

Negative Economic Sentiment Index Based on Finnish News Titles*

Aleksi Avela^{1,†} And Markku Lehmus²

August 28, 2023

Abstract

We construct an index for measuring negative economic sentiment in Finland by using news titles collected from the Finnish broadcasting company Yle's archive. Our approach uses supervised machine learning text classification for detecting news titles featuring negative economic sentiment, and the monthly aggregated proportional frequencies of those titles are then used for defining the index. We find a negative correlation between our index and the consumer confidence index by Statistics Finland, and more remarkably, our index seems to lead the consumer confidence index, somewhat, by one month. We also show that our index correlates positively with Finnish stock market volatility. In addition, based on a simple VAR model, we examine how certain macro variables respond to changes in economic sentiment and show that our index could prove helpful in assessing the current and near-future state of the Finnish economy.

Keywords: *economic sentiment, natural language processing, machine learning, naive Bayes classifier, economic forecasting*

JEL codes: *D80, E27, E37*

*The authors gratefully acknowledge Jetro Anttonen, Olli Ropponen, Pauliina Ilmonen, the seminar participants at the Bank of Finland, and the two anonymous reviewers for the valuable feedback and comments. The authors also wish to sincerely thank Yle for providing an access to their news data archive, and ETLA Economic Research for providing the facilities for the first stages of this work. Aleksi Avela acknowledges support from the Academy of Finland via the Finnish Centre of Excellence in Randomness and Structures (decision number 346308), and the personal grant from The Emil Aaltonen Foundation (Nuoren tutkijan apuraha, numero 230014).

†Corresponding author: Aleksi Avela (aleksi.avela@aalto.fi)

¹ Department of Mathematics and Systems Analysis, Aalto University School of Science, P.O. Box 11100, FI-00076 Aalto, Finland (email: aleksi.avela@aalto.fi)

² Bank of Finland, P.O. Box 160, 00101 Helsinki, Finland (email: markku.lehmus@bof.fi)

1 Introduction

Economic sentiment indicators (i.e., soft data) have become increasingly important in economic forecasting and, especially, in nowcasting applications. One of the main advantages of soft indices compared to hard macroeconomic data (such as GDP) is that they become available earlier and thus can help providing a timely assessment of the current and near-future state of the economy, see, e.g., Bortoli et al. (2018). For instance, Angelini et al. (2011) show that, when nowcasting GDP of the current economic quarter, the further away we are from the end of the present quarter, the larger the relative importance of qualitative data is compared to quantitative data.

In addition to rapid availability, indicators that measure sentiments, anticipations, or uncertainty levels of people involved (e.g., by surveying consumers or purchasing managers) can capture some valuable information that would not (yet) be visible in hard data, see, e.g., Ristolainen et al. (2021). Take, for instance, the effects of uncertainty on the economy. For example, Bernanke (1983) shows that economic uncertainty can encourage firms to postpone investment and hiring. Moreover, uncertainty may increase precautionary spending by households and raise the cost of finance (Gilchrist et al., 2014; Pástor and Veronesi, 2013), as well as increase managerial risk-aversion (Panousi and Papanikolaou, 2012), thus discouraging new investment projects. The ramifications of the effects mentioned can take months or even years to materialize to be visible in hard data.

Although the existing survey-based soft indices are already available earlier than hard economic data, there have still been efforts in sentiment literature to develop even faster approaches for measuring economic sentiment. Mainly, these methods include automated news coverage based indicators that model economic sentiment or uncertainty by measuring frequencies of certain words or phrases in the news articles. These indicators can be based on logical rules constructed by humans, on sentiment dictionaries, or on machine learning methods.

Approaches that utilize news data include, for example, Bortoli et al. (2018), who nowcast French GDP by labeling news articles based on a sentiment dictionary, and Larsen and Thorsrud (2019), who apply Latent Dirichlet Allocation (LDA) topic modeling approach (Blei et al., 2003) for identifying topics in the news coverage and study the role of those topics in forecasting and explaining economic fluctuations. LDA is also applied by Bybee et al. (2020), who use topical information retrieved from Wall Street Journal articles as an input signal in an economic timeseries model, and by Ristolainen et al. (2021), who show that financial crises are foreshadowed by changes observed in the narrative information of newspaper article titles.

News coverage offers high-frequency data that can be argued to describe (as well as to have an effect on) the general economic sentiment and it is therefore widely applied in sentiment literature. Nevertheless, other text data sources than news articles can also be applied in sentiment analysis. For example, Hassan et al. (2019) measure the share of U.S. firms' quarterly earnings conference calls that include discussion about political risks. They use this measure for analyzing the effects of political risks on individual U.S. firms, for example, in the form of reduction of new hirings and investments.

Some of the most applied survey-based soft economic indices in Finland include the consumer confidence index measured by Statistics Finland and the eurozone Purchasing Managers Indices (PMIs). Globally, the Economic Policy Uncertainty (EPU) index by Baker et al. (2016) is a widely used (automatic) soft economic indicator. The EPU index reflects the frequency of articles that include the following triple: (i) 'economic' or 'economy', (ii) 'uncertain' or 'uncertainty', and (iii) one or more of 'congress', 'deficit', 'Federal Reserve', 'legislation', 'regulation' or 'White House'. Based on the approach introduced by Baker et al. (2016), uncertainty indices have also been developed for Sweden (Armeliu et al., 2017) and the Netherlands (Kok et al., 2015), for instance.

Our approach for modeling negative economic sentiment in Finland is based on measuring the frequency of negative sentiment featuring in Yle's economic news coverage. We assume that the overall public sentiment about negative economic development in Finland is significantly correlated with sentiment of the considered economic news coverage - yet, without presuming the direction of possible causality. Yle (i.e., the Finnish Broadcasting Company) is 99.9% owned by the Finnish state, tax-funded, and does not receive any advertising revenue (Yle, 2020). In 2021, Yle reached 94% of the

population weekly (Yle, 2021). Thus, we believe that their news coverage is a reliable source for an index describing the collective domestic economic sentiment.

It should be noted that our index is a measure of *negative* economic sentiment in Finland, and the only way that it acknowledges *positive* sentiment is merely as a lack of negative sentiment. The reason that we are focusing particularly on negative sentiment is that usually the aim of soft indices is to help assessing and managing potential future risks in economic development. That is, notable increments in the level of negative sentiment is the signal we are trying to capture. In fact, there are also previous works in sentiment literature that focus particularly on negative sentiment. For example, Tetlock (2007) studies the connection between pessimistic sentiment in a popular Wall Street Journal column and the stock market. He shows that elevated pessimistic sentiment in media foreshadows downward pressure on market prices, and, on the other hand, that abnormally low or high pessimism forecasts high trading volumes. Likewise, the EPU index by Baker et al. (2016), and other similar uncertainty indices, are based on measuring the level of *uncertainty* without (directly) considering the opposite.

We utilize supervised machine learning in order to develop a text classifier for detecting news items featuring negative economic sentiment. That is, we let the classification model learn the significant words and their respective weights instead of selecting these words by hand or by using predefined sentiment dictionaries. The weightings of relevant words can later be used as a sanity check for the indicator. When discussing the results, we also inspect the words that seem to be the driving factors behind observed increments in negative sentiment, thus providing tangible narrative for the detected peaks.

By applying a machine learning approach, our model can learn varying negative sentiment weights for different words without the need to manually go through the words and to assemble a negative sentiment dictionary (and sentiment weights) by ourselves. That said, we still have to annotate training data for the classifier by hand, i.e., to create a labeled training set of news items. The annotation process and labeling system are further discussed in the following section.

The contribution of our paper is two-fold. First, we apply natural language processing and text classification techniques to the morphologically complex Finnish language. Second, we aim at developing a novel soft indicator that could help assessing the current state of domestic economic sentiment. We validate our index by comparing it to a few other indicators measuring sentiment and/or uncertainty in the Finnish economy. However, our objective is not to develop a substitute for these measures, but rather to study the possibility of including text data in macroeconomic analyses to enrich our understanding of the current and near-future state of the Finnish economy. Section 2 describes the methods and data used in the paper, Section 3 evaluates and discusses the results, and the final section concludes the work.

2 Methods

The idea of the proposed index is to detect news items that embody negative sentiment on the domestic economy, and consequently, to describe the overall negative economic sentiment in Finland as the share of negative economic news aggregated over some time periods, for example, on a monthly basis. It is important to notice that a single value of this indicator (or, arguably, of any other soft indicator) is not informative as such, but the aim is to evaluate its development over time. For a time period including a set of news items $d_j \in T$, the value of our negative sentiment index S_T is defined as

$$(1) \quad S_T = \frac{\sum_{d_j \in T} [\delta_e(d_j)\delta_n(d_j)]}{|T|},$$

where $\delta_e(d_j) = 1$ if news item d_j is labeled to represent economic news and zero otherwise, and $\delta_n(d_j) = 1$ if news item d_j is labeled to feature negative economic sentiment and zero otherwise. The normalizing term $|T|$ is the total number of news items in the set T .

The sentiment classification in our model is binary, i.e., all negative news are *equally negative*. There are two reasons behind this approach. First, negative economic sentiment is a subjective concept as such, and we did not want to make the training data annotation any more complicated than necessary. Second, the idea is that every time there are some *more negative* news, they are also likely to generate a large number of some other *less negative* news and thus raise the value of the aggregated negative sentiment index. As news within a given time period are dependent on each other, we believe that aggregated binary classification is an adequate approach.

2.1 Data collection and annotation process

The data, for both training and testing the classifier as well as for producing the sentiment index, has been obtained from the Finland's national public broadcasting company Yle. Each data row consists of the title of the news item, a topic tag attached to it, and the release date. There are a few reasons why we do not use the whole articles but rather only the titles. Considering that the classification model we are proposing for this task is rather simple, we assume that usually the title includes almost all of the relevant information of sentiment of the news item. Moreover, processing only the titles makes labeling much faster for both humans and the classifier.

The complete data set covers all Yle's news titles from the start of 2010 to September 2021. The annotated data set consists of 12,155 news titles of which 10,000 observations are used for training the classifier and 2,155 observations for testing it. Following a standard convention for classification of chronological data, the data has been divided into training and test sets also in a chronological manner. That is, observations in the first four and half months of the annotated set are used for training the classifier and the remaining one month is used for testing. In many machine learning applications, the train-test split is done randomly, but as here the classifier is trying to classify observations in the past and in the future based on the training data in the present, it would gain unnecessary advantage in the test set if the training set was selected randomly (given that news topics are time dependent).

First, we read some titles and then decided on the rules which we were going to follow in the annotation. Naturally, we discussed these rules and updated them when necessary throughout the whole annotation process. In addition, we performed consistency checks and discussed the labels with our colleagues. As said, there are two data features, and both of them are labeled with binary classification: (i) whether we think that the news item considers (domestic) economy, and (ii) whether we view its economic sentiment to be negative (only considered if the title is classified to be economic news). In our approach, negative sentiment means that the title must refer to something that has, or will have, straight-forward and obvious negative effects on the domestic economy or on households, for example, a furlough or termination of employment, a bankruptcy, a strike, or an economic downturn.

As our main motivation is to explore the ability of news data to depict current economic sentiment, we defined the annotation rules to consider news that can be viewed to imply (or to cause) negative economic sentiment in the short-term horizon. For instance, while the correlation between strikes and longer run economic performance is unclear, strikes naturally imply disruptions in production in the short run (think, e.g., losses of labor hours in factories). Strikes may also be a sign of economic policy disputes. For these reasons, news about strikes are assumed to imply negative sentiment in the very near future. On the other hand, when considering the sentiment of individual people, it is intuitively clear how, e.g., news about furloughs or layoffs can (at least temporarily) increase negative sentiment and uncertainty among employees.

We split the classification process into two consecutive steps (i.e., to first detect the economic topic and then negative sentiment) due to the imbalanced nature of the data (the share of negative economic news from the training data is only about 4%). We could have also developed a single classifier for detecting the economic news with negative sentiment directly, however, then the classification would have been quite notably imbalanced. Many studies have shown that data imbalance often compromises the performance of most standard classifiers, see, e.g., He and Garcia (2009). After splitting the classification steps, the separate classification tasks are closer to being balanced (in the first step the share of economic news is about 15%, and in the second step the share of negative titles from the economic news is about 29%). Moreover, we believe that separating these two classification steps helps both classifiers to focus on the relevant words, respectively.

2.2 Text classifiers and data transformations

Our approach is based on modeling the news titles as bag-of-words vectors, where the index i in a document vector represents the count of word w_i in that document. As bag-of-words vectors do not measure similarity between different words, forms of words are usually normalized with stemming or lemmatization so that similar words account for the same index in the vectors. The Finnish language has a complex system of word affixes, and thus word form normalization is especially important, see, for example, Singh and Gupta (2016). We apply a simple truncating n -stemmer as it is an efficient approach for morphologically complex languages and effortless to implement. The n -stemmer preserves only n characters from the start of each word and discards the rest. We set the value of preserved characters as $n = 7$ after experimenting with a few various values.

The Finnish language includes a number of compound words that cannot be separated by traditional tokenization that uses white spaces to separate the words. For example, ‘economic crisis’ is in Finnish ‘talouskriisi’, where ‘talous’ means ‘economic’ and ‘kriisi’ means ‘crisis’. We use a small set of relevant words (such as ‘kriisi’) and split every compound word that contains any of these words. In addition, we use a Finnish dictionary for finding and separating other compound words by testing whether a word can be divided into two parts such that both parts could be found from the dictionary, however, by excluding short words from these comparisons. We also apply stopword removal which is a common pre-processing step in text classification for feature space reduction and for decreasing noise in the data. The stopwords consist of both very common and very infrequent words which are assumed to hold only little semantic information.

We use lists of words that are replaced by certain tags, for example, names of (large) Finnish companies are replaced by a tag ‘Finnish Firm’. There are two reasons for this. First, we aim to help the first classifier (which objective is to detect economic news) by grouping relevant words together. A majority of these word sets are selected to have a positive correlation with the economic news label, such as the group of Finnish companies. However, some of the constructed sets might also have a negative correlation, for example, we group together sports and names of sports leagues. The second reason is that we can then block the second classifier (considering sentiment of the title) from using these tagged words in order to prevent the classifier from learning, e.g., that a certain company name would correlate with negative economic sentiment. Other replaced lists of words consist of numbers, foreign countries, Finnish cities, and names of Finnish political parties. Word forms representing ‘Finland’ or ‘Finnish’ and ‘economy’ are also replaced with respective tags.

Classification in our index consists of two separate binary text classifiers, as shown in Equation 1, and we use a naive Bayes classifier with a data-based class prior for both. Even though the feature independence assumption made by naive Bayes is strong, studies have shown that naive Bayes classifiers are able to produce desirable results even in applications (such as text classification) where the features cannot be assumed to be independent, see, e.g., Domingos and Pazzani (1996) and Rennie et al. (2003). We utilize an implementation of naive Bayes featured in the scikit-learn package (Pedregosa et al., 2011), which is a well-established machine learning library designed for Python.

In our context, certain distinctive advantages of naive Bayes classifier are the intuitively understandable parameters, that is, the word weights, that we can use for deeper analysis when trying to understand the observed changes in economic sentiment, as well as the ability of naive Bayes classifier to update the word weight estimates if new labeled data becomes available without the need to train the whole model again from the scratch. These features make naive Bayes a transparent approach that can be somewhat easily updated if new labeled data is collected.

Naive Bayes classifier $\delta(\cdot)$ assigns an observation d_j to class c_k (where, in binary classification, $c_k \in \{0, 1\}$) that maximizes the posterior log probability:

$$(2) \quad \delta(d_j) = \arg \max_{c_k} \log \hat{P}(c_k) + \log \hat{P}(d_j|c_k),$$

where $\widehat{P}(c_k)$ is the class prior probability estimate (based on the class frequency in the training data) and $\widehat{P}(d_j|c_k)$ is the class likelihood estimate. The marginal likelihood is omitted as it is a constant with respect to the maximization problem. Based on the *naive* independence assumption, the log likelihood estimate is computed as

$$(3) \quad \log \widehat{P}(d_j|c_k) = \sum_{i=1}^n f_{i,j} \log \widehat{P}(w_i|c_k),$$

where n is the size of the training vocabulary, $f_{i,j}$ is the frequency of word w_i in document d_j , and $\widehat{P}(w_i|c_k)$ is an estimate of the probability of word w_i appearing in class c_k .

We apply traditional preprocessing steps to the bag-of-words vectors before computing the word probability estimates (i.e., the word weights). We first transform the vectors with tf-idf weighting and L_2 -normalization, and then apply Laplace smoothing to the transformed word frequencies, as these techniques have been shown to improve the performance of naive Bayes classifier, see, e.g., Rennie et al. (2003). The word weights for each class are then computed as maximum likelihood estimates based on the transformed and smoothed vectors.

The tf-idf weighting consists of a sublinear tf-transformation, such that each term frequency tf in a vector is replaced with $\ln(\text{tf}+1)$, as well as of an inverse document frequency (idf) transformation, which decreases the weight of a word based on its frequency in the training data. The idf-weighting for word w_i is defined as $\text{idf}(w_i) = \ln \frac{N_d}{\text{df}(w_i)}$, where N_d is the total number of training documents and $\text{df}(w_i)$ is the count of documents that include word w_i . Finally, the purpose of Laplace smoothing is to prevent zero word probability estimates by adding a pseudo observation containing every word in the vocabulary to each class.

3 Results and discussion

Evaluation of our model and its results is two-fold. First, we have to consider whether the text classifiers produce satisfying results. After that, we need to find a way to validate the results of our sentiment index.

As an initial sanity check, we can inspect those words that get the highest weights with respect to the two classification tasks. Words that have a high weight for a topic considering economy consist, for instance, of tags representing Finnish firms and economy, as well as of words 'industry', 'employees', and 'trade'. On the other hand, considering (negative) sentiment of the title, the set of words with the largest weights include, for instance, different forms of 'lay-off', 'strike', 'crisis', 'bankruptcy', and 'sanction'. These sets of words seem intuitively justifiable, which is encouraging for the applicability of our indicator.

Instead of plain accuracy score, we apply balanced accuracy (BA) for evaluating the classification performance due to the imbalance in our dataset. Balanced accuracy is defined as the average of true positive rate (number of true positives divided by the number of positive observations) and true negative rate (number of true negatives divided by the number of negative observations). We assess the two classifiers both separately and also chained together, that is, as shown in the numerator of Equation 1.

The classification performance is shown Table 1, where in-sample results correspond to classification of the training set and out-of-sample results to classification of the test set, that is, observations which the classifiers did not encounter in the training phase. In addition, the area under the curve (AUC) measures of the classifiers are reported in Table 2. The AUCs are based on the receiver operating characteristic (ROC) curves which are presented in Appendix A.

Table 1: BA scores (and accuracy scores) of the classifiers.

Classifier	In-sample	Out-of-sample
Economic News	0.87 (0.95)	0.72 (0.87)
Neg. Economic News	0.92 (0.99)	0.69 (0.95)
Chained	0.84 (0.98)	0.62 (0.94)

Table 2: AUC measures of the classifiers.

Classifier	In-sample	Out-of-sample
Economic News	0.98	0.89
Neg. Economic News	0.97	0.72

In our application, however, where the goal is to produce a time-series index, we are not necessarily interested in the classification performance on the observation level, but on the aggregate level. Therefore, we may also evaluate the classification performance by comparing the aggregated classifications with the aggregated ground truth labels. The (Pearson) correlation coefficients between the levels (and changes) of daily aggregated time series of the classification results and the ground truth labels are shown in Table 3.

Table 3: Correlation on levels (and on changes) between the daily aggregated classifications and the ground truth labels.

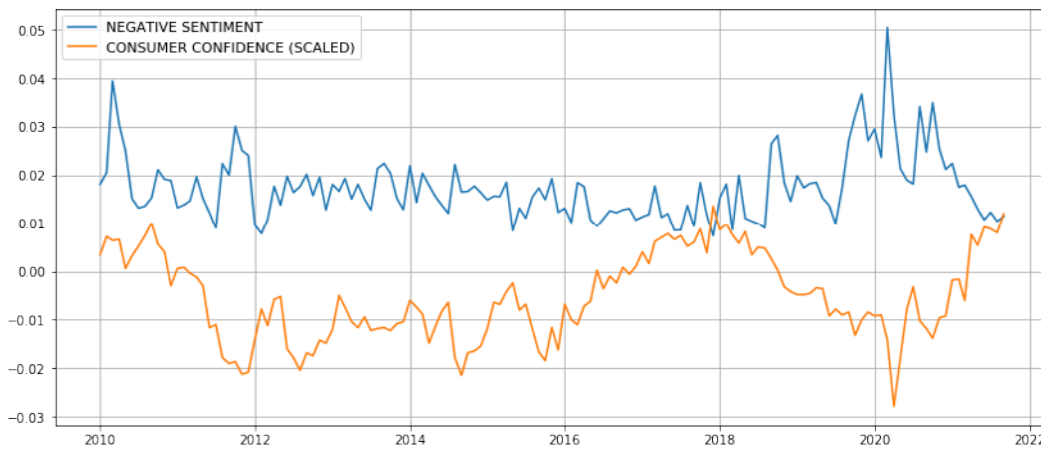
Classifier	In-sample	Out-of-sample
Economic News	0.93 (0.91)	0.90 (0.73)
Chained	0.89 (0.84)	0.77 (0.58)

The labeled data set consists in total of some 12,000 news titles, and as this set corresponds to only five full months in time, we decided to evaluate the aggregates on a daily level in order to obtain more meaningful results. On the other hand, it can be assumed that aggregating the classifications over longer periods can produce more robust results, and it should be reminded that our final negative sentiment index is aggregated on a monthly basis.

To assess the applicability of our index, it can, for instance, be compared to the consumer confidence index measured by Statistics Finland, as shown in Figure 1. Their index is one of the most commonly used soft indices to consider the state of the Finnish economy. Our index correlates negatively with the consumer confidence index, with a correlation value of -0.33 on levels and -0.15 on changes. Another important result is that our sentiment index seems to lead the consumer confidence index, somewhat, by one month. The correlation value of the one-month lead of the sentiment index and the consumer confidence index is -0.36 on levels (and -0.23 on changes). Our index, for instance, recognizes the COVID-19 crisis one month earlier than the consumer confidence index.

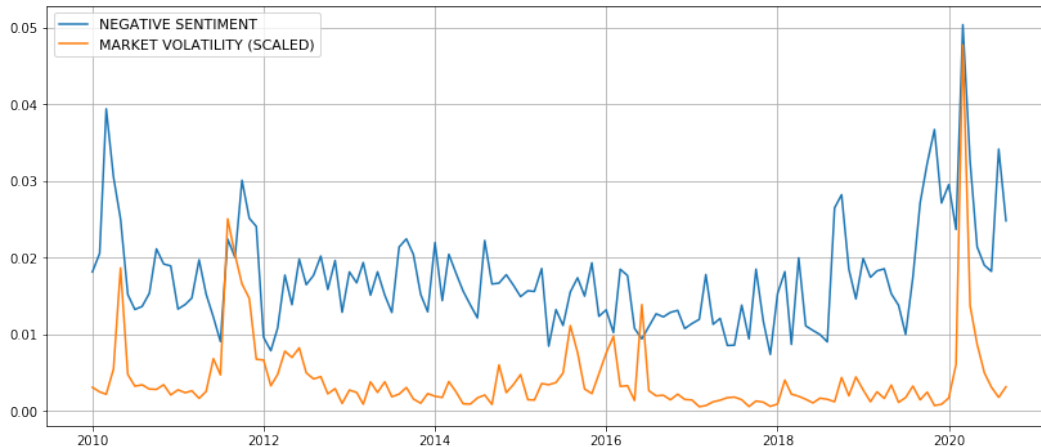
While the negative correlation between these two indices is valuable, we do not intentionally aim at matching the changes in the consumer confidence index. The purpose of our index is to be a measure of negative sentiment in the economy, but not the other way around. Thus, from the other perspective, increasing the number of positive news is not necessarily reflected in our index, but can instead be reflected in the consumer confidence index. This is probably one of the reasons for lesser correlation between the sentiment index and the consumer confidence index around years 2014–2017, when the Finnish economy first contracted but finally started to grow after the long shadow of the financial crisis and its after-shock, the euro crisis.

Figure 1: Our negative sentiment index and the consumer confidence index by Statistics Finland have a negative correlation. For example, COVID-19 generates clear peaks in both indices.



The sentiment index can be compared to other economic indicators, too. For example, our index has a correlation of 0.44 on levels (and 0.42 on changes) with the volatility index of the Finnish stock market as reported by the Bank of Finland. Stock market volatility is often associated with economic uncertainty and instability, and, on the other hand, uncertainty is strongly affiliated with negative economic sentiment, and thus there should be some correlation between the two phenomena. These two indicators are shown alongside in Figure 2.

Figure 2: Negative sentiment index and the Finnish stock market volatility index as reported by the Bank of Finland.



Although the correlation coefficients (on levels and on changes) between our index and stock market volatility and consumer confidence index (concurrent and leading) are not particularly large in magnitude, they are intuitively in the right direction, and all of them, except concurrent correlation on changes with consumer confidence, are statistically significant.¹ Yet, it should be reminded that the main motivation for these comparisons is to validate our index and to perform sanity checks. As said, our index does not aim at matching any other soft indicator, and we view our index more as a possible source of supplementary information in addition to existing measures rather than simply as a substitute.

The comparison of sentiment and market volatility also provides some justification for the observed negative sentiment peak in early 2010. By looking at the news titles that correspond to the increased negative sentiment in March 2010, it seems that the temporary negative sentiment was mostly due to broad strikes in Finnish ports and transport that generated

¹ Based on (Bonferroni corrected, $c = 6$) two-tailed permutation tests with $\alpha = \frac{0.05}{c}$.

uncertainty regarding foreign trade. Later in May 2010, there were also concerns about the real economy. The other two clear spikes in stock market volatility, namely, the European debt crisis in autumn 2011 and the COVID-19 in early 2020, are also visible in our sentiment index.

Interestingly, increased market volatility in early 2010 is already foreshadowed, a month or two earlier, in the negative sentiment index, but on the other hand, the European debt crisis in late 2011 materializes instantly in the volatility index but generates elevated negative sentiment with a few months delay. It seems that, depending on the type and context of the negative shock, news data based information may have different levels of leading information.

Naturally, the largest negative shocks in the span of our data contribute heavily to the correlations between our index and the evaluated counterparts. While it is certainly desirable that these indices have strong connections during major negative events, it might still be interesting to inspect their correlations in periods without any large negative events. The two most notable negative shocks in our data are the European debt crisis in the beginning of the 2010s and the COVID-19 in 2020. The turnaround point in the European debt crisis is usually considered to be the so-called “whatever-it-takes” speech by the ECB’s then-president M. Draghi in the summer of 2012. Based on this, we shorten the considered indicators to range from July 2012 to December 2019 in order to evaluate our index within a time period without any large negative shocks.

Within this shorter time interval, our index has statistically significant² correlations on levels with the consumer confidence index (-0.37 for both concurrent and leading correlation). However, correlations on changes with the consumer confidence index and correlations with the stock market volatility index are much smaller in magnitude. A possible explanation for the lesser correlations is the fact that the Finnish economy was in almost continuous upturn from around 2015 to the beginning of 2019. As stated earlier, our index is a measure of negative sentiment in the economy and not the other way around, and therefore it is not surprising that, particularly with stock market volatility, during this interval the correlations are smaller.

Another important remark is that usually there are (at least) two interesting phases in any crisis, namely, the beginning of the crisis and later on its wane. While detecting the start of an economic downturn may, at least in hindsight, seem trivial, recognizing the latter phase, where the economy commences to heal, is of great importance as well and arguably not so straightforward. Judging, for example, by Figures 1 and 2, our index seems accurate also in identifying the turning points of the major crises; a feature which, in this shorter interval considered, is not displayed either.

We can also use the sets of words associated with different peaks in the negative sentiment index to get a grasp of phenomena that are behind the periods of elevated negative sentiment. It seems, at least based on our approach, that the instability generated by the European debt crisis in late 2011 materialized in the domestic economic sentiment in the form of economic uncertainty and concerns about employment. In September 2018, the government was aiming to loosen the conditions for lay-offs in small businesses, which initially caused broad objections from trade unions. This negative sentiment peak can also be seen in our index, where words regarding economic policy, job actions, and employment have high frequencies in the titles.

The elevated negative sentiment in around 2019 was probably mostly caused by the parliamentary crises in that year (first due to the failed reform of health and social services in early 2019, and later due to the incidents concerning the labor dispute in the Finnish postal service in late 2019) which both eventually led to resignations of the cabinets. On top of that, in autumn 2019, there were notable issues in air traffic (e.g., the collapse of the Thomas Cook Group) which generated concerns also in the Finnish economy. Last, the global COVID-19 crisis in early 2020 appears in our index mainly as uncertainties about employment and financial conditions, and, later in the same year, also as negative news concerning the domestic real economy.

Finally, we analyze the relationship of our negative economic sentiment index with output and employment in a simple VAR model ($N = 139$). This can be seen as a validation of the indicator we have developed, but also, by using this tool we

² Based on (Bonferroni corrected, $c = 6$) two-tailed permutation tests with $\alpha = \frac{0.05}{c}$.

can examine how the selected macro variables respond to elevated negative economic sentiment. We add a trend indicator of output and (trend) employment provided by Statistics Finland to our VAR specification, consisting also of lags of two periods (i.e., months) for each of the variables. The order of the variables is assumed to be the following: economic sentiment, output, and employment. Thus, we assume that negative sentiment precedes changes in production which then affect employment with a lag.

According to standard impulse response functions calculated from the VAR model (using Cholesky-type identification), an increase (of one S.D. innovation) in negative sentiment is associated with a decline in trend indicator of output, as shown in Figure 3, and with a decline in employment, as shown in Figure 4. The considered times series as well as all the impulse responses of the model variables are shown in Appendices B and C, respectively. The peak negative effect in output comes in five months, however, the impact of a negative sentiment shock is persistent and decreases only slowly in time. The negative effect on employment is even more persistent, staying clearly negative for 24 months in time as well, implying, as it could be assumed, that heightened negative economic sentiment and economic instability may foreshadow long-lasting scars in employment.

Figure 3: Response of (log) output to one S.D. innovation in negative sentiment (95% confidence bands).

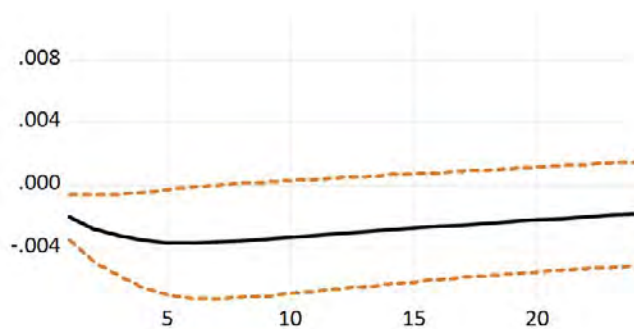
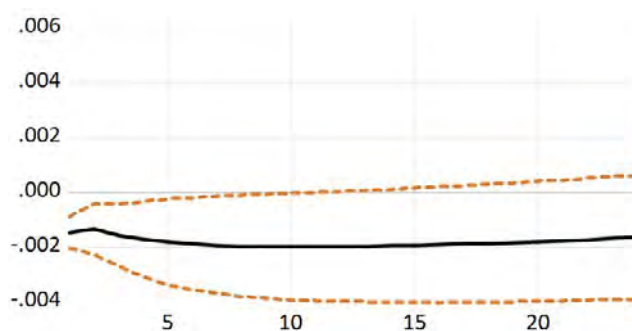


Figure 4: Response of (log) employment to one S.D. innovation in negative sentiment (95% confidence bands).



The VAR results work as a sanity check for our results; to assess whether the patterns and signs of the impulse responses are as expected. The hump-shaped response in output is in our view plausible while it can be expected that at least part of the effects of the shock will fade out in time. The effect of negative sentiment on employment is slightly more persistent, yet this pattern is credible as the impact of the shock slowly fades out in this case as well. Thus, the results, while based on a simple VAR model, suggest that our sentiment index could have potential to be used in further policy analysis.

The results of our work are encouraging, and they are in line with the evidence shown in previous literature. The results suggest that information obtained from news data could even lead some other widely used soft indicators and therefore could be applied in nowcasting as well as in analysis of connections between sentiment and other macro outcomes. However, additional evidence for these conclusions should be gathered in future research. Moreover, although news data based

approaches have been widely studied previously, our work contributes to the use of news data in the Finnish language in economic sentiment analysis.

Our work highlights that even the simplest machine learning models seem to be able to retrieve valuable information from news data. That said, even though our classifiers fared reasonably well in the out-of-sample set, it is clear that there still is room for improvement. Although naive Bayes has many desirable properties, some other, more sophisticated, learning algorithms could offer better classification performance. Finally, about the evaluation of our results, it should be reminded that spurious correlations between time series are known to occur. That said, we analyzed the correlations both on levels and on changes and tested for their significance, and these results suggest that there could be some real dependencies.

4 Conclusion

In this paper, we introduced an approach for measuring negative economic sentiment in Finland based on the frequency of negative sentiment featuring in Finnish news titles. To automate the classification steps, we applied two transformed naive Bayes text classifiers. The first classifier was designed for detecting economic news, and the second for detecting negative sentiment in the economic news.

We then applied this modeling approach to news titles obtained from the Finnish broadcasting company Yle and evaluated the results. We compared our sentiment index with the consumer confidence index by Statistics Finland and found a negative correlation between the two series. Yet, an even more interesting result was that the sentiment index seems to lead the consumer confidence index by one month. This is important, as the consumer confidence index is currently one of the best methods for assessing the state of economic sentiment in the national economy, and the results suggest that our index could provide information even earlier than that. Thus, one of the conclusions of our exercise is that news titles may contain information that can be useful in forecasting purposes as well. Moreover, we showed that our index has a positive correlation with Finnish stock market volatility which is a commonly applied measure of economic instability and sentiment as well.

Our index could be helpful in other kinds of macroeconomic analysis, too. In the final part of evaluation and discussion of our results, we constructed a simple VAR model to assess the usefulness of the sentiment index in policy analysis. This can also be seen as a validation of the developed index. While based on a simple VAR model, the results are intuitive and hence support the possibility to use the developed index in further policy analysis. Moreover, the results suggest that the sentiment index could also be used to consider relationships between negative economic sentiment and important macro variables such as production, employment, and investments.

Arguably, one of the main advantages of an automated news-based sentiment index is that it can capture changes in economic sentiment even before results of survey-based indicators are available - and without the need to manually conduct any surveys. On top of that, as we showed in discussion of our results, when a soft index is built on machine learning classification of news items in this manner, we can also utilize the word weights assigned by the classifier as well as the observed word frequencies in periods of heightened negative sentiment in order to provide possible explanations for the changes in economic sentiment. In addition, because of this, one possible extension of such news-based indicator is to provide (soft) context when studying economic events in (recent) history.

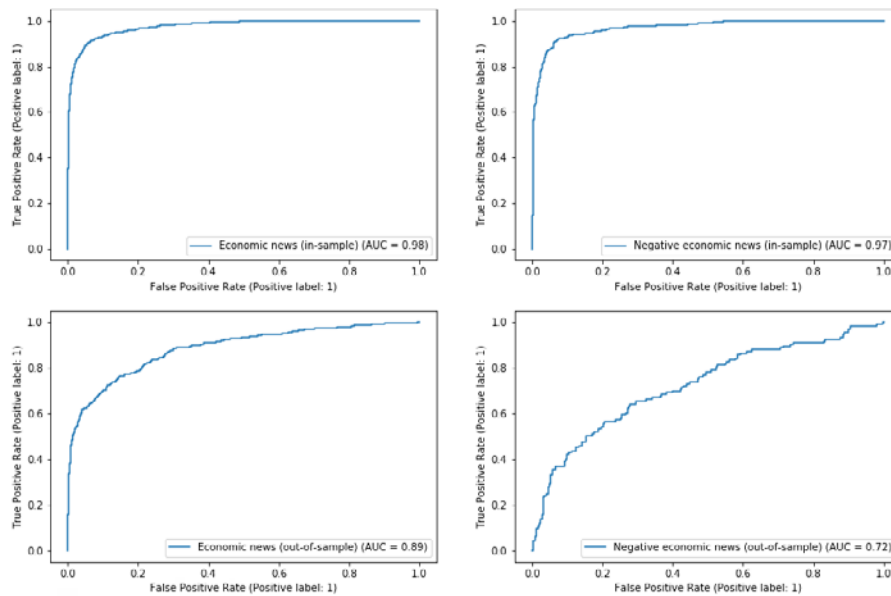
On the other hand, there are also certain risks of spurious conclusions associated with machine learning based indicators that should be carefully considered. First of all, having just one news outlet as a data source may be problematic, although we did provide justifications for our selection. In further research, a similar index could also be developed by utilizing multiple news sources simultaneously. Moreover, machine learning (text) classifiers are prone to learn some kinds of biases, e.g., to associate some words with a certain class due to random variations in data. However, we tried our best to tackle this issue, for example, by blocking the classifier from learning associations of negative sentiment with any political parties or (large) companies.

We applied a fairly simple text classifier that processes the news titles as 'bag-of-words' without the ability to reflect any more complicated structures of text. We also used a simple stemming algorithm in this work, and some other, more sophisticated stemmer, could possibly be experimented with in our future research as well. Yet, using a simple model is certainly justifiable, especially, when there is a limited amount of data available for training the classifier. Future prospects for our index include experimenting with other classification algorithms, utilizing multiple news sources, and considering full news articles instead of just the titles. As a final remark, it should be noted that, when working with automated soft indicators, any conclusions drawn should be made by humans rather than by relying on the indicator as such.

References

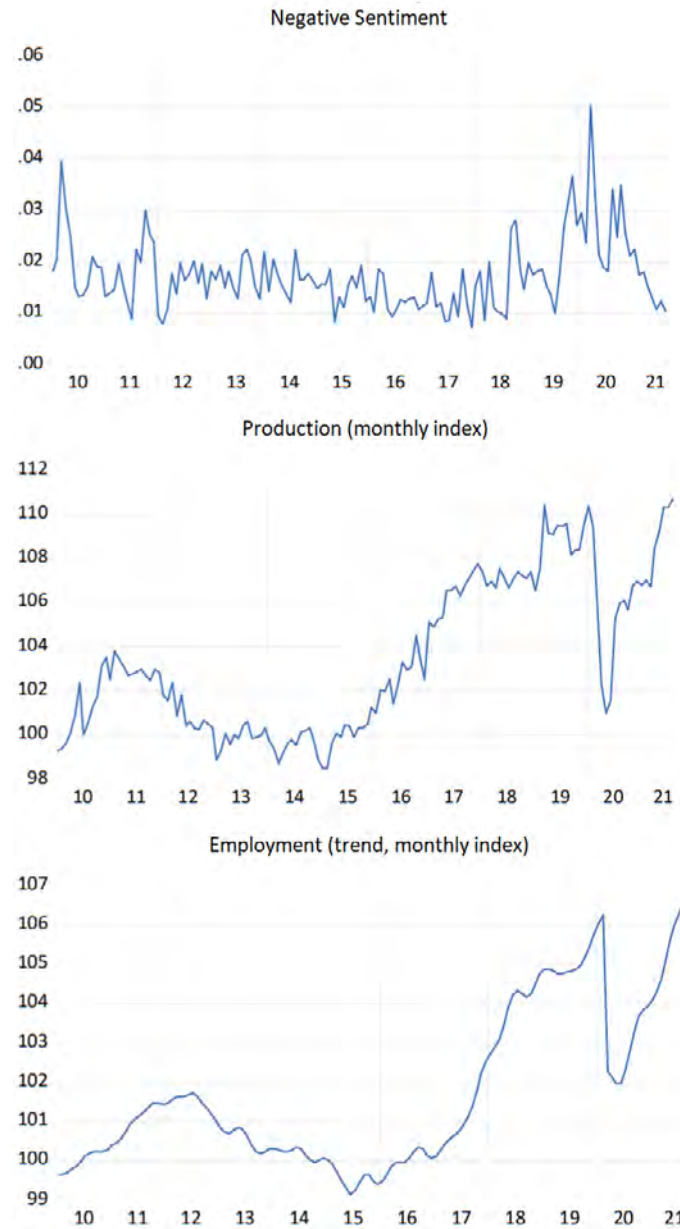
- Angelini, E., Camba-Mendez, G., Giannone, D., Reichlin, L., and Rünstler, G. (2011). "Short-term forecasts of euro area GDP growth". *The Econometrics Journal*, 14(1):C25–C44.
- Armeliu, H., Hull, I., and Stenbacka Köhler, H. (2017). "The timing of uncertainty shocks in a small open economy". *Economics Letters*, 155:31–34.
- Baker, S. R., Bloom, N., and Davis, S. J. (2016). "Measuring economic policy uncertainty". *The Quarterly Journal of Economics*, 131(4):1593–1636.
- Bernanke, B. S. (1983). "Irreversibility, uncertainty, and cyclical investment". *The Quarterly Journal of Economics*, 98(1):85–106.
- Blei, D. M., Ng, A. Y., and Jordan, M. I. (2003). "Latent Dirichlet Allocation". *Journal of Machine Learning Research*, 3:993–1022.
- Bortoli, C., Combes, S., and Renault, T. (2018). "Nowcasting GDP growth by reading newspapers". *Economie et Statistique*, 505(1): 17–33.
- Bybee, L., Kelly, B. T., Manela, A., and Xiu, D. (2020). "The structure of economic news". *NBER Working Paper*, 26648.
- Domingos, P. and Pazzani, M. (1996). "Beyond independence: Conditions for optimality of the simple Bayesian classifier". *Proceedings of the 13th International Conference on Machine Learning*, (ICML 1996):105–112.
- Gilchrist, S., Sim, J. W., and Zakrajšek, E. (2014). "Uncertainty, financial frictions, and investment dynamics". *NBER Working Paper*, 20038.
- Hassan, T. A., Hollander, S., Van Lent, L., and Tahoun, A. (2019). "Firm-level political risk: Measurement and effects". *The Quarterly Journal of Economics*, 134(4):2135–2202.
- He, H. and Garcia, E. A. (2009). "Learning from imbalanced data". *IEEE Transactions on Knowledge and Data Engineering*, 21(9):1263–1284.
- Kok, S., Kroese, L., and Parlevliet, J. (2015). "Beleidsonzekerheid in Nederland". *Economisch Statistische Berichten*, 100(4715):464–467.
- Larsen, V. H. and Thorsrud, L. A. (2019). "The value of news for economic developments". *Journal of Econometrics*, 210(1):203–218.
- Panousi, V. and Papanikolaou, D. (2012). "Investment, idiosyncratic risk, and ownership". *The Journal of Finance*, 67(3):1113–1148.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. (2011). "Scikit-learn: Machine learning in Python". *Journal of Machine Learning Research*, 12:2825–2830.
- Pástor, L. and Veronesi, P. (2013). "Political uncertainty and risk premia". *Journal of Financial Economics*, 110(3):520–545.
- Rennie, J. D., Shih, L., Teevan, J., and Karger, D. R. (2003). "Tackling the poor assumptions of naive Bayes text classifiers". *Proceedings of the 20th International Conference on Machine Learning*, (ICML 2003):616–623.
- Ristolainen, K., Roukka, T., and Nyberg, H. (2021). "A thousand words tell more than just numbers: Financial crises and historical headlines". Available at SSRN: <https://ssrn.com/abstract=3961534>.
- Singh, J. and Gupta, V. (2016). "Text stemming: Approaches, applications, and challenges". *ACM Computing Surveys*, 49(3):1–46.
- Tetlock, P. C. (2007). "Giving content to investor sentiment: The role of media in the stock market". *The Journal of finance*, 62(3):1139–1168.
- Yle (2020). This is Yle. <https://yle.fi/aihe/artikkeli/about-yle/this-is-yle>. Accessed: 2023-02-01.
- Yle (2021). Yle's year 2021. <https://yle.fi/aihe/yles-year-2021>. Accessed: 2023-02-01.

Appendix A: ROC curves and AUC measures of the two applied classifiers evaluated on observation level in training (in-sample) and test (out-of-sample) sets



Classifier 'Economic news' detects news titles that consider economy and classifier 'Negative economic news' detects titles with negative sentiment from observations that have been labeled as economic news.

Appendix B: Negative sentiment index, production, and employment (trend)



Negative economic sentiment index presented in this paper and trend indicators of output and employment as provided by Statistics Finland.

Appendix C: VAR model responses to Cholesky one S.D. innovations

