)_boost_(1)_with(out)
4	1,124517	1,425513	ML(F)_boost_(1)_with()
5	1,201777	1,435941	ML(F)_forest_(3)_with()
6	1,199803	1,439757	ML(F)_forest_(6)_with()
7	1,194405	1,439928	ML(F)_forest_(3)_with(out)
8	1,170353	1,444595	ML(F)_boost_(3)_with(out)
9	1,196304	1,446442	ML(F)_linear_(3)_with()
10	1,176439	1,447547	ML(F)_boost_(6)_with(out)
11	1,167845	1,450207	ML(F)_lasso_(6)_with()
12	1,168287	1,450318	ML(N)_lasso_(6)_with()
13	1,224480	1,451118	ML(F)_elastic_(3)_with(out)
14	1,184407	1,452495	ML(F)_ridge_(6)_with(out)
15	1,171973	1,452535	ML(F)_lasso_(3)_with()
16	1,216030	1,452684	ML(F)_forest_(1)_with()
17	1,227103	1,452776	ML(F)_linear_(3)_with(out)
18	1,172812	1,453143	ML(N)_lasso_(3)_with()
19	1,179487	1,453204	ML(F)_forest_(6)_with(out)
20	1,175809	1,457299	ML(F)_lasso_(6)_with(out)
21	1,193172	1,458218	ML(F)_ridge_(3)_with(out)
22	1,176328	1,458241	ML(F)_lasso_(3)_with(out)
23	1,177430	1,460320	ML(N)_lasso_(6)_with(out)
24	1,179842	1,460395	ML(N)_ridge_(6)_with()
25	1,184777	1,464437	ML(N)_ridge_(3)_with()
26	1,180024	1,465625	ML(N)_lasso_(3)_with(out)
27	1,188694	1,467924	ML(N)_ridge_(6)_with(out)
28	1,177218	1,468823	ML(F)_ridge_(6)_with()
29	1,186012	1,469993	ML(F)_ridge_(3)_with()
30	1,195440	1,477558	ML(N)_ridge_(3)_with(out)
31	1,186569	1,479160	ML(N)_lasso_(1)_with(out)
32	1,189510	1,481278	ML(N)_lasso_(1)_with()
33	1,196035	1,481569	ML(F)_ridge_(1)_with(out)
34	1,246728	1,483125	ARXF(3) without
35	1,212922	1,484695	ML(N)_elastic_(3)_with()
36	1,212571	1,488480	ARF(3) without
37	1,212571	1,488480	ARN(3) without
38	1,202610	1,489304	ML(N)_ridge_(1)_with(out)
39	1,201695	1,490782	ML(N)_ridge_(1)_with()
40	1,225520	1,494650	ML(F)_elastic_(3)_with()
41	1,209616	1,499860	ML(F)_ridge_(1)_with()
42	1,221019	1,500921	ML(F)_lasso_(1)_with(out)
43	1,232061	1,502988	ARF(3) with
44	1,232061	1,502988	ARN(3) with
45	1,202562	1,504703	ML(F)_lasso_(1)_with()
46	1,233326	1,506211	ARXN(3) with
47	1,250133	1,507134	ARXF(3) with
48	1,240559	1,511749	ML(N)_elastic_(6)_with()
49	1,217081	1,516494	ML(F)_linear_(6)_with()
50	1,235652	1,520107	ML(F)_linear_(1)_with()
51	1,249239	1,521939	ML(F)_elastic_(6)_with()

Frequency Index [Full]

	MAE	RMSE	model
1	1,167817	1,450216	ML(N)_lasso_(6)_with()
2	1,167434	1,450494	ML(F)_lasso_(6)_with()
3	1,172009	1,452545	ML(F)_lasso_(3)_with()
4	1,172059	1,452562	ML(N)_lasso_(3)_with()
5	1,173475	1,453352	ML(F)_lasso_(6)_with(out)
6	1,175679	1,457067	ML(N)_lasso_(6)_with(out)
7	1,176587	1,458721	ML(F)_lasso_(3)_with(out)
8	1,176912	1,459328	ML(N)_lasso_(3)_with(out)
9	1,179842	1,460396	ML(N)_ridge_(6)_with()
10	1,183731	1,463047	ML(N)_ridge_(3)_with()
11	1,188694	1,467924	ML(N)_ridge_(6)_with(out)
12	1,188800	1,468302	ML(N)_ridge_(3)_with(out)
13	1,186569	1,479160	ML(F)_lasso_(1)_with()
14	1,186569	1,479160	ML(N)_lasso_(1)_with(out)
15	1,186569	1,479160	ML(F)_lasso_(1)_with(out)
16	1,201606	1,479238	ML(F)_ridge_(1)_with(out)
17	1,197083	1,480485	ML(F)_ridge_(3)_with()
18	1,189336	1,481561	ML(N)_lasso_(1)_with()
19	1,212714	1,484391	ML(F)_elastic_(3)_with()
20	1,212871	1,484689	ML(N)_elastic_(3)_with()
21	1,176924	1,487958	ARXF(1) with
22	1,212571	1,488480	ARF(3) without
23	1,212571	1,488480	ARN(3) without
24	1,207062	1,488503	ML(F)_ridge_(6)_with()
25	1,202610	1,489304	ML(N)_ridge_(1)_with(out)
26	1,201695	1,490782	ML(N)_ridge_(1)_with()
27	1,232083	1,495558	ML(F)_ridge_(3)_with(out)
28	1,209616	1,499860	ML(F)_ridge_(1)_with()
29	1,244430	1,500720	ML(F)_ridge_(6)_with(out)
30	1,232061	1,502988	ARF(3) with
31	1,232061	1,502988	ARN(3) with
32	1,209037	1,503346	ML(F)_boost_(6)_with()
33	1,220935	1,504005	ML(N)_boost_(3)_with(out)
34	1,242179	1,509228	ARXN(3) with
35	1,240438	1,511436	ML(F)_elastic_(6)_with()
36	1,240496	1,511734	ML(N)_elastic_(6)_with()
37	1,207065	1,523360	ML(N)_boost_(6)_with(out)
38	1,214916	1,523693	ML(N)_boost_(1)_with()
39	1,232820	1,525497	ARXF(3) with
40	1,218241	1,526211	ARXN(1) with
41	1,265023	1,527250	ARF(6) without
42	1,265023	1,527250	ARN(6) without
43	1,248260	1,527287	ARF(6) with
44	1,248260	1,527287	ARN(6) with
45	1,252552	1,527369	ML(F)_boost_(6)_with(out)
46	1,223353	1,527857	ML(N)_forest_(1)_with()
47	1,243161	1,530010	ML(N)_boost_(3)_with()
48	1,239849	1,531712	ARXN(6) with
49	1,253039	1,534090	ML(F)_boost_(1)_with()
50	1,256289	1,543284	ARXF(6) with
51	1,239952	1,548442	ML(N)_boost_(6)_with()

Frequency Index [Optimal]

	MAE	RMSE	model
1	1,167817	1,450216	ML(F)_lasso_(6)_with()
2	1,167390	1,450477	ML(N)_lasso_(6)_with()
3	1,182570	1,451362	ML(F)_boost_(6)_with()
4	1,186524	1,452207	ML(F)_ridge_(6)_with(out)
5	1,172144	1,452618	ML(F)_lasso_(3)_with()
6	1,171881	1,452677	ML(N)_lasso_(3)_with()
7	1,173345	1,453147	ML(N)_lasso_(6)_with(out)
8	1,175031	1,455931	ML(N)_lasso_(3)_with(out)
9	1,175679	1,457067	ML(F)_lasso_(6)_with(out)
10	1,175893	1,458392	ML(N)_ridge_(6)_with(out)
11	1,177236	1,459945	ML(F)_lasso_(3)_with(out)
12	1,178303	1,460825	ML(N)_ridge_(6)_with()
13	1,185590	1,461644	ML(F)_ridge_(3)_with(out)
14	1,183731	1,463047	ML(N)_ridge_(3)_with()
15	1,183247	1,463739	ML(F)_ridge_(3)_with()
16	1,178195	1,464443	ML(F)_ridge_(6)_with()
17	1,188784	1,468298	ML(N)_ridge_(3)_with(out)
18	1,225179	1,478975	ML(F)_boost_(1)_with()
19	1,186493	1,479139	ML(F)_lasso_(1)_with()
20	1,186569	1,479160	ML(N)_lasso_(1)_with(out)
21	1,186569	1,479160	ML(F)_lasso_(1)_with(out)
22	1,189716	1,481035	ML(N)_lasso_(1)_with()
23	1,230629	1,481087	ML(F)_boost_(6)_with(out)
24	1,257016	1,482269	ML(F)_forest_(6)_with(out)
25	1,212779	1,484455	ML(F)_elastic_(3)_with()
26	1,212943	1,484721	ML(N)_elastic_(3)_with()
27	1,212571	1,488480	ARF(3) without
28	1,212571	1,488480	ARN(3) without
29	1,202610	1,489304	ML(N)_ridge_(1)_with(out)
30	1,202610	1,489304	ML(F)_ridge_(1)_with(out)
31	1,203424	1,492148	ML(N)_ridge_(1)_with()
32	1,203424	1,492148	ML(F)_ridge_(1)_with()
33	1,171856	1,502802	ML(N)_boost_(1)_with(out)
34	1,232061	1,502988	ARF(3) with
35	1,232061	1,502988	ARN(3) with
36	1,255531	1,504768	ML(F)_forest_(1)_with()
37	1,237768	1,506056	ML(F)_linear_(3)_with()
38	1,240547	1,511761	ML(N)_elastic_(6)_with()
39	1,247305	1,516422	ML(F)_forest_(6)_with()
40	1,181882	1,516945	ML(N)_boost_(6)_with(out)
41	1,272628	1,518041	ML(F)_boost_(1)_with(out)
42	1,249271	1,521884	ML(F)_elastic_(6)_with()
43	1,286886	1,522597	ML(F)_linear_(1)_with()
44	1,299336	1,524179	ML(F)_forest_(3)_with(out)
45	1,265023	1,527250	ARF(6) without
46	1,265023	1,527250	ARN(6) without
47	1,248260	1,527287	ARF(6) with
48	1,248260	1,527287	ARN(6) with
49	1,200616	1,532331	ML(N)_boost_(6)_with()
50	1,278635	1,535887	ML(F)_boost_(3)_with()
51	1,197931	1,537182	ML(N)_boost_(1)_with()

Sentiment Index

	MAE	RMSE	model
1	1,149102	1,447485	ML(N)_boost_(6)_with(out)
2	1,168075	1,450251	ML(F)_lasso_(6)_with()
3	1,172144	1,452618	ML(F)_lasso_(3)_with()
4	1,173853	1,456294	ML(N)_lasso_(6)_with()
5	1,176111	1,457947	ML(N)_lasso_(3)_with()
6	1,176587	1,458721	ML(F)_lasso_(6)_with(out)
7	1,177236	1,459945	ML(F)_lasso_(3)_with(out)
8	1,179842	1,460396	ML(F)_ridge_(6)_with()
9	1,143583	1,460410	ML(N)_boost_(3)_with(out)
10	1,121626	1,461950	ML(N)_ridge_(6)_with(out)
11	1,183731	1,463047	ML(F)_ridge_(3)_with()
12	1,121486	1,466509	ML(N)_ridge_(3)_with(out)
13	1,188694	1,467924	ML(F)_ridge_(6)_with(out)
14	1,188800	1,468302	ML(F)_ridge_(3)_with(out)
15	1,174341	1,470901	ML(F)_boost_(1)_with()
16	1,192667	1,476862	ML(N)_ridge_(3)_with()
17	1,149571	1,477925	ML(N)_forest_(6)_with(out)
18	1,194629	1,478271	ML(N)_ridge_(6)_with()
19	1,186569	1,479160	ML(N)_lasso_(1)_with(out)
20	1,186569	1,479160	ML(F)_lasso_(1)_with(out)
21	1,186569	1,479160	ML(N)_lasso_(3)_with(out)
22	1,186569	1,479160	ML(N)_lasso_(6)_with(out)
23	1,176128	1,479907	ML(N)_forest_(3)_with(out)
24	1,188529	1,480799	ML(N)_lasso_(1)_with()
25	1,189510	1,481278	ML(F)_lasso_(1)_with()
26	1,210804	1,482889	ML(F)_elastic_(3)_with()
27	1,207351	1,483138	ML(N)_elastic_(3)_with()
28	1,212571	1,488480	ARF(3) without
29	1,212571	1,488480	ARN(3) without
30	1,202610	1,489304	ML(N)_ridge_(1)_with(out)
31	1,202610	1,489304	ML(F)_ridge_(1)_with(out)
32	1,201695	1,490782	ML(N)_ridge_(1)_with()
33	1,201695	1,490782	ML(F)_ridge_(1)_with()
34	1,165489	1,497839	ML(F)_boost_(6)_with(out)
35	1,197037	1,502365	ML(N)_boost_(1)_with()
36	1,232061	1,502988	ARF(3) with
37	1,232061	1,502988	ARN(3) with
38	1,167798	1,503052	ML(N)_boost_(1)_with(out)
39	1,189333	1,503752	ML(N)_boost_(6)_with()
40	1,188221	1,509783	ML(F)_boost_(6)_with()
41	1,172855	1,515576	ML(N)_forest_(1)_with()
42	1,245892	1,520418	ML(N)_elastic_(6)_with()
43	1,248296	1,520999	ML(F)_elastic_(6)_with()
44	1,265023	1,527250	ARF(6) without
45	1,265023	1,527250	ARN(6) without
46	1,248260	1,527287	ARF(6) with
47	1,248260	1,527287	ARN(6) with
48	1,207241	1,531239	ML(F)_boost_(3)_with(out)
49	1,220338	1,533530	ML(F)_boost_(1)_with(out)
50	1,206338	1,538531	ML(N)_elastic_(1)_with()
51	1,206339	1,544431	ML(N)_forest_(1)_with(out)

52	1,265023	1,527250	ARF(6) without
53	1,265023	1,527250	ARN(6) without
54	1,248260	1,527287	ARF(6) with
55	1,248260	1,527287	ARN(6) with
56	1,206472	1,534333	ARXF(1) with
57	1,229144	1,536154	ML(N)_boost_(1)_with()
58	1,213635	1,540727	ML(N)_elastic_(1)_with()
59	1,252696	1,544174	ML(F)_elastic_(6)_with(out)
60	1,230881	1,545068	ML(N)_boost_(1)_with(out)
61	1,222815	1,546896	ML(N)_boost_(6)_with()
62	1,245305	1,549661	ML(N)_boost_(3)_with()
63	1,299499	1,554889	ML(F)_elastic_(1)_with(out)
64	1,233655	1,554957	ML(F)_elastic_(1)_with()
65	1,261695	1,555201	ML(F)_linear_(6)_with(out)
66	1,234333	1,555683	ARF(1) with
67	1,234333	1,555683	ARN(1) with
68	1,270226	1,556798	ML(N)_boost_(3)_with(out)
69	1,301356	1,557732	ML(F)_linear_(1)_with(out)
70	1,257727	1,559228	ARXN(1) with
71	1,312389	1,565300	ML(N)_linear_(3)_with()
72	1,269238	1,571873	ML(N)_boost_(6)_with(out)
73	1,236332	1,575526	ARF(1) without
74	1,236332	1,575526	ARN(1) without
75	1,267335	1,576898	ARXF(6) with
76	1,299259	1,578379	ARXN(6) with
77	1,328539	1,578736	ML(N)_elastic_(3)_with(out)
78	1,318578	1,579425	ML(N)_linear_(6)_with()
79	1,331493	1,580231	ML(N)_linear_(3)_with(out)
80	1,319762	1,581197	ARXF(1) without
81	1,312538	1,592887	ML(F)_forest_(1)_with(out)
82	1,327353	1,596856	ARXF(6) without
83	1,377821	1,605042	ML(N)_elastic_(6)_with(out)
84	1,383859	1,609551	ML(N)_linear_(6)_with(out)
85	1,364962	1,624180	ARXN(3) without
86	1,435699	1,672915	ARXN(6) without
87	1,349619	1,677371	ML(N)_linear_(1)_with()
88	1,430035	1,737350	ML(N)_elastic_(1)_with(out)
89	1,433352	1,741421	ML(N)_linear_(1)_with(out)
90	1,410095	1,742075	ML(N)_forest_(3)_with()
91	1,474656	1,774815	ARXN(1) without
92	1,458883	1,793489	ML(N)_forest_(1)_with()
93	1,504357	1,799137	ML(N)_forest_(3)_with(out)
94	1,582487	1,855586	ML(N)_forest_(6)_with()
95	1,578126	1,907448	ML(N)_forest_(1)_with(out)
96	1,605237	1,946544	ML(N)_forest_(6)_with(out)

52	1,233550	1,554671	ML(F)_elastic_(1)_with()
53	1,279932	1,554784	ML(F)_boost_(3)_with(out)
54	1,233750	1,555222	ML(N)_elastic_(1)_with()
55	1,234333	1,555683	ARF(1) with
56	1,234333	1,555683	ARN(1) with
57	1,278420	1,565794	ML(F)_boost_(3)_with()
58	1,287116	1,575109	ML(F)_forest_(6)_with()
59	1,236332	1,575526	ARF(1) without
60	1,236332	1,575526	ARN(1) without
61	1,276690	1,582997	ML(F)_forest_(3)_with(out)
62	1,234183	1,583563	ML(N)_forest_(1)_with(out)
63	1,298588	1,583578	ML(F)_forest_(6)_with(out)
64	1,293255	1,584634	ML(F)_boost_(1)_with(out)
65	1,264029	1,590903	ML(N)_forest_(3)_with(out)
66	1,288686	1,595469	ML(N)_forest_(3)_with()
67	1,259138	1,602851	ML(N)_boost_(1)_with(out)
68	1,329207	1,608074	ML(F)_linear_(3)_with()
69	1,294436	1,610827	ML(N)_forest_(6)_with()
70	1,349936	1,616718	ML(F)_linear_(1)_with()
71	1,299260	1,620935	ML(N)_linear_(3)_with()
72	1,323033	1,629077	ML(N)_linear_(6)_with()
73	1,358530	1,637375	ML(F)_linear_(6)_with()
74	1,287342	1,648334	ML(N)_forest_(6)_with(out)
75	1,284115	1,671458	ML(N)_linear_(1)_with()
76	1,340450	1,693701	ML(N)_elastic_(3)_with(out)
77	1,347143	1,702787	ML(N)_linear_(3)_with(out)
78	1,321165	1,715454	ML(N)_elastic_(1)_with(out)
79	1,416006	1,715649	ML(F)_forest_(3)_with()
80	1,325571	1,720670	ML(N)_linear_(1)_with(out)
81	1,425697	1,721469	ML(N)_elastic_(6)_with(out)
82	1,432962	1,729232	ML(N)_linear_(6)_with(out)
83	1,435084	1,734791	ML(F)_elastic_(6)_with(out)
84	1,394228	1,737473	ML(F)_elastic_(3)_with(out)
85	1,376721	1,739377	ARXN(3) without
86	1,448791	1,741249	ML(F)_elastic_(1)_with(out)
87	1,348081	1,741602	ARXN(1) without
88	1,403694	1,749376	ML(F)_linear_(3)_with(out)
89	1,449690	1,750513	ML(F)_linear_(6)_with(out)
90	1,455821	1,752770	ML(F)_linear_(1)_with(out)
91	1,432823	1,758016	ARXF(3) without
92	1,430073	1,760204	ML(F)_forest_(1)_with()
93	1,477100	1,775252	ARXF(1) without
94	1,487104	1,788117	ARXF(6) without
95	1,523006	1,832615	ARXN(6) without
96	1,540806	1,876092	ML(F)_forest_(1)_with(out)

52	1,280181	1,548688	ML(F)_forest_(3)_with()
53	1,282717	1,549591	ML(F)_linear_(6)_with()
54	1,336363	1,550991	ML(F)_boost_(3)_with(out)
55	1,233592	1,554802	ML(F)_elastic_(1)_with()
56	1,233823	1,555234	ML(N)_elastic_(1)_with()
57	1,246636	1,555279	ARXN(6) with
58	1,234333	1,555683	ARF(1) with
59	1,234333	1,555683	ARN(1) with
60	1,218547	1,567888	ARXN(1) with
61	1,236332	1,575526	ARF(1) without
62	1,236332	1,575526	ARN(1) without
63	1,345435	1,581289	ML(F)_forest_(1)_with(out)
64	1,278363	1,582124	ARXF(1) with
65	1,280856	1,585060	ARXN(3) with
66	1,300578	1,586402	ARXF(3) with
67	1,306951	1,594972	ARXF(6) with
68	1,242189	1,600332	ML(N)_forest_(1)_with()
69	1,287486	1,606332	ML(N)_linear_(3)_with()
70	1,279263	1,611219	ML(N)_forest_(6)_with()
71	1,253689	1,615993	ML(N)_boost_(3)_with()
72	1,303210	1,623499	ML(N)_linear_(6)_with()
73	1,271056	1,631918	ML(N)_boost_(3)_with(out)
74	1,368730	1,641387	ML(F)_elastic_(3)_with(out)
75	1,375607	1,650290	ML(F)_linear_(3)_with(out)
76	1,325969	1,666256	ML(N)_forest_(6)_with(out)
77	1,404770	1,667047	ML(F)_linear_(1)_with(out)
78	1,404870	1,669745	ML(F)_elastic_(1)_with(out)
79	1,289771	1,670485	ML(N)_linear_(1)_with()
80	1,351890	1,675044	ML(N)_forest_(3)_with()
81	1,318736	1,677362	ML(N)_forest_(1)_with(out)
82	1,320956	1,686803	ML(N)_elastic_(3)_with(out)
83	1,418898	1,687735	ARXF(3) without
84	1,355290	1,688039	ML(N)_forest_(3)_with(out)
85	1,327512	1,693737	ML(N)_linear_(3)_with(out)
86	1,440855	1,715517	ARXF(1) without
87	1,422806	1,721793	ML(N)_elastic_(6)_with(out)
88	1,384294	1,726424	ARXN(3) without
89	1,428436	1,728018	ML(F)_elastic_(6)_with(out)
90	1,431889	1,729984	ML(N)_linear_(6)_with(out)
91	1,436429	1,739400	ML(F)_linear_(6)_with(out)
92	1,366305	1,745235	ML(N)_elastic_(1)_with(out)
93	1,370418	1,751100	ML(N)_linear_(1)_with(out)
94	1,395279	1,769152	ARXN(1) without
95	1,509105	1,814681	ARXF(6) without
96	1,531049	1,818040	ARXN(6) without

52	1,222058	1,551307	ML(N)_forest_(6)_with()
53	1,232986	1,551477	ML(F)_elastic_(1)_with()
54	1,265365	1,555020	ARXN(6) with
55	1,234333	1,555683	ARF(1) with
56	1,234333	1,555683	ARN(1) with
57	1,256019	1,559881	ARXN(3) with
58	1,213122	1,563607	ML(F)_boost_(3)_with()
59	1,255635	1,569203	ML(N)_boost_(3)_with()
60	1,258986	1,569329	ML(N)_forest_(3)_with()
61	1,236332	1,575526	ARF(1) without
62	1,236332	1,575526	ARN(1) without
63	1,271690	1,582446	ML(F)_forest_(1)_with()
64	1,281942	1,594384	ML(F)_linear_(3)_with()
65	1,234589	1,606042	ML(N)_elastic_(6)_with(out)
66	1,288555	1,612672	ML(F)_forest_(6)_with(out)
67	1,245298	1,616076	ML(N)_linear_(6)_with(out)
68	1,265191	1,616266	ML(N)_linear_(3)_with()
69	1,291272	1,623074	ML(N)_linear_(6)_with()
70	1,154530	1,627723	ML(N)_elastic_(1)_with(out)
71	1,325630	1,631455	ML(F)_forest_(6)_with()
72	1,157322	1,632408	ML(N)_linear_(1)_with(out)
73	1,321390	1,633827	ML(F)_forest_(3)_with(out)
74	1,241069	1,639846	ML(N)_elastic_(3)_with(out)
75	1,167882	1,640647	ML(N)_linear_(1)_with()
76	1,296597	1,648926	ML(F)_linear_(1)_with()
77	1,252805	1,651416	ML(N)_linear_(3)_with(out)
78	1,348574	1,654141	ML(F)_forest_(3)_with()
79	1,292603	1,655689	ML(F)_elastic_(3)_with(out)
80	1,302580	1,657132	ARXF(3) without
81	1,350840	1,658244	ML(F)_linear_(6)_with()
82	1,352353	1,662516	ARXF(3) with
83	1,302195	1,668458	ML(F)_linear_(3)_with(out)
84	1,252122	1,682730	ARXN(1) without
85	1,389816	1,685332	ARXF(6) with
86	1,385364	1,695487	ARXN(1) with
87	1,401086	1,701270	ARXF(1) with
88	1,363722	1,710341	ARXF(6) without
89	1,322308	1,720751	ARXN(3) without
90	1,360988	1,724785	ML(F)_elastic_(1)_with(out)
91	1,364899	1,731412	ML(F)_linear_(1)_with(out)
92	1,400418	1,733329	ML(F)_elastic_(6)_with(out)
93	1,412884	1,745694	ML(F)_linear_(6)_with(out)
94	1,396307	1,758819	ARXF(1) without
95	1,379594	1,791431	ARXN(6) without
96	1,450946	1,837701	ML(F)_forest_(1)_with(out)

Appendix H5. IM models

Frequency Index [Basic]

	MAE	RMSE	model
1	0,406205	0,632387	ML(F)_boost_(6)_with(out)
2	0,407414	0,634332	ML(F)_boost_(3)_with(out)
3	0,420058	0,639366	ML(F)_elastic_(1)_with()
4	0,420074	0,639377	ML(N)_elastic_(1)_with()
5	0,426621	0,639718	ML(N)_lasso_(3)_with(out)
6	0,426621	0,639718	ML(F)_lasso_(3)_with(out)
7	0,426621	0,639718	ML(N)_lasso_(6)_with(out)
8	0,426621	0,639718	ML(F)_lasso_(6)_with(out)
9	0,428976	0,641126	ML(N)_lasso_(1)_with()
10	0,428976	0,641126	ML(F)_lasso_(1)_with()
11	0,428976	0,641126	ML(N)_lasso_(1)_with(out)
12	0,428976	0,641126	ML(F)_lasso_(1)_with(out)
13	0,428976	0,641126	ML(N)_lasso_(3)_with()
14	0,428976	0,641126	ML(F)_lasso_(3)_with()
15	0,428976	0,641126	ML(N)_lasso_(6)_with()
16	0,428976	0,641126	ML(F)_lasso_(6)_with()
17	0,405220	0,641250	ML(N)_boost_(3)_with(out)
18	0,410442	0,641491	ML(N)_forest_(3)_with(out)
19	0,409567	0,643126	ML(N)_boost_(1)_with()
20	0,429613	0,644347	ML(N)_ridge_(3)_with()
21	0,429613	0,644347	ML(F)_ridge_(3)_with()
22	0,429613	0,644347	ML(N)_ridge_(6)_with()
23	0,429613	0,644347	ML(F)_ridge_(6)_with()
24	0,412631	0,646267	ML(N)_boost_(6)_with(out)
25	0,419509	0,646669	ML(F)_boost_(3)_with()
26	0,420937	0,646950	ML(F)_boost_(6)_with()
27	0,435827	0,647956	ML(F)_ridge_(3)_with(out)
28	0,435827	0,647956	ML(F)_ridge_(6)_with(out)
29	0,435527	0,648329	ML(N)_ridge_(1)_with()
30	0,435527	0,648329	ML(F)_ridge_(1)_with()
31	0,416601	0,648337	ML(N)_forest_(6)_with(out)
32	0,436260	0,649993	ML(N)_ridge_(3)_with(out)
33	0,436233	0,650009	ML(N)_ridge_(6)_with(out)
34	0,434216	0,651416	ARF(1) with
35	0,434216	0,651416	ARN(1) with
36	0,442200	0,654006	ML(N)_forest_(1)_with()
37	0,451781	0,655697	ML(F)_ridge_(1)_with(out)
38	0,451997	0,655781	ML(N)_ridge_(1)_with(out)
39	0,422744	0,658447	ML(N)_boost_(3)_with()
40	0,454777	0,658524	ARF(1) without
41	0,454777	0,658524	ARN(1) without
42	0,427893	0,663238	ML(N)_boost_(6)_with()
43	0,454854	0,672791	ML(N)_forest_(3)_with()
44	0,416898	0,674143	ARF(3) without
45	0,416898	0,674143	ARN(3) without
46	0,448624	0,676304	ML(N)_boost_(1)_with(out)
47	0,424178	0,681047	ML(F)_elastic_(3)_with()
48	0,424149	0,681068	ML(N)_elastic_(3)_with()
49	0,451988	0,685166	ML(N)_forest_(6)_with()
50	0,429979	0,691614	ARF(3) with
51	0,429979	0,691614	ARN(3) with

Frequency Index [Full]

	MAE	RMSE	model
1	0,4026233	0,6226804	ML(N)_boost_(3)_with(out)
2	0,4026233	0,6226804	ML(N)_boost_(6)_with(out)
3	0,4536021	0,6340633	ML(F)_forest_(3)_with(out)
4	0,4091087	0,6358065	ML(N)_boost_(3)_with()
5	0,4091827	0,6359098	ML(N)_boost_(6)_with()
6	0,4352437	0,6360337	ML(N)_forest_(3)_with(out)
7	0,4359503	0,6367039	ML(N)_forest_(6)_with(out)
8	0,4172284	0,6389266	ML(F)_boost_(3)_with(out)
9	0,4200838	0,6393286	ML(F)_elastic_(1)_with()
10	0,4201084	0,6394953	ML(N)_elastic_(1)_with()
11	0,4266212	0,6397183	ML(N)_lasso_(3)_with(out)
12	0,4266212	0,6397183	ML(F)_lasso_(3)_with(out)
13	0,4266212	0,6397183	ML(N)_lasso_(6)_with(out)
14	0,4266212	0,6397183	ML(F)_lasso_(6)_with(out)
15	0,4346408	0,6408681	ML(N)_forest_(6)_with()
16	0,4289758	0,6411265	ML(N)_lasso_(1)_with()
17	0,4289758	0,6411265	ML(F)_lasso_(1)_with()
18	0,4289758	0,6411265	ML(N)_lasso_(1)_with(out)
19	0,4289758	0,6411265	ML(F)_lasso_(1)_with(out)
20	0,4289758	0,6411265	ML(N)_lasso_(3)_with()
21	0,4289758	0,6411265	ML(F)_lasso_(3)_with()
22	0,4289758	0,6411265	ML(N)_lasso_(6)_with()
23	0,4289758	0,6411265	ML(F)_lasso_(6)_with()
24	0,4327437	0,6419509	ML(F)_boost_(6)_with(out)
25	0,4227845	0,6426367	ML(F)_ridge_(3)_with()
26	0,4250200	0,6434533	ML(F)_boost_(6)_with()
27	0,4296135	0,6443474	ML(N)_ridge_(3)_with()
28	0,4296135	0,6443474	ML(N)_ridge_(6)_with()
29	0,4296135	0,6443474	ML(F)_ridge_(6)_with()
30	0,4360426	0,6468003	ML(F)_ridge_(6)_with(out)
31	0,4355270	0,6483290	ML(N)_ridge_(1)_with()
32	0,4355270	0,6483290	ML(F)_ridge_(1)_with()
33	0,4346134	0,6484941	ML(F)_ridge_(3)_with(out)
34	0,4361109	0,6488992	ML(N)_ridge_(3)_with(out)
35	0,4360990	0,6489703	ML(N)_ridge_(6)_with(out)
36	0,4268005	0,6508852	ML(F)_boost_(3)_with()
37	0,4342156	0,6514161	ARF(1) with
38	0,4342156	0,6514161	ARN(1) with
39	0,4549391	0,6533133	ML(F)_ridge_(1)_with(out)
40	0,4548693	0,6552319	ML(N)_forest_(1)_with()
41	0,4423583	0,6553626	ML(N)_forest_(3)_with()
42	0,4519967	0,6557805	ML(N)_ridge_(1)_with(out)
43	0,4547766	0,6585241	ARF(1) without
44	0,4547766	0,6585241	ARN(1) without
45	0,4140461	0,6615187	ML(N)_boost_(1)_with(out)
46	0,4379785	0,6731596	ML(F)_boost_(1)_with()
47	0,4271287	0,6732541	ML(N)_boost_(1)_with()
48	0,4168976	0,6741430	ARF(3) without
49	0,4168976	0,6741430	ARN(3) without
50	0,4595990	0,6758944	ML(F)_boost_(1)_with(out)
51	0,4776472	0,6766131	ML(F)_forest_(1)_with()

Frequency Index [Optimal]

	MAE	RMSE	model
1	0,384419	0,608980	ARXF(1) with
2	0,404268	0,627349	ML(N)_boost_(6)_with(out)
3	0,410699	0,629522	ML(F)_boost_(6)_with(out)
4	0,431072	0,630478	ML(F)_boost_(3)_with()
5	0,430836	0,636497	ML(N)_forest_(3)_with(out)
6	0,409330	0,636728	ML(N)_boost_(3)_with(out)
7	0,420039	0,639322	ML(F)_elastic_(1)_with()
8	0,420085	0,639411	ML(N)_elastic_(1)_with()
9	0,426621	0,639718	ML(N)_lasso_(3)_with(out)
10	0,426621	0,639718	ML(F)_lasso_(3)_with(out)
11	0,426621	0,639718	ML(N)_lasso_(6)_with(out)
12	0,426621	0,639718	ML(F)_lasso_(6)_with(out)
13	0,428976	0,641126	ML(N)_lasso_(1)_with()
14	0,428976	0,641126	ML(F)_lasso_(1)_with()
15	0,428976	0,641126	ML(N)_lasso_(1)_with(out)
16	0,428976	0,641126	ML(F)_lasso_(1)_with(out)
17	0,428976	0,641126	ML(N)_lasso_(3)_with()
18	0,428976	0,641126	ML(F)_lasso_(3)_with()
19	0,428976	0,641126	ML(N)_lasso_(6)_with()
20	0,428976	0,641126	ML(F)_lasso_(6)_with()
21	0,439671	0,644019	ML(N)_forest_(1)_with()
22	0,418863	0,644242	ML(F)_boost_(6)_with()
23	0,429613	0,644347	ML(N)_ridge_(3)_with()
24	0,429613	0,644347	ML(F)_ridge_(3)_with()
25	0,429613	0,644347	ML(N)_ridge_(6)_with()
26	0,429613	0,644347	ML(F)_ridge_(6)_with()
27	0,435527	0,648329	ML(N)_ridge_(1)_with()
28	0,435527	0,648329	ML(F)_ridge_(1)_with()
29	0,436049	0,648408	ML(N)_ridge_(3)_with(out)
30	0,436011	0,648457	ML(N)_ridge_(6)_with(out)
31	0,436273	0,648616	ML(F)_ridge_(3)_with(out)
32	0,436273	0,648616	ML(F)_ridge_(6)_with(out)
33	0,480644	0,651023	ML(F)_boost_(1)_with(out)
34	0,440633	0,651270	ML(N)_forest_(3)_with()
35	0,434216	0,651416	ARF(1) with
36	0,434216	0,651416	ARN(1) with
37	0,420464	0,651749	ML(N)_boost_(6)_with()
38	0,459926	0,652350	ML(F)_forest_(6)_with(out)
39	0,462580	0,652800	ML(F)_forest_(3)_with(out)
40	0,451997	0,655781	ML(N)_ridge_(1)_with(out)
41	0,451365	0,656113	ML(F)_ridge_(1)_with(out)
42	0,460042	0,656226	ML(F)_boost_(1)_with()
43	0,449159	0,657163	ML(F)_forest_(3)_with()
44	0,479795	0,657950	ML(F)_forest_(1)_with()
45	0,442607	0,658044	ML(F)_forest_(6)_with()
46	0,454777	0,658524	ARF(1) without
47	0,454777	0,658524	ARN(1) without
48	0,425828	0,658632	ML(N)_boost_(3)_with()
49	0,460643	0,661890	ML(N)_forest_(1)_with(out)
50	0,467268	0,663220	ML(F)_boost_(3)_with(out)
51	0,430426	0,665856	ML(N)_boost_(1)_with()

Sentiment Index

	MAE	RMSE	model
1	0,4121630	0,6352807	ML(F)_boost_(6)_with(out)
2	0,4312737	0,6391794	ML(N)_forest_(6)_with(out)
3	0,4266212	0,6397183	ML(N)_lasso_(3)_with(out)
4	0,4266212	0,6397183	ML(F)_lasso_(3)_with(out)
5	0,4266212	0,6397183	ML(N)_lasso_(6)_with(out)
6	0,4266212	0,6397183	ML(F)_lasso_(6)_with(out)
7	0,4206703	0,6410272	ML(N)_elastic_(1)_with()
8	0,4207063	0,6410765	ML(F)_elastic_(1)_with()
9	0,4289758	0,6411265	ML(N)_lasso_(1)_with()
10	0,4289758	0,6411265	ML(F)_lasso_(1)_with()
11	0,4289758	0,6411265	ML(N)_lasso_(1)_with(out)
12	0,4289758	0,6411265	ML(F)_lasso_(1)_with(out)
13	0,4289758	0,6411265	ML(N)_lasso_(3)_with()
14	0,4289758	0,6411265	ML(F)_lasso_(3)_with()
15	0,4289758	0,6411265	ML(N)_lasso_(6)_with()
16	0,4289758	0,6411265	ML(F)_lasso_(6)_with()
17	0,4183519	0,6412840	ML(F)_boost_(6)_with()
18	0,4296135	0,6443474	ML(N)_ridge_(3)_with()
19	0,4296135	0,6443474	ML(F)_ridge_(3)_with()
20	0,4296135	0,6443474	ML(N)_ridge_(6)_with()
21	0,4296135	0,6443474	ML(F)_ridge_(6)_with()
22	0,4171365	0,6457150	ML(F)_boost_(3)_with(out)
23	0,4254619	0,6460223	ML(N)_boost_(3)_with(out)
24	0,4298950	0,6481575	ARXF(1) with
25	0,4355270	0,6483290	ML(N)_ridge_(1)_with()
26	0,4355270	0,6483290	ML(F)_ridge_(1)_with()
27	0,4360487	0,6484079	ML(F)_ridge_(3)_with(out)
28	0,4360112	0,6484567	ML(F)_ridge_(6)_with(out)
29	0,4364831	0,6493329	ML(N)_ridge_(3)_with(out)
30	0,4364594	0,6493554	ML(N)_ridge_(6)_with(out)
31	0,4251952	0,6495321	ML(N)_forest_(3)_with(out)
32	0,4342156	0,6514161	ARF(1) with
33	0,4342156	0,6514161	ARN(1) with
34	0,4238117	0,6527138	ML(N)_boost_(6)_with(out)
35	0,4271821	0,6543788	ML(N)_boost_(3)_with()
36	0,4276900	0,6556863	ML(F)_boost_(3)_with()
37	0,4519967	0,6557805	ML(F)_ridge_(1)_with(out)
38	0,4525089	0,6570268	ML(N)_ridge_(1)_with(out)
39	0,4547766	0,6585241	ARF(1) without
40	0,4547766	0,6585241	ARN(1) without
41	0,4528835	0,6603872	ML(N)_forest_(1)_with()
42	0,4361189	0,6651309	ML(N)_boost_(6)_with()
43	0,4510229	0,6733655	ML(N)_forest_(6)_with()
44	0,4168976	0,6741430	ARF(3) without
45	0,4168976	0,6741430	ARN(3) without
46	0,4431948	0,6786021	ML(N)_forest_(3)_with()
47	0,4243836	0,6821399	ML(F)_elastic_(3)_with()
48	0,4258098	0,6825982	ML(N)_elastic_(3)_with()
49	0,4540476	0,6875064	ML(N)_boost_(1)_with()
50	0,4299793	0,6916137	ARF(3) with
51	0,4299793	0,6916137	ARN(3) with

52	0,473940	0,695667	ML(F)_boost_(1)_with()	**52**	0,5117416	0,6795394	ML(F)_forest_(6)_with(out)	**52**	0,428222	0,667046	ML(N)_boost_(1)_with(out)	**52**	0,4300352	0,6972369	ARF(6) without
53	0,430035	0,697237	ARF(6) without	**53**	0,4242136	0,6810351	ML(F)_elastic_(3)_with()	**53**	0,500690	0,671164	ML(F)_linear_(1)_with()	**53**	0,4300352	0,6972369	ARN(6) without
54	0,430035	0,697237	ARN(6) without	**54**	0,4242288	0,6811942	ML(N)_elastic_(3)_with()	**54**	0,443369	0,672952	ARXF(6) with	**54**	0,4523484	0,6981316	ML(F)_boost_(1)_with(out)
55	0,471855	0,698241	ML(F)_forest_(3)_with(out)	**55**	0,4617102	0,6822067	ML(N)_forest_(1)_with(out)	**55**	0,416898	0,674143	ARF(3) without	**55**	0,4728498	0,6984056	ML(N)_linear_(1)_with()
56	0,444862	0,702930	ML(F)_elastic_(6)_with()	**56**	0,4299793	0,6916137	ARF(3) with	**56**	0,416898	0,674143	ARN(3) without	**56**	0,4448910	0,7039327	ML(F)_elastic_(6)_with()
57	0,444836	0,702965	ML(N)_elastic_(6)_with()	**57**	0,4299793	0,6916137	ARN(3) with	**57**	0,511504	0,675510	ML(F)_forest_(1)_with(out)	**57**	0,4466002	0,7042158	ML(N)_elastic_(6)_with()
58	0,472411	0,705253	ML(F)_forest_(6)_with(out)	**58**	0,5168510	0,6933958	ML(F)_forest_(1)_with(out)	**58**	0,464214	0,678694	ML(N)_forest_(6)_with(out)	**58**	0,4386536	0,7080858	ARXF(3) with
59	0,496213	0,709522	ML(N)_forest_(1)_with(out)	**59**	0,4300352	0,6972369	ARF(6) without	**59**	0,424201	0,681024	ML(F)_elastic_(3)_with()	**59**	0,4690431	0,7109220	ML(F)_forest_(6)_with(out)
60	0,510014	0,709598	ML(N)_linear_(1)_with()	**60**	0,4300352	0,6972369	ARN(6) without	**60**	0,424151	0,681094	ML(N)_elastic_(3)_with()	**60**	0,4484193	0,7131703	ARF(6) with
61	0,456531	0,710106	ML(F)_forest_(3)_with()	**61**	0,4793928	0,6982061	ML(F)_forest_(3)_with()	**61**	0,458444	0,681519	ARXF(3) with	**61**	0,4484193	0,7131703	ARN(6) with
62	0,448419	0,713170	ARF(6) with	**62**	0,4923485	0,7004591	ML(F)_forest_(6)_with()	**62**	0,429979	0,691614	ARF(3) with	**62**	0,4812269	0,7210908	ML(N)_forest_(1)_with(out)
63	0,448419	0,713170	ARN(6) with	**63**	0,4448646	0,7028852	ML(F)_elastic_(6)_with()	**63**	0,429979	0,691614	ARN(3) with	**63**	0,4825236	0,7228980	ML(F)_linear_(1)_with()
64	0,495223	0,713478	ML(F)_boost_(1)_with(out)	**64**	0,4449159	0,7030796	ML(N)_elastic_(6)_with()	**64**	0,454747	0,694040	ML(N)_linear_(1)_with()	**64**	0,4549508	0,7245360	ML(F)_forest_(6)_with()
65	0,467718	0,719320	ARXF(3) with	**65**	0,5009219	0,7074232	ML(F)_linear_(1)_with()	**65**	0,430035	0,697237	ARF(6) without	**65**	0,4561530	0,7258171	ML(F)_boost_(1)_with()
66	0,498590	0,724048	ML(F)_linear_(1)_with()	**66**	0,4484193	0,7131703	ARF(6) with	**66**	0,430035	0,697237	ARN(6) without	**66**	0,4641214	0,7293116	ML(F)_linear_(3)_with()
67	0,529347	0,727381	ARXF(1) with	**67**	0,4484193	0,7131703	ARN(6) with	**67**	0,451102	0,700597	ML(N)_elastic_(1)_with(out)	**67**	0,4656926	0,7302460	ML(N)_linear_(3)_with()
68	0,494589	0,729877	ML(F)_linear_(3)_with()	**68**	0,5543255	0,7296641	ML(F)_elastic_(1)_with(out)	**68**	0,451538	0,701523	ML(N)_linear_(1)_with(out)	**68**	0,4497287	0,7322701	ML(F)_forest_(3)_with()
69	0,506511	0,732710	ML(F)_elastic_(3)_with(out)	**69**	0,5587813	0,7353317	ML(F)_linear_(1)_with(out)	**69**	0,503103	0,702676	ML(F)_linear_(3)_with()	**69**	0,4816977	0,7362258	ML(F)_forest_(3)_with(out)
70	0,510174	0,736934	ML(F)_linear_(3)_with(out)	**70**	0,5013147	0,7369357	ML(F)_linear_(6)_with()	**70**	0,444885	0,702898	ML(F)_elastic_(6)_with()	**70**	0,4785982	0,7408451	ML(F)_elastic_(3)_with(out)
71	0,526646	0,739757	ML(F)_elastic_(6)_with(out)	**71**	0,5137362	0,7387956	ML(F)_linear_(3)_with()	**71**	0,444840	0,702991	ML(N)_elastic_(6)_with()	**71**	0,4802512	0,7416181	ML(N)_boost_(1)_with(out)
72	0,505218	0,742648	ML(F)_forest_(6)_with()	**72**	0,4873818	0,7413726	ML(N)_linear_(1)_with()	**72**	0,448419	0,713170	ARF(6) with	**72**	0,4808171	0,7427842	ML(F)_linear_(3)_with(out)
73	0,489955	0,742933	ARXF(6) with	**73**	0,5404665	0,7435484	ARXF(1) with	**73**	0,448419	0,713170	ARN(6) with	**73**	0,4868180	0,7458306	ML(N)_linear_(6)_with()
74	0,531745	0,744887	ML(F)_linear_(6)_with(out)	**74**	0,5711673	0,7501260	ML(F)_elastic_(6)_with(out)	**74**	0,514788	0,717595	ML(F)_linear_(6)_with()	**74**	0,4849998	0,7465259	ML(F)_elastic_(6)_with(out)
75	0,503021	0,746112	ML(F)_linear_(6)_with()	**75**	0,5747577	0,7556997	ARXF(1) without	**75**	0,491526	0,721994	ML(N)_forest_(6)_with()	**75**	0,4636203	0,7469687	ARXF(6) with
76	0,516056	0,749629	ML(F)_elastic_(1)_with(out)	**76**	0,5780019	0,7571499	ML(F)_linear_(6)_with(out)	**76**	0,449686	0,724516	ARXN(6) with	**76**	0,5164080	0,7506321	ML(N)_elastic_(1)_with(out)
77	0,517132	0,753306	ML(F)_forest_(1)_with()	**77**	0,5770296	0,7637942	ARXF(6) without	**77**	0,483836	0,735857	ARXN(3) with	**77**	0,4955724	0,7552106	ML(F)_linear_(6)_with(out)
78	0,520209	0,754806	ML(F)_linear_(1)_with(out)	**78**	0,5911223	0,7651704	ML(F)_elastic_(3)_with(out)	**78**	0,574081	0,737228	ML(F)_elastic_(1)_with(out)	**78**	0,5207235	0,7553559	ML(N)_linear_(1)_with(out)
79	0,541120	0,755297	ML(N)_elastic_(1)_with(out)	**79**	0,5540560	0,7664581	ARXF(6) with	**79**	0,466797	0,739502	ML(N)_elastic_(3)_with(out)	**79**	0,4834067	0,7584576	ML(F)_linear_(6)_with()
80	0,544587	0,759169	ML(N)_linear_(1)_with(out)	**80**	0,5987392	0,7727767	ML(F)_linear_(3)_with(out)	**80**	0,580560	0,744133	ML(F)_linear_(1)_with(out)	**80**	0,4902022	0,7654611	ML(N)_elastic_(3)_with(out)
81	0,532675	0,762683	ARXF(3) without	**81**	0,5996935	0,7750829	ARXF(3) without	**81**	0,470345	0,744721	ML(N)_linear_(3)_with(out)	**81**	0,4973345	0,7666826	ML(F)_forest_(1)_with()
82	0,536891	0,771907	ARXF(6) without	**82**	0,5028604	0,7782968	ML(N)_elastic_(1)_with(out)	**82**	0,487734	0,745077	ML(N)_linear_(3)_with()	**82**	0,5044996	0,7682161	ARXF(3) without
83	0,526639	0,779216	ML(N)_linear_(3)_with()	**83**	0,5051077	0,7822601	ML(N)_linear_(1)_with(out)	**83**	0,562163	0,745332	ML(F)_elastic_(6)_with(out)	**83**	0,5144301	0,7697836	ML(N)_elastic_(6)_with(out)
84	0,540387	0,782393	ML(F)_forest_(1)_with(out)	**84**	0,5906329	0,7836862	ARXN(1) with	**84**	0,579669	0,756861	ML(F)_linear_(6)_with(out)	**84**	0,4940704	0,7704399	ML(N)_linear_(3)_with(out)
85	0,576478	0,792424	ARXN(1) without	**85**	0,5321625	0,7860700	ARXN(3) with	**85**	0,581198	0,761479	ML(F)_elastic_(3)_with(out)	**85**	0,5209245	0,7761109	ML(N)_linear_(6)_with(out)
86	0,563256	0,797531	ARXF(1) without	**86**	0,5520875	0,8207894	ARXN(6) with	**86**	0,489231	0,762077	ARXN(3) without	**86**	0,5051981	0,7804642	ML(F)_elastic_(1)_with(out)
87	0,532023	0,798615	ML(N)_linear_(6)_with()	**87**	0,5586316	0,8221608	ARXN(1) without	**87**	0,485369	0,769183	ML(N)_elastic_(6)_with(out)	**87**	0,5248398	0,7837537	ARXN(3) without
88	0,546448	0,808310	ARXN(6) with	**88**	0,5316134	0,8238451	ML(N)_linear_(3)_with()	**88**	0,609004	0,769232	ARXF(1) without	**88**	0,5075538	0,7843113	ML(F)_linear_(1)_with(out)
89	0,585805	0,830813	ARXN(3) with	**89**	0,5442560	0,8253489	ML(N)_linear_(6)_with()	**89**	0,588785	0,769712	ML(F)_linear_(3)_with(out)	**89**	0,5463065	0,7911624	ARXN(3) with
90	0,577595	0,833841	ML(N)_elastic_(3)_with(out)	**90**	0,6297902	0,8264200	ARXF(3) with	**90**	0,503280	0,773460	ML(N)_linear_(6)_with()	**90**	0,5323999	0,7915495	ARXF(6) without
91	0,602864	0,838976	ARXN(3) without	**91**	0,5643644	0,8484350	ML(N)_elastic_(6)_with(out)	**91**	0,489170	0,776805	ML(N)_linear_(6)_with(out)	**91**	0,5555654	0,8042903	ML(F)_forest_(1)_with(out)
92	0,585901	0,841812	ML(N)_linear_(3)_with(out)	**92**	0,5601789	0,8550748	ML(N)_elastic_(3)_with(out)	**92**	0,539497	0,777233	ARXN(1) without	**92**	0,6255703	0,8184053	ARXN(1) with
93	0,577245	0,847609	ML(N)_elastic_(6)_with(out)	**93**	0,5706314	0,8583912	ML(N)_linear_(6)_with(out)	**93**	0,598406	0,784635	ARXF(3) without	**93**	0,6088911	0,8192401	ARXN(1) without
94	0,619378	0,855971	ARXN(6) without	**94**	0,5688344	0,8683992	ML(N)_linear_(3)_with(out)	**94**	0,491355	0,784978	ARXN(6) without	**94**	0,5827710	0,8253804	ARXF(1) without
95	0,587800	0,857573	ML(N)_linear_(6)_with(out)	**95**	0,5863600	0,8813325	ARXN(3) without	**95**	0,597394	0,787436	ARXN(1) with	**95**	0,6337458	0,8714645	ARXN(6) with
96	0,741916	0,914032	ARXN(1) with	**96**	0,5873234	0,8846536	ARXN(6) without	**96**	0,607858	0,788278	ARXF(6) without	**96**	0,6458582	0,8879052	ARXN(6) without

Appendix H6. NX models

Frequency Index [Basic]

	MAE	RMSE	model
1	1,085830	1,327868	ML(F)_forest_(3)_with()
2	1,070187	1,331696	ML(F)_ridge_(3)_with(out)
3	1,070620	1,334002	ML(F)_ridge_(3)_with()
4	1,070620	1,334002	ML(N)_ridge_(3)_with()
5	1,074335	1,334914	ML(N)_ridge_(3)_with(out)
6	1,082651	1,335253	ML(N)_lasso_(1)_with(out)
7	1,072756	1,337415	ML(N)_ridge_(6)_with(out)
8	1,069756	1,339763	ML(N)_ridge_(6)_with()
9	1,078244	1,344338	ML(N)_lasso_(3)_with()
10	1,078244	1,344338	ML(F)_lasso_(3)_with()
11	1,078244	1,344338	ML(N)_lasso_(3)_with(out)
12	1,078244	1,344338	ML(F)_lasso_(3)_with(out)
13	1,078244	1,344338	ML(N)_lasso_(6)_with()
14	1,078244	1,344338	ML(F)_lasso_(6)_with()
15	1,078244	1,344338	ML(N)_lasso_(6)_with(out)
16	1,078244	1,344338	ML(F)_lasso_(6)_with(out)
17	1,077053	1,344564	ML(N)_ridge_(1)_with(out)
18	1,089204	1,344866	ML(N)_lasso_(1)_with()
19	1,123317	1,345729	ML(F)_forest_(6)_with(out)
20	1,084054	1,351401	ML(N)_ridge_(1)_with()
21	1,085141	1,351917	ML(F)_ridge_(1)_with(out)
22	1,134821	1,352868	ML(F)_forest_(1)_with()
23	1,099912	1,354918	ML(F)_ridge_(1)_with()
24	1,090229	1,355007	ML(F)_ridge_(6)_with(out)
25	1,082148	1,356481	ML(F)_elastic_(3)_with()
26	1,082214	1,356567	ML(N)_elastic_(3)_with()
27	1,089663	1,357243	ML(F)_boost_(1)_with()
28	1,097954	1,357832	ML(F)_forest_(6)_with()
29	1,094524	1,358474	ML(F)_lasso_(1)_with()
30	1,084704	1,359827	ML(F)_boost_(1)_with(out)
31	1,088381	1,361075	ML(F)_ridge_(6)_with()
32	1,088751	1,362812	ARF(3) without
33	1,088751	1,362812	ARN(3) without
34	1,081052	1,363072	ML(F)_boost_(3)_with(out)
35	1,142391	1,363109	ML(F)_forest_(3)_with(out)
36	1,094611	1,368039	ARF(3) with
37	1,094611	1,368039	ARN(3) with
38	1,096762	1,369455	ML(F)_boost_(3)_with()
39	1,105439	1,378936	ML(N)_elastic_(1)_with()
40	1,150078	1,379816	ML(F)_forest_(1)_with(out)
41	1,127952	1,381774	ML(N)_boost_(3)_with(out)
42	1,125229	1,382106	ML(N)_boost_(6)_with(out)
43	1,103552	1,383054	ML(F)_lasso_(1)_with(out)
44	1,111844	1,386526	ML(F)_elastic_(1)_with()
45	1,124966	1,388603	ML(F)_boost_(6)_with()
46	1,128701	1,389588	ML(F)_linear_(3)_with()
47	1,113627	1,390247	ARF(1) with
48	1,113627	1,390247	ARN(1) with
49	1,116177	1,391076	ML(N)_elastic_(3)_with(out)
50	1,135096	1,392478	ML(F)_boost_(6)_with(out)
51	1,117347	1,392613	ML(N)_linear_(3)_with(out)

Frequency Index [Full]

	MAE	RMSE	model
1	1,075282	1,332195	ML(F)_ridge_(3)_with(out)
2	1,070620	1,334002	ML(F)_ridge_(3)_with()
3	1,060007	1,334712	ML(N)_ridge_(6)_with(out)
4	1,074212	1,335091	ML(F)_ridge_(6)_with(out)
5	1,080039	1,337086	ML(N)_lasso_(1)_with(out)
6	1,069756	1,339763	ML(F)_ridge_(6)_with()
7	1,069756	1,339763	ML(N)_ridge_(6)_with()
8	1,085879	1,341725	ML(N)_lasso_(1)_with()
9	1,078244	1,344338	ML(F)_lasso_(1)_with()
10	1,078244	1,344338	ML(F)_lasso_(1)_with(out)
11	1,078244	1,344338	ML(N)_lasso_(3)_with()
12	1,078244	1,344338	ML(F)_lasso_(3)_with()
13	1,078244	1,344338	ML(N)_lasso_(3)_with(out)
14	1,078244	1,344338	ML(F)_lasso_(3)_with(out)
15	1,078244	1,344338	ML(N)_lasso_(6)_with()
16	1,078244	1,344338	ML(F)_lasso_(6)_with()
17	1,078244	1,344338	ML(N)_lasso_(6)_with(out)
18	1,078244	1,344338	ML(F)_lasso_(6)_with(out)
19	1,077053	1,344564	ML(N)_ridge_(1)_with(out)
20	1,077053	1,344564	ML(F)_ridge_(1)_with(out)
21	1,077053	1,344564	ML(N)_ridge_(3)_with(out)
22	1,090436	1,349552	ML(N)_ridge_(3)_with()
23	1,084054	1,351401	ML(N)_ridge_(1)_with()
24	1,084054	1,351401	ML(F)_ridge_(1)_with()
25	1,082059	1,356395	ML(F)_elastic_(3)_with()
26	1,082135	1,356491	ML(N)_elastic_(3)_with()
27	1,088751	1,362812	ARF(3) without
28	1,088751	1,362812	ARN(3) without
29	1,094611	1,368039	ARF(3) with
30	1,094611	1,368039	ARN(3) with
31	1,108295	1,369447	ARXF(1) with
32	1,080117	1,369820	ML(N)_forest_(3)_with(out)
33	1,105309	1,378719	ML(F)_elastic_(1)_with()
34	1,142872	1,385776	ML(N)_boost_(1)_with()
35	1,111844	1,386549	ML(N)_elastic_(1)_with()
36	1,113627	1,390247	ARF(1) with
37	1,113627	1,390247	ARN(1) with
38	1,137517	1,392489	ML(F)_boost_(3)_with(out)
39	1,136030	1,392879	ML(F)_boost_(6)_with(out)
40	1,131663	1,394018	ML(N)_forest_(6)_with(out)
41	1,117680	1,396834	ARF(1) without
42	1,117680	1,396834	ARN(1) without
43	1,139451	1,400566	ML(N)_forest_(3)_with()
44	1,139769	1,400703	ML(F)_boost_(1)_with(out)
45	1,157037	1,409865	ML(N)_boost_(1)_with(out)
46	1,152608	1,414008	ML(F)_elastic_(6)_with()
47	1,152754	1,414109	ML(N)_elastic_(6)_with()
48	1,134544	1,415435	ML(F)_boost_(3)_with()
49	1,133282	1,417732	ML(F)_boost_(6)_with()
50	1,159673	1,423739	ARF(6) with
51	1,159673	1,423739	ARN(6) with

Frequency Index [Optimal]

	MAE	RMSE	model
1	1,0706204	1,3340022	ML(F)_ridge_(3)_with()
2	1,0805504	1,3391428	ML(F)_ridge_(3)_with(out)
3	1,0697564	1,3397633	ML(F)_ridge_(6)_with()
4	1,0793003	1,3414621	ML(F)_ridge_(6)_with(out)
5	1,0915491	1,3420009	ML(F)_boost_(3)_with()
6	1,0876910	1,3442026	ML(F)_ridge_(1)_with()
7	1,0782440	1,3443378	ML(N)_lasso_(1)_with(out)
8	1,0782440	1,3443378	ML(F)_lasso_(1)_with(out)
9	1,0782440	1,3443378	ML(N)_lasso_(3)_with(out)
10	1,0782440	1,3443378	ML(F)_lasso_(3)_with(out)
11	1,0782440	1,3443378	ML(N)_lasso_(6)_with(out)
12	1,0782440	1,3443378	ML(F)_lasso_(6)_with(out)
13	1,0770525	1,3445636	ML(N)_ridge_(1)_with(out)
14	1,0770525	1,3445636	ML(N)_ridge_(3)_with(out)
15	1,0770525	1,3445636	ML(N)_ridge_(6)_with(out)
16	1,0772823	1,3446283	ML(F)_lasso_(3)_with()
17	1,0772823	1,3446283	ML(F)_lasso_(6)_with()
18	1,0887716	1,3448072	ML(F)_ridge_(1)_with(out)
19	1,0780193	1,3454595	ML(N)_lasso_(1)_with()
20	1,0781960	1,3454859	ML(F)_lasso_(1)_with()
21	1,0785475	1,3456262	ML(N)_lasso_(3)_with()
22	1,0785475	1,3456262	ML(N)_lasso_(6)_with()
23	1,1014302	1,3481484	ML(F)_boost_(3)_with(out)
24	1,0876145	1,3511611	ML(N)_ridge_(3)_with()
25	1,0876145	1,3511611	ML(N)_ridge_(6)_with()
26	1,0845062	1,3514316	ML(N)_ridge_(1)_with()
27	1,1079970	1,3563623	ML(F)_boost_(1)_with()
28	1,0821115	1,3564354	ML(F)_elastic_(3)_with()
29	1,0822139	1,3565777	ML(N)_elastic_(3)_with()
30	1,1011033	1,3572515	ML(F)_boost_(6)_with()
31	1,1154847	1,3606373	ML(F)_boost_(6)_with(out)
32	1,1183630	1,3622829	ML(F)_boost_(1)_with(out)
33	1,0887513	1,3628118	ARF(3) without
34	1,0887513	1,3628118	ARN(3) without
35	1,0946113	1,3680387	ARF(3) with
36	1,0946113	1,3680387	ARN(3) with
37	1,1026844	1,3697605	ML(N)_boost_(6)_with(out)
38	1,1053257	1,3787728	ML(F)_elastic_(1)_with()
39	1,1054322	1,3789440	ML(N)_elastic_(1)_with()
40	1,1136270	1,3902469	ARF(1) with
41	1,1136270	1,3902469	ARN(1) with
42	1,1364992	1,3963732	ML(N)_boost_(3)_with()
43	1,1176804	1,3968341	ARF(1) without
44	1,1176804	1,3968341	ARN(1) without
45	1,1301241	1,3986112	ML(N)_boost_(1)_with()
46	1,1306396	1,3990285	ML(N)_boost_(3)_with(out)
47	1,1239331	1,4006688	ML(N)_boost_(1)_with(out)
48	1,0868078	1,4044763	ML(N)_forest_(3)_with(out)
49	1,1360164	1,4045080	ML(N)_boost_(6)_with()
50	1,1527677	1,4141710	ML(N)_elastic_(6)_with()
51	1,1625179	1,4234902	ML(F)_elastic_(6)_with()

Sentiment Index

	MAE	RMSE	model
1	1,0729933	1,3108964	ML(N)_forest_(1)_with()
2	1,0471170	1,3160604	ARXN(3) with
3	1,0548208	1,3196000	ML(N)_ridge_(3)_with(out)
4	1,0542433	1,3222341	ML(N)_ridge_(6)_with(out)
5	1,0662981	1,3251335	ML(F)_ridge_(3)_with(out)
6	1,0646152	1,3275710	ML(F)_ridge_(6)_with(out)
7	1,0745214	1,3288785	ARXN(1) with
8	1,0408815	1,3289607	ML(N)_forest_(1)_with(out)
9	1,0829367	1,3293092	ML(N)_boost_(3)_with(out)
10	1,0552668	1,3302081	ML(N)_forest_(3)_with(out)
11	1,0706132	1,3339945	ML(N)_ridge_(3)_with()
12	1,0706202	1,3339977	ML(F)_ridge_(3)_with()
13	1,0819653	1,3347281	ML(F)_lasso_(1)_with(out)
14	1,0378017	1,3376492	ML(N)_forest_(6)_with(out)
15	1,0524172	1,3393047	ML(N)_forest_(6)_with()
16	1,0697496	1,3397559	ML(N)_ridge_(6)_with()
17	1,0697561	1,3397590	ML(F)_ridge_(6)_with()
18	1,0883312	1,3436393	ML(F)_lasso_(1)_with()
19	1,0782440	1,3443378	ML(N)_lasso_(1)_with()
20	1,0782440	1,3443378	ML(N)_lasso_(1)_with(out)
21	1,0782440	1,3443378	ML(N)_lasso_(3)_with()
22	1,0782440	1,3443378	ML(F)_lasso_(3)_with()
23	1,0782440	1,3443378	ML(N)_lasso_(3)_with(out)
24	1,0782440	1,3443378	ML(F)_lasso_(3)_with(out)
25	1,0782440	1,3443378	ML(N)_lasso_(6)_with()
26	1,0782440	1,3443378	ML(F)_lasso_(6)_with()
27	1,0782440	1,3443378	ML(N)_lasso_(6)_with(out)
28	1,0782440	1,3443378	ML(F)_lasso_(6)_with(out)
29	1,0770525	1,3445636	ML(N)_ridge_(1)_with(out)
30	1,0770525	1,3445636	ML(F)_ridge_(1)_with(out)
31	1,0695701	1,3464928	ML(N)_boost_(3)_with()
32	1,0840539	1,3514006	ML(N)_ridge_(1)_with()
33	1,0840539	1,3514006	ML(F)_ridge_(1)_with()
34	1,1047411	1,3514566	ARXN(6) with
35	1,0783859	1,3520796	ML(N)_elastic_(3)_with()
36	1,0818418	1,3551079	ML(F)_elastic_(3)_with()
37	1,1019368	1,3562192	ML(N)_boost_(1)_with()
38	1,1055162	1,3568371	ML(N)_boost_(6)_with(out)
39	1,0615290	1,3583775	ML(N)_elastic_(3)_with(out)
40	1,0887513	1,3628118	ARF(3) without
41	1,0887513	1,3628118	ARN(3) without
42	1,0669465	1,3628974	ML(N)_linear_(3)_with(out)
43	1,0946113	1,3680387	ARF(3) with
44	1,0946113	1,3680387	ARN(3) with
45	1,1017220	1,3714686	ML(N)_boost_(6)_with()
46	1,1054524	1,3730752	ML(F)_boost_(3)_with()
47	1,1106816	1,3732481	ML(N)_boost_(1)_with(out)
48	1,1003873	1,3743971	ML(N)_elastic_(1)_with()
49	1,1295559	1,3779992	ML(F)_boost_(3)_with(out)
50	1,1033849	1,3787264	ARXF(1) with
51	1,1331282	1,3834007	ML(F)_boost_(1)_with(out)

52	1,117680	1,396834	ARF(1) without
53	1,117680	1,396834	ARN(1) without
54	1,137719	1,400741	ML(N)_boost_(3)_with()
55	1,116162	1,402607	ML(N)_linear_(3)_with()
56	1,144248	1,406940	ML(N)_boost_(6)_with()
57	1,148920	1,407908	ARXN(3) without
58	1,152705	1,414099	ML(F)_elastic_(6)_with()
59	1,152774	1,414172	ML(N)_elastic_(6)_with()
60	1,153692	1,417457	ML(N)_boost_(1)_with()
61	1,159673	1,423739	ARF(6) with
62	1,159673	1,423739	ARN(6) with
63	1,139056	1,425703	ML(F)_linear_(1)_with()
64	1,161269	1,427084	ML(F)_elastic_(3)_with(out)
65	1,164271	1,430909	ML(F)_linear_(3)_with(out)
66	1,186397	1,433247	ARF(6) without
67	1,186397	1,433247	ARN(6) without
68	1,141383	1,434602	ML(N)_linear_(1)_with()
69	1,184856	1,435084	ML(N)_boost_(1)_with(out)
70	1,174095	1,440030	ML(N)_elastic_(1)_with(out)
71	1,174914	1,442021	ML(N)_linear_(1)_with(out)
72	1,192896	1,445654	ARXN(1) without
73	1,172527	1,446945	ARXF(3) without
74	1,198274	1,451606	ARXN(6) with
75	1,182650	1,465822	ML(F)_elastic_(1)_with(out)
76	1,184318	1,469854	ML(N)_forest_(3)_with()
77	1,186196	1,470113	ML(F)_linear_(1)_with(out)
78	1,202013	1,474269	ARXN(1) with
79	1,245000	1,478001	ML(N)_elastic_(6)_with(out)
80	1,229880	1,480044	ML(N)_linear_(6)_with()
81	1,214050	1,480107	ARXN(3) with
82	1,193428	1,480162	ARXF(1) without
83	1,255187	1,486377	ML(N)_linear_(6)_with(out)
84	1,248578	1,494193	ARXN(6) without
85	1,243234	1,496777	ML(F)_linear_(6)_with()
86	1,237804	1,504243	ARXF(3) with
87	1,208105	1,510415	ML(N)_forest_(3)_with(out)
88	1,172030	1,518318	ML(N)_forest_(6)_with()
89	1,236328	1,534136	ML(N)_forest_(1)_with()
90	1,254459	1,565244	ARXF(1) with
91	1,282073	1,578861	ML(N)_forest_(6)_with(out)
92	1,337183	1,587539	ML(F)_elastic_(6)_with(out)
93	1,304253	1,594349	ARXF(6) with
94	1,348428	1,601708	ML(F)_linear_(6)_with(out)
95	1,318713	1,602097	ML(N)_forest_(1)_with(out)
96	1,357200	1,633264	ARXF(6) without

52	1,130999	1,427416	ML(N)_linear_(3)_with()
53	1,132623	1,431493	ML(N)_forest_(1)_with(out)
54	1,186397	1,433247	ARF(6) without
55	1,186397	1,433247	ARN(6) without
56	1,182221	1,437334	ARXN(3) with
57	1,206490	1,446048	ML(F)_forest_(3)_with(out)
58	1,144883	1,446957	ML(N)_linear_(1)_with()
59	1,213467	1,451172	ML(N)_boost_(3)_with()
60	1,151628	1,454092	ML(N)_forest_(6)_with()
61	1,179579	1,456388	ML(F)_boost_(1)_with()
62	1,192925	1,456631	ML(N)_forest_(1)_with()
63	1,167326	1,457147	ML(N)_boost_(6)_with(out)
64	1,197742	1,464014	ARXN(3) without
65	1,185090	1,464703	ML(F)_forest_(6)_with(out)
66	1,198386	1,465096	ML(N)_elastic_(3)_with(out)
67	1,178100	1,467183	ML(N)_boost_(3)_with(out)
68	1,203825	1,469021	ML(N)_boost_(6)_with()
69	1,203033	1,469296	ML(N)_linear_(3)_with(out)
70	1,220455	1,472585	ML(F)_linear_(1)_with()
71	1,191030	1,474855	ARXN(1) without
72	1,197819	1,480645	ML(N)_elastic_(1)_with(out)
73	1,213908	1,481745	ML(F)_linear_(3)_with()
74	1,202345	1,484988	ML(N)_linear_(1)_with(out)
75	1,206410	1,488181	ARXN(1) with
76	1,190862	1,496656	ML(F)_forest_(1)_with(out)
77	1,231895	1,501403	ML(N)_linear_(6)_with()
78	1,194058	1,502843	ML(F)_forest_(6)_with()
79	1,218938	1,518289	ARXF(3) with
80	1,259187	1,534251	ARXN(6) with
81	1,292238	1,549715	ML(F)_forest_(3)_with()
82	1,272058	1,550982	ML(F)_linear_(6)_with()
83	1,227910	1,551088	ML(F)_forest_(1)_with()
84	1,273374	1,558234	ARXF(6) with
85	1,310963	1,596706	ML(F)_elastic_(1)_with(out)
86	1,324969	1,597973	ML(N)_elastic_(6)_with(out)
87	1,307321	1,598674	ARXF(1) without
88	1,336339	1,609246	ML(N)_linear_(6)_with(out)
89	1,321529	1,609699	ML(F)_linear_(1)_with(out)
90	1,327833	1,635624	ARXF(3) without
91	1,326244	1,636779	ML(F)_elastic_(3)_with(out)
92	1,338025	1,652943	ML(F)_linear_(3)_with(out)
93	1,369524	1,667520	ML(F)_elastic_(6)_with(out)
94	1,386305	1,686840	ML(F)_linear_(6)_with(out)
95	1,401928	1,686985	ARXN(6) without
96	1,401721	1,695225	ARXF(6) without

52	1,1596728	1,4237387	ARF(6) with
53	1,1596728	1,4237387	ARN(6) with
54	1,1394511	1,4329186	ML(N)_forest_(3)_with()
55	1,1863967	1,4332474	ARF(6) without
56	1,1863967	1,4332474	ARN(6) without
57	1,1673525	1,4508027	ML(N)_linear_(3)_with()
58	1,1601813	1,4526114	ML(N)_forest_(6)_with(out)
59	1,1766626	1,4540294	ML(F)_forest_(3)_with()
60	1,1641789	1,4552003	ML(F)_forest_(6)_with(out)
61	1,2024587	1,4562368	ARXF(1) with
62	1,1746619	1,4654711	ML(F)_forest_(6)_with()
63	1,1951261	1,4655416	ML(F)_linear_(3)_with()
64	1,1967961	1,4687282	ML(F)_forest_(3)_with(out)
65	1,1660076	1,4716300	ML(N)_linear_(1)_with()
66	1,2148662	1,4721471	ML(F)_linear_(1)_with()
67	1,1818848	1,4733095	ML(N)_forest_(1)_with()
68	1,1785028	1,4793365	ML(N)_forest_(1)_with(out)
69	1,1945795	1,4825607	ML(F)_forest_(1)_with()
70	1,2099149	1,4915188	ML(N)_linear_(6)_with()
71	1,2068217	1,4957635	ML(F)_forest_(1)_with(out)
72	1,2090737	1,4976305	ARXN(3) without
73	1,2228982	1,5037636	ARXF(3) with
74	1,2204245	1,5137196	ML(N)_elastic_(3)_with(out)
75	1,2195705	1,5172058	ARXN(1) without
76	1,2248887	1,5191053	ML(N)_linear_(3)_with(out)
77	1,2294444	1,5321329	ML(N)_elastic_(1)_with(out)
78	1,2000049	1,5342734	ML(N)_forest_(6)_with()
79	1,2323141	1,5377777	ML(N)_linear_(1)_with(out)
80	1,2695717	1,5385909	ML(F)_linear_(6)_with()
81	1,2682952	1,5504312	ARXN(6) with
82	1,2957183	1,5511160	ML(N)_elastic_(6)_with(out)
83	1,2957644	1,5563795	ARXN(6) without
84	1,3030537	1,5574170	ML(N)_linear_(6)_with(out)
85	1,3390352	1,5992765	ARXF(6) with
86	1,3420624	1,6268457	ARXN(3) with
87	1,3147104	1,6325398	ARXN(1) with
88	1,3565635	1,6341912	ARXF(1) without
89	1,3468965	1,6351583	ML(F)_elastic_(3)_with(out)
90	1,3634457	1,6427973	ML(F)_elastic_(1)_with(out)
91	1,3549161	1,6457066	ML(F)_linear_(3)_with(out)
92	1,3712457	1,6530436	ML(F)_linear_(1)_with(out)
93	1,3678401	1,6544306	ARXF(3) without
94	1,3751617	1,6830552	ML(F)_elastic_(6)_with(out)
95	1,3871484	1,6962012	ML(F)_linear_(6)_with(out)
96	1,4476574	1,7789062	ARXF(6) without

52	1,1205726	1,3835881	ML(F)_boost_(1)_with()
53	1,0956489	1,3847421	ML(N)_linear_(3)_with()
54	1,1121368	1,3849641	ML(F)_elastic_(1)_with()
55	1,1136270	1,3902469	ARF(1) with
56	1,1136270	1,3902469	ARN(1) with
57	1,0823452	1,3940779	ML(N)_elastic_(1)_with(out)
58	1,1176804	1,3968341	ARF(1) without
59	1,1176804	1,3968341	ARN(1) without
60	1,1160001	1,3980940	ARXN(3) without
61	1,0866963	1,3983095	ML(N)_linear_(1)_with(out)
62	1,1189396	1,4023408	ML(F)_forest_(3)_with(out)
63	1,1386686	1,4030192	ML(N)_forest_(3)_with()
64	1,1626749	1,4033430	ML(F)_boost_(6)_with(out)
65	1,1242804	1,4058619	ARXF(3) with
66	1,0943956	1,4092760	ML(N)_linear_(1)_with()
67	1,1354904	1,4093765	ML(F)_forest_(6)_with(out)
68	1,1499505	1,4099219	ML(N)_elastic_(6)_with()
69	1,1532063	1,4129409	ML(F)_elastic_(6)_with()
70	1,1063151	1,4132901	ARXN(1) without
71	1,1339388	1,4189093	ML(F)_forest_(1)_with(out)
72	1,1599313	1,4217375	ML(F)_linear_(3)_with()
73	1,1402822	1,4224199	ML(N)_elastic_(6)_with(out)
74	1,1596728	1,4237387	ARF(6) with
75	1,1596728	1,4237387	ARN(6) with
76	1,1517553	1,4313019	ML(N)_linear_(6)_with(out)
77	1,1863967	1,4332474	ARF(6) without
78	1,1863967	1,4332474	ARN(6) without
79	1,1464002	1,4441545	ML(N)_linear_(6)_with()
80	1,1995480	1,4449268	ARXF(3) without
81	1,1818220	1,4498167	ML(F)_linear_(1)_with()
82	1,1765939	1,4543298	ML(F)_forest_(3)_with()
83	1,1689597	1,4589217	ML(F)_boost_(6)_with()
84	1,1808590	1,4610985	ARXF(6) with
85	1,2278560	1,4712788	ML(F)_elastic_(3)_with(out)
86	1,2328650	1,4775159	ML(F)_linear_(3)_with(out)
87	1,1780743	1,4800553	ML(F)_forest_(1)_with()
88	1,2465627	1,4827237	ARXF(6) without
89	1,2132332	1,4881754	ML(F)_elastic_(1)_with(out)
90	1,1956045	1,4886070	ARXN(6) without
91	1,2177685	1,4941973	ML(F)_linear_(1)_with(out)
92	1,2185376	1,4959822	ARXF(1) without
93	1,2314048	1,4985009	ML(F)_linear_(6)_with()
94	1,2265257	1,5001978	ML(F)_forest_(6)_with()
95	1,2835127	1,5441532	ML(F)_elastic_(6)_with(out)