Working Papers | 84

8.12.2020

It's in the News: Developing a Real Time Index for Economic Uncertainty Based on Finnish News Titles

Aleksi Avela
Etla Economic Research
aleksi.avela@etla.fi

Markku Lehmus
Etla Economic Research
markku.lehmus@etla.fi

Suggested citation:
Avela, Aleksi & Lehmus, Markku (8.12.2020). "It's in the News: Developing a Real Time Index for Economic Uncertainty Based on Finnish News Titles".

ETLA Working Papers No 84.
http://pub.etla.fi/ETLA-Working-Papers-84.pdf

Abstract

Uncertainty may affect economic behavior of individuals and firms in a wide variety of ways, with typically negative consequences for economic growth. It is due to this fact, combined with rising political uncertainty observed lately in many countries, that uncertainty has gained increasing attention in economic literature, too. In this paper, we construct a measure of economic uncertainty for Finland based on Finnish news titles, collected from the YLE's (the Finnish broadcasting company) website. To construct the index, we utilize machine learning and natural language processing (NLP) techniques, and in this paper, specifically, a transformed naive Bayes text classifier. On basis of the model evaluation, the constructed uncertainty index seems helpful in giving a timely assessment of the current state of the Finnish economy. We find a strong negative correlation between our index and the consumer confidence index by Statistics Finland, and most remarkably, our index seems to lead the consumer confidence index by one month.

Tiivistelmä

Taloudellisen epävarmuuden arvioiminen suomalaisesta uutisdatasta

Taloudellinen epävarmuus voi vaikuttaa monin tavoin yksilöiden sekä yritysten toimintaan. Epävarmuus päätöksenteossa heikentää usein talouskasvua ja on sen takia herättänyt paljon mielenkiintoa taloustutkijoiden keskuudessa, erityisesti kun poliittisen epävarmuuden kasvua on viime aikoina havaittu useissa maissa. Tässä tutkimuksessa kehitämme taloudellista epävarmuutta arvioivan indeksin, joka perustuu negatiivisten talousuutisten tunnistamiseen Ylen verkkosivuilta kerätystä uutisvirrasta. Hyödynnämme mallissa luonnollisen kielen käsittelyä (eng. natural langauge processing) ja kehitämme muokatun naiivin Bayesilaisen mallin otsikoiden luokittelua varten. Tutkimuksessa osoitetaan, että kehittämämme malli tarjoaa hyödyllistä tietoa Suomen talouden tilasta ja taloudellisen epävarmuuden kehityksestä. Kehittämämme epävarmuusindeksi korreloi voimakkaan negatiivisesti Tilastokeskuksen kuluttajaluottamusindeksin kanssa ja jopa ennakoi sen muutoksia.

B.Sc. (Tech.) **Aleksi Avela** is a Research Assistant at ETLA Economic Research.

PhD (Econ.) **Markku Lehmus** is a Head of forecasting at ETLA Economic Research.

TkK **Aleksi Avela** on Elinkeinoelämän tutkimuslaitoksen tutkimusharjoittelija.

VTT **Markku Lehmus** on Elinkeinoelämän tutkimuslaitoksen ennustepäällikkö.

Acknowledgements: We would like to thank Jetro Anttonen and Olli Ropponen for valuable comments and conversations.

Kiitokset: Haluaisimme kiittää Jetro Anttosta ja Olli Ropposta hyödyllisistä kommenteista sekä keskusteluista.

Keywords: Economic uncertainty, Nowcasting, Machine learning, Natural language processing, Naive Bayes

Asiasanat: Taloudellinen epävarmuus, Nykyhetken ennustaminen, Koneoppimismallit, Luonnollisen kielen käsittely, Naiivit Bayesilaiset mallit

JEL: C45, C53, C61, E71

1 Introduction

Uncertainty may affect economic behavior of individuals and firms in a wide variety of ways. It gives, for instance, firms an incentive to delay investment and hiring when investment projects are costly to undo or workers are costly to hire and fire, as shown by Bernanke (1983). It also increases precautionary spending by households and raises the cost of finance (e.g., Gilchrist et al., 2014, and Pastor and Veronesi, 2013). In addition, it may add to managerial risk-aversion (Panousi and Papanikolaou, 2012), thus discouraging new investment projects.

It is due to these mechanisms, combined with elevated political uncertainty that has lately been observed in many countries, that economic uncertainty has gained increasing attention among economists and public commentators alike. Fortunately, researchers have also invented new ways of measuring uncertainty. One of the most sited papers on this field is Baker et al. (2016) who use newspaper coverage frequency to develop an index of economic policy uncertainty. They find that elevated policy uncertainty in the United States and Europe in recent years had material harmful effects on macroeconomic performance. Based on the methods presented by Baker et al., uncertainty indices have also been developed for Sweden (Armelius et al., 2017) and the Netherlands (Kok et al., 2015), for instance.

Regardless of the growing interest in economic uncertainty, our paper is the first well-documented attempt to develop an uncertainty index for Finland. So far, the only uncertainty index developed for Finland is an index constructed by the Finland Chamber of Commerce (2020). Their index is based on the methods developed by Baker et al. but it is a relatively simple and crude version compared to the original index as well as to the index we develop in this paper.[1] As far as we know, our work is also the first news data based uncertainty index that includes machine learning properties.

A measure of uncertainty can be thought as soft data, - in contrast to hard data such as GDP or retail trade volumes. Soft indices do not measure economic outcomes directly but indirectly, hence they can be said to have a looser link to the real state of the economy. They are typically based on surveys, which is the case with one of the most commonly used measures among economic forecasters, PMIs (purchase manager's indices).

Instead of quantitative measures, soft indices often provide qualitative assessments reflecting sentiment or expectations. Yet their value comes from the fact that they become available earlier than hard data and hence they can be used to give a timely assessment of the current state of the economy, as explained by Keeney et al. (2012). This is also one of the motivations in our paper in which we construct a measure of economic uncertainty for

1. The indices can be compared using some simple metrics, such as their (negative) correlation with economic growth (rate of change of GDP). The correlation between our uncertainty index and economic growth in Finland is -0.65, whereas the Finland Chamber of Commerce reported a correlation of -0.17 between their index and economic growth in Finland.

Finland based on Finnish news titles. On this task, we utilize natural language processing (NLP) techniques and in this paper in particular, a transformed naive Bayes text classifier.

Natural language processing (NLP) is a sub-field of artificial intelligence, machine learning and linguistics in which the objective is to develop models that can produce, classify or otherwise process natural languages, that is, languages that humans speak. NLP models are largely based on the syntax of the text and are usually unable to detect the semantics of the sentences, that is, they process the text documents as sets of individual words.

In classification tasks, each word will have its own weight for belonging to each of the predefined categories. In modern applications, statistical NLP models have surpassed rule-based models that consisted of large sets of heuristic logical rules defined by humans. Modern computers have enabled building these heavy statistical models that usually use large sets of text documents as training data. However, it could be argued that even the state-of-the-art NLP models are still behind the abilities of the best solutions in some other sub-areas of artificial intelligence and machine learning.

A certain distinctive factor in NLP is its language-specificity: even though the underlying ideas scale quite well for different languages, the exact models cannot be used for separate languages. Implementing a model on a new language requires at least collecting the corpus on that language and labeling the data. Yet, usually this is not even enough as one has to also consider the special quirks of that given language. In Finnish, these aspects consist of its diverse set of grammatical cases, large number of compound words and varying word orders. In addition, for widely spoken languages there are lots of work already done, e.g., in the form of pretrained models and prelabeled corpora, that can be utilized in the model development. For small languages, such as Finnish, there often is a very limited base to build on.

Contribution of our paper is two-fold. First, we use novel modeling techniques and apply them to the Finnish language. Second, we aim at providing an indicator which gives a timely assessment of the current state of the Finnish economy. The paper is organized as follows. Section 2 describes the methods used in the paper, section 3 discusses the results, and the final section concludes.

2 Methods

The Economy Policy Uncertainty (EPU) index developed by Baker et al. (2016) is based on calculating the frequencies of certain predefined sets of words in the news articles. These words include e.g., ”economy”, ”uncertainty” and ”deficit”. Unlike the example set by the EPU index, our aim is to develop a model that could identify all news titles that imply a straight-forward and clear negative impact on the Finnish economy. After that, our economy uncertainty index would depend on the share of the negative economy news from the whole news feed aggregated on given time periods. Our objective is to utilize machine learning techniques for classifying the negative economic news items, meaning that we let the model learn the significant words instead of selecting the words by hand.

Classification in our model is binary, that is, a piece of economy news is either negative or not. Thus, all negative news are ”equally negative” from our model’s perspective. There are two reasons behind this idea. First, ”negative news” is itself such a subjective concept that we did not want to make the data labeling any more complicated than necessary. In addition, the idea is that every time there are some ”more negative” news, these will generate a large number of some other ”less negative” news and thus raise the value of the uncertainty index.

To clarify this idea, let us consider an example: a title saying that Finland is about to be in a recession, and a title saying that a Finnish company is planning to lay off a few employees, are both considered to be negative economy news from the perspective of the model we propose. While this is true, it is clear that the former is much worse news than the latter and will probably cause more uncertainty for the Finnish economy. However, if there are news titles like the first one - and Finland is about to be in a recession - there most certainly will also be other news about furloughs and bankruptcies etc. that will increase the value of our uncertainty index. That is, we believe that news are significantly dependent on each other and therefore binary classification is sufficient.

2.1 Naive Bayes text classifier

Many of the state-of-the-art NLP models are based on neural networks (e.g., Joulin et al., 2016). However, as training neural networks requires large amounts of data, the model we propose for the classification task is based on a naive Bayes classifier. In machine learning classification applications, the general objective is to develop and train models that can assign samples into the right classes, for instance, to identify news items that belong to the class of news that imply a negative impact on the Finnish economy.

For naive Bayes classification, there has to be a fixed set of classes in which the samples are classified into and a weight parameter vector $\overline{w}_c$ for each predefined class $c \in [c_1, c_2, ..., c_m]$, such that

$$\overline{w}_c = (w_{c,1}, w_{c,2}, ..., w_{c,n}), \tag{1}$$

where n is the number of distinct words in all of the training text documents. That is, for each predefined class c and for each word i there is a weight $w_{c,i}$ that describes how strongly word i is associated to class c.

A naive Bayes classifier considers a text document as an unordered set of words and assigns the text into the class with the maximum posterior probability for the given word weights. These weight estimates $\hat{w}_{c,i}$ for word i in class c are the parameters that the model learns from the data. The classifier is considered to be naive as it assumes that the words are independent and as the prior distribution of the class probabilities is usually considered to be uniform.

The parameter estimate $\hat{w}_{c,i}$ for word i in class c is defined as the probability of word i occurring in class c in the training data. In the most simple case, $\hat{w}_{c,i}$ is defined as

$$\hat{w}_{c,i} = P(i|c) = \frac{N_{c,i}}{N_c}, \tag{2}$$

where $N_{c,i}$ is the count of word i in the documents in class c and N_c is the total number of words in the documents in class c. Following the example set by Rennie et al. (2003), we replace these estimates by smoothed posterior probabilities, but this simple version is a good example for illustrating the core idea of a naive Bayes classifier. The word count smoothing - as well as the other data transformations - are further discussed in section 2.4. A naive Bayes classifier assigns a label $l(d) \in [c_1, c_2, ..., c_m]$ for a text document $d = [d_1, d_2, ..., d_n]$ as:

$$l(d) = \arg\max_c \sum_{i=1}^{n} d_i \log \hat{w}_{c,i}, \tag{3}$$

where n is the number of words in the vocabulary and d_i is the count of word i in document d. It is notable that for a single text document a majority of the word frequencies d_i will usually be zero.

The method of considering a text document as an unordered set of words is called the Bag-of-Words (BoW) representation of the text (e.g., Joulin et al., 2016). BoW does not preserve the information about the word order, which obviously can be a major drawback in many cases, but in classification tasks BoW is considered to be a good and an efficient way to model text documents. For example, if our data consisted of the following documents:

$$\begin{cases} D_1 & = \text{"Mark has three dogs and three cats"} \\ D_2 & = \text{"Lisa likes cats and dogs"} \\ D_3 & = \text{"Tony likes dogs"}, \end{cases}$$

our vocabulary would include the words ('Mark', 'has', 'three', 'dogs', 'and', 'cats, 'Lisa', 'likes', 'Tony') and the BoW representations of the documents would be:

$$\begin{cases} BoW(D_1) & = [1,1,2,1,1,1,0,0,0] \\ BoW(D_2) & = [0,0,0,1,1,1,1,1,0] \\ BoW(D_3) & = [0,0,0,1,0,0,0,1,1]. \end{cases}$$

Our classification model consists of three layers: data preprocessing, transforming the data as described by Rennie et al. (2003) and finally applying the naive Bayes classifier. In the preprocessing phase, we tokenize the texts, remove stopwords, stem tokens and apply some other heuristics that are believed to enhance the model's ability to learn the relevant features. We then transform the data based on the steps proposed by Rennie et al. which they showed to improve a regular naive Bayes classifier. In addition, we apply feature weighting for the transformed data as presented by Forman (2003) and Timonen (2012).

We follow quite closely the example set by Rennie et al. with our classifier, too, but moreover, we apply a few other methods for this type of sentiment classification. Namely, we split the classification process into two subsequent steps where the texts are first classified into "Economy News" and "Other News", and only the economy news are further classified with respect to their sentiments. We discuss this approach more in section 2.2. The resulting classification model is referred to as the Feature-weighted and Transformed Weight-normalized Naive Bayes (FTWNB) classifier. In the following sections we present the data and our model.

2.2 Data collection and annotation process

Supervised machine learning models always require annotated training data, that is, data that has been labeled beforehand. In our case we had to carry out the annotation process ourselves. Determining a sentiment of a text is always somewhat subjective. There has to be clear guidelines for how the data is annotated by humans so that the model can be able to learn the right way to classify it.

The data for both training and testing the model has been collected from the website of the Finland's national public broadcasting company Yle. Each row of the data consists of the title of the news item and a "topic tag" attached to it. 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 noticed that usually the title includes almost all of the relevant information about the sentiment of the news item. Frankly, we rather need to remove words from the titles than have more information. In addition, processing only the titles makes classification done by humans and the model itself much faster.

First, we read some titles and then decided on the rules which we were going to follow in the annotation. Naturally, we discussed about these rules and updated them when necessary throughout the whole annotation process. There are three data features, and each feature is labeled with a binary classification: 1) whether the news item is related to Finland, 2) whether it is about (Finnish or any other) economy and 3) whether its sentiment is negative (only considered if the title is classified to be about economy). The negative sentiment means that the title must include something that will have a straight-forward and obvious negative impact on the national economy, e.g., a furlough, a bankruptcy, a strike or an economic downturn.

The reason why we chose to use binary labels in the sentiment annotation rather than some label scale (e.g., to label the economy news into ”very negative”, ”negative”, ”neutral” and so on) is that the concept of ”bad news” is already so subjective and difficult to compress into well-defined rules that we did not want to make the labeling system any more complex than necessary. Furthermore, considering the final economy uncertainty index, we believe that a binary classification is sufficient as the number of negative economy news titles will be aggregated on daily or monthly basis (depending on the length of the data set) and as the news titles are heavily dependent on each other.

If we had labeled the data into ”Domestic Negative Economy News” and ”Other News” - the former one being the only category that we are truly interested in - a large part of valuable data for learning would have been lost. Using the three data features (”Domestic”, ”Economy” and ”Negative”) separately enables the model to have more training data, as the model can, for example, learn significant vocabulary associated with economy from news titles that are not necessarily related to Finland. This also applies to the manner in which the news titles are classified, as the classifier assigns its predictions with respect to the three features separately. As a matter of fact, the model consists of three separate classifiers for each feature.

In addition to the benefits with respect to the model's learning ability, splitting the classification process into consecutive steps is also necessary due to the imbalanced nature of the data as the share of negative economy news from the data is only about five percent. After splitting the classification, in the first step the share of economy news is about 18%, and in the second step the share of negative news titles from the economy news is about 28%, so now the classification steps are much closer to being balanced.

Another effect of having imbalanced data is that the goodness of the classifier cannot be evaluated by accuracy: if positive samples (the negative economy news titles are positive samples in this case as they are the ones we are trying to detect) make up 5% of the data, classifying all samples as negative would yield an accuracy of 95%. Fortunately, there are many alternative scoring metrics for accuracy. In this paper we use the balanced accuracy score (BA) which is the average of recall on positive samples and recall on negative samples.

2.3 Preparing the data

Splitting the text documents into tokens - that is, into separate words - and removing stopwords (words that do not hold any semantic information, such as ”about”, ”here” and ”so”) are quite standard procedures in statistical natural language processing. Text stemming means reducing words into their stem forms, whilst lemmatazing means replacing the words with their dictionary forms. For example, the stem of a word ”hiring” is ”hir” but its lemma is ”to hire”. Typically, NLP models use stemming rather than lemmatization, as finding the stem of a word is more effortless and the stem form can actually preserve more of the word’s information.

There are some existing methods for text stemming in English, but in Finnish this is a little more complicated. There are two main reasons for this. First, comparing to English - the most studied language in the field of NLP - Finnish is a small language and there are very few existing methods and applications for lemmatazing or stemming it. The other reason is that in Finnish there is a very complex system of word affixes and setting a word back to its lemma may be really difficult.

Moreover, in Finnish there are lots of compound words that do not have a space between them and thus cannot be separated in traditional tokenization. For example, ”economic crisis” is in Finnish ”talouskriisi”, where ”talous” means ”economic” and ”kriisi” means ”crisis”. We use a small dictionary of relevant words (such as ”kriisi”) and split every compound word that contains any of these words into two tokens. We stem tokens by taking only a constant number of characters from the start of the token and discarding the rest. This is a simple - yet very efficient and practical - way of removing the word affixes. However, as this number is fixed for the whole data, some longer words may lose more information and some shorter words may still preserve all or some parts of their affixes.

The classifier does not ”know” anything about the relations of words or how close the meanings of different words are to each other: from the classifier’s perspective two words either are equal or then they are not. Stemming improves the model’s ability to learn, as a larger part of the words actually look exactly the same and therefore account for the same index in the Bag-of-Word (BoW) representation. For example, the Finnish words ”irtisanominen” (termination) and ”irtisanoa” (to terminate) would look exactly the same if the number of characters used as a stemming length was eight or smaller. The number we use as the stemming length is seven. Obviously, the selection of this number as well as this procedure itself are heuristics, but we found this stemming length to enhance the model’s learning ability for the majority of the vocabulary.

We also apply some other heuristics in order to improve the model's ability to learn relevant features and try to avoid learning irrelevant or biased properties. The heuristics consist of lists of words that are replaced by some tags, for example, names of Finnish companies are replaced by a tag "Finnish Firm". The reason for this is to prevent the model from learning that some specific company name would have a high correlation with the negative sentiment of the title. As the data set that the model uses for learning is relatively short from the economy's perspective (some half a year of news titles), this correlation may be true in the training set, but we do not want that the model learns this kind of "wrong" correlation. Other replaced lists of words consist of foreign countries, Finnish cities and names of Finnish political parties.

2.4 Data transformations and feature weighting

A naive Bayes classifier is based on counting the occurrences of different tokens - that is, different words - in different categories and learning the weights for each token with respect to each class. A naive Bayes classifier then assigns a sample into the class with the highest posterior probability given the feature vector of the sample. In our case, these feature vectors are transformed and weighted Bag-of-Words (BoW) representations of the news titles.

We build our model by following closely the implementation of Transformed Weight-normalized Complement Naive Bayes (TWCNB) classifier (Rennie et al., 2003), which fixes many biases and weaknesses of a regular naive Bayes classifier that are caused by imbalanced data and dependencies between the words that a naive Bayes does not take into consideration as it assumes every word in a sentence to be independent.

The complement feature is the only property in the TWCNB that we do not utilize with our model. Rennie et al. introduced this feature for dealing with skewed data in multi-class classification: instead of training the classifier with data in class c, they train it with data in every other class except c. The classifier then assigns a document into the class in which it the least probably does not belong. They show that this method causes the classifier to be more effective and less biased. However, we have binary data and therefore the complement feature is unnecessary for our model. The following data transformations are described as presented by Rennie et al. (2003).

First, the data is transformed with a commonly used Term Frequency (TF) - Inverse Document Frequency (IDF) transformation. The TF-transformation maps the number $d_{i,j}$ of token i in document j with a concave function. The idea is that - since the words in a text are obviously dependent on each other - every new occurrence of a given word in a document is less significant than the previous one. We use the log-function for the TF-transformation:

$$\mathrm{TF}(d_{i,j}) = \log(d_{i,j} + 1), \tag{4}$$

where we add one to the count $d_{i,j}$ to avoid problems in the case where $d_{i,j} = 0$. Frankly, in our case - where the documents are news titles and thus consist of only about 10-15 words on average - usually each word in the document appears only once. This is called the *TF=1 challenge* (Timonen, 2012) and it means that in our data only for about 10% of the word counts this transformation actually transforms data. Nevertheless, we noticed that the TF-transformation still enhances the model’s ability to learn and decided to include it in the data transformations. Furthermore, the *TF=1 challenge* will be addressed later in this section by feature weighting.

The other part of the TF-IDF -transformation is the Inverse Document Frequency transformation. The IDF-transformation reduces the weights of tokens that appear in many documents. This is necessary, as frequent tokens cannot be used for determining the classes of the documents, but as the training sets always consist of some random variation there can otherwise be some unwanted correlations. This is done as

$$\mathrm{IDF}(d_{i,j}) = d_{i,j} \log \frac{N_d}{\sum_k \delta_{i,k}}, \tag{5}$$

where N_d is the number of documents in the training data and in the denominator $\delta_{i,k}$ is 1 if token i appears in document k and 0 otherwise.

A quite similar bias to the one with the frequent words, is that longer texts include more words and thus the significance of a certain word in a longer text is smaller than in a shorter text. In order to restrain this bias, the token counts are normalized by the lengths of the texts where they feature in:

$$\mathrm{NORM}(d_{i,j}) = \frac{d_{i,j}}{\sqrt{\sum_k (d_{k,j})^2}}. \tag{6}$$

By applying the transformations presented in the equations 4, 5 and 6, we get the transformed word occurrence counts that are used for estimating the token weights. That is, the transformed count $\widetilde{d}_{i,j}$ of word i in document j is defined as

$$\widetilde{d}_{i,j} = \mathrm{NORM}(\mathrm{IDF}(\mathrm{TF}(d_{i,j}))). \tag{7}$$

Instead of using empirical probabilities for word occurrences as described earlier in section 2.1, we use Laplace smoothing which is a method for smoothing categorical data. One of the reasons for smoothing is to prevent any of the probabilities from going to zero. Empirical probability of token i in class c is

$$P(i|c) = \frac{N_{c,i}}{N_c}, \tag{8}$$

where $N_{c,i}$ is the number of occurrences of token i in class c and N_c is the total number of tokens in class c. Instead, we add a pseudo-count α to each token in order to create the smoothed posterior probabilities for the token occurrences. That is, the weight estimate $\hat{w}_{c,i}$ of token i in class c is defined as

$$\hat{w}_{c,i} = \frac{\sum_{l(k)=c} \widetilde{d}_{i,k} + \alpha}{\sum_{l(k)=c} \left(\sum_t \widetilde{d}_{t,k} + \alpha \right)}, \tag{9}$$

where $l(k) = c$ indicates that the label of document k is c. That is, we take a sum over all documents that are labeled to be in class c. α is the pseudo-count which is usually heuristically chosen to be $\alpha = 1$. Our experiments showed as well that $\alpha = 1$ is an efficient choice and it produced good results. That is, we follow this common manner and set the pseudo-count to be equal to one.

Finally, the weights are normalized as

$$w_{c,i} = \frac{\log(\hat{w}_{c,i})}{\sum_k \log(\hat{w}_{c,k})}. \tag{10}$$

This is necessary as the imbalanced nature of the data can cause either one of the classes to have higher dependencies than the other and cause the weights to shift closer to that class as shown by Rennie et al. (2003).

As mentioned earlier, in our data each word in a given title usually appears only once. The frequency of a word in the text is an important feature in a naive Bayes classifier, but when the documents are short - as they are in our data - the frequencies do not hold too much information. This is referred to as the *TF=1 challenge* (Timonen, 2012). One solution to the *TF=1 challenge* is feature weighting and selection (Forman, 2003 and Timonen, 2012). The idea is to utilize some other methods than solely term frequency for determining which tokens are more significant than the others. The objective is to either weight the tokens by some method or to discard the tokens that do not reach a given threshold level in significance.

We use Bi-Normal Separation (BNS) as presented by Forman (2003) for feature weighting. The BNS-weight $\phi_{c,i}$ for token i in class c is defined as

$$\phi_{c,i} = |F^{-1}\left(\frac{D_{c,i}}{D_i}\right) - F^{-1}\left(\frac{D_{c,\neg i}}{D_{\neg i}}\right)|, \tag{11}$$

where F^{-1} is the inverse Normal cumulative distribution function. $D_{c,i}$ is the count of documents that include token i in class c, $D_{c,\neg i}$ is the count of documents without token i in class c, D_i is the count of all documents including token i and $D_{\neg i}$ is the count of all documents without token i. In order to avoid problems of taking the inverse Normal of zero or one, both distributions are limited into a range of [0.0001, 0.9999]. The idea of BNS is to measure the difference between the distributions of documents in that class with and without a certain token and consequently infer the significance of that given token with respect to that class.

2.5 The uncertainty index

Our model consists of three separate classifiers that are implemented as described in the earlier sections. The classified binary features are: ”Domestic News”, ”Economy News” and ”Negative Economy News”. A positively classified sample requires all of the three features to be positive. The classifiers are trained independently, however, for training the ”Negative Economy News” classifier we use only data labeled to be economy news in order to enhance the classifier's ability to learn significant features with respect to the negative sentiment.

The classifier developed here is referred to as the Feature-weighted and Transformed Weight-normalized Naive Bayes (FTWNB) classifier, differing slightly from the Transformed Weight-normalized Complement Naive Bayes (TWCNB) classifier developed by Rennie et al. (2003) and augmented with feature weighting (Forman, 2003 and Timonen, 2012). The classifier assigns a label $l^*(d) \in (-1, 1)$ to the document $d = (d_1, ..., d_n)$ with respect to each binary feature according to:

$$l^*(d) = \arg\max_c \left[\sum_{i=1}^{n} d_i w_{c,i} \phi_{c,i}\right], \tag{12}$$

where d_i is the count of token i in document d, $w_{c,i}$ is the weight learned by the model for token i in class c and $\phi_{c,i}$ is the BNS-weight for token i in class c.

In Figure 1 is shown an example histogram of how one of the classifiers separates the news titles. The values on the x-axis correspond to the posterior probabilities with respect to each title such that class c_D is "Domestic News": $P(c_D) = \sum_{i=1}^{n} d_i w_{c_D,i} \phi_{c_D,i}$. The dashed line indicates the probability of 0.5. The samples on right are classified to be positive and on left to be negative samples. The ground truth labels are represented with colors. The histogram shows that the distributions overlap a little, which causes some inevitable errors in the classification.

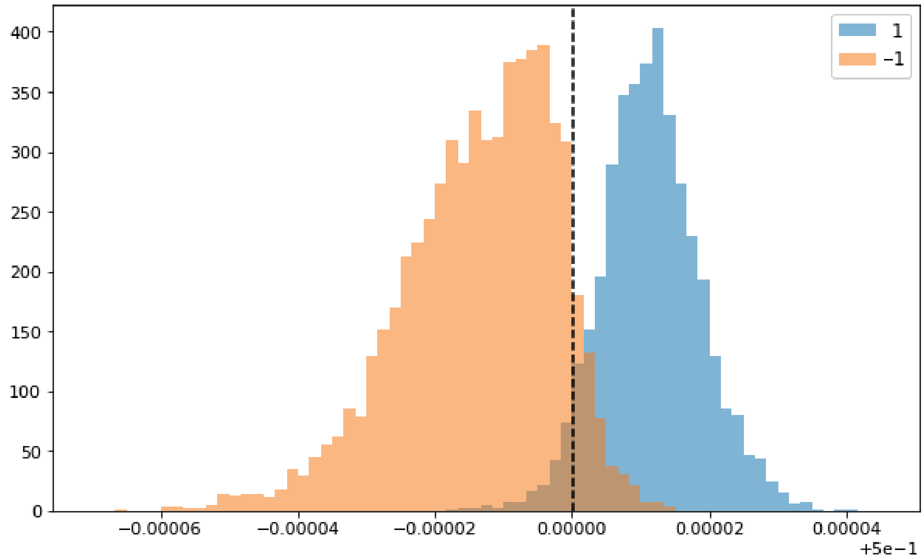


Figure 1: Sample separation for the "Domestic News" classifier in the training data. 1 = domestic news title, -1 = other.

As stated earlier, our idea is that the share of the negative economy news titles within a time period would describe the economic uncertainty in that given period. However, it is worth noting that a single value of the uncertainty index is not informative, but rather the interesting phenomenon is the development of the index over time. We define the uncertainty index in the period i:

$$\mathrm{UNC}(i) = \left(\frac{N_{i,\text{Negative}}}{N_{i,\text{Total}}} \right)^2, \tag{13}$$

where $N_{i,\text{Negative}}$ and $N_{i,\text{Total}}$ are the numbers of negative domestic economy news titles as classified by the model and the total number of titles, respectively, in the period i. That is, we believe that the process behind the uncertainty is nonlinear. So for example, if the share of negative news titles doubles the value of our uncertainty index quadruples.

3 Results

Evaluation of the results of our model is two-fold. First, we have to consider whether our text classifier produces satisfying results in the test data set. After that, we need to find a way to validate the results of our uncertainty index. The second part is far more tricky, as the feature we are measuring, that is, the economic uncertainty, is such a subjective phenomenon itself.

The data used for training as well as testing the model is collected from the website of the Finnish broadcasting company Yle. The complete data set covers news titles from the start of the year 2018 to the September of 2020. The annotated data set consists of some 12 000 news titles of which about 10 000 samples are used for training the model and 2 000 samples for testing it. Following a standard convention for classification with chronological data, the data has been divided into training and test sets also in chronological manner. In many machine learning applications, this split is done randomly, but as the model is trying to classify future samples it would gain unnecessary advantage if the training set was selected randomly.

As stated earlier, we do not use accuracy as a scoring metric for our classifier due to the biases it causes for an imbalanced data set. Instead, we evaluate the goodness of the results with the balanced accuracy score (BA), which is defined as

$$\mathrm{BA} = \frac{\frac{\mathrm{TP}}{\mathrm{P}} + \frac{\mathrm{TN}}{\mathrm{N}}}{2}, \tag{14}$$

where TP is the number of true positives, TN is the number of true negatives and P and N are the numbers of positive and negative samples in the data. That is, BA score is the average of recall on positive samples and recall on negative samples.

The BA scores (as well as accuracy scores) of the three classifiers are shown in Table 1. All three classifiers fit well in the training data (in-sample). Scores in the test data (out-of-sample) are naturally a little lower, although, they are decent as well. The *Domestic news* -classifier achieves the best result in both sets - probably due to the fact that this classification task is the closest of being balanced. Classifying the negative sentiment of a news title is clearly the most difficult task as the naive Bayes and the BoW approaches cannot detect the sentiment of a sentence, but only the sentiments of separate words.

Classifier	**In-sample**	**Out-of-sample**
Domestic news	0.94 (0.94)	0.84 (0.84)
Economy news	0.89 (0.95)	0.76 (0.88)
Neg. economy news	0.91 (0.97)	0.72 (0.92)

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

A good rule of thumb is to test models as they were supposed to be used. That is, we should compare the aggregated time series produced by the classifiers and our annotation. In addition, in this particular occasion we are not necessarily even interested in the absolute levels in these time series but rather the correlation between them. Namely, our uncertainty index depends on the share of the negative news titles, but - as stated earlier - a single value of the index is not informative; we are interested in the changes in the index over time. That is, instead of accuracy we can use correlation for evaluating the results in the aggregated test data set. The correlation coefficients between the daily aggregated time series of the classification results and the ground truth labels are shown in Table 2.

Classifier	**Daily agg. correlation**
Domestic news	0.99
Economy news	0.93
Negative economy news	0.86
Dom. neg. economy news	0.83

Table 2: Correlation coefficients between the daily aggregated classifications and ground truth labels.

As shown, the results correlate significantly with the ground truth. Two of the aggregated time series are also shown in Figures 2 and 3. It is worth noting that the accuracy of the third classifier relies on the second classifier as in order to classify a title to be "Negative economy news" it has to be first labeled under "Economy news". Naturally, for the same reasons, the the fourth classifier relies on both the first and the second classifier.

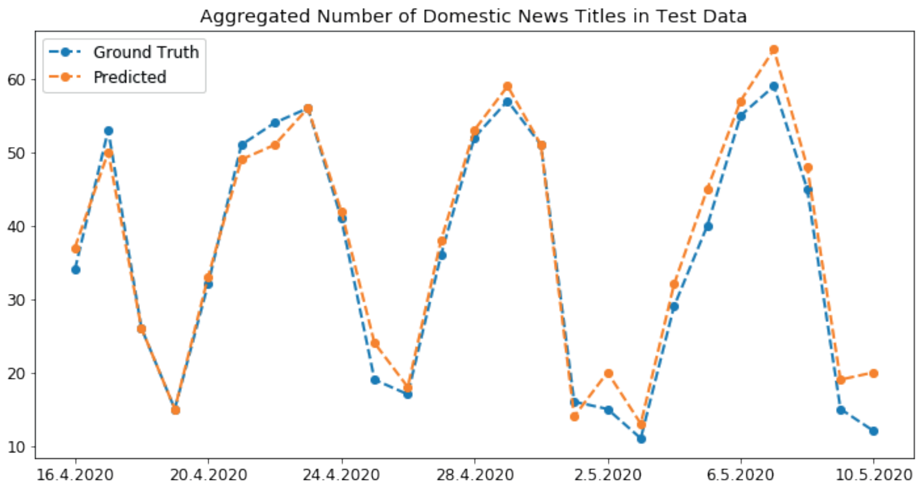


Figure 2: The correlation between the classification results and the ground truth is 0.99.

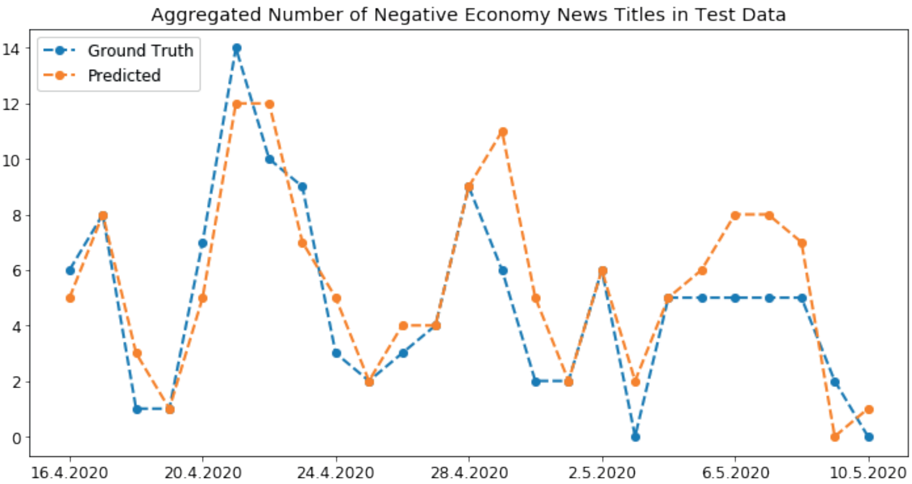


Figure 3: The correlation between the classification results and the ground truth is 0.86.

So how to validate the uncertainty index we have developed? As stated before, the numerical values of our index do not represent itself any particular economic phenomenon. However, as it is typically the case with soft indices, it is the change of the index value that aims at providing useful information on some economic development of interest. In our case the interest is to observe the current state of the economy, at best earlier than what is observed using other macro data.

To assess the ability of our index to achieve this aim, it can, for instance, be compared to the consumer confidence index measured by Statistics Finland as shown in Figure 4. Their index is one of the most commonly used soft indices to consider the current state of the Finnish economy.

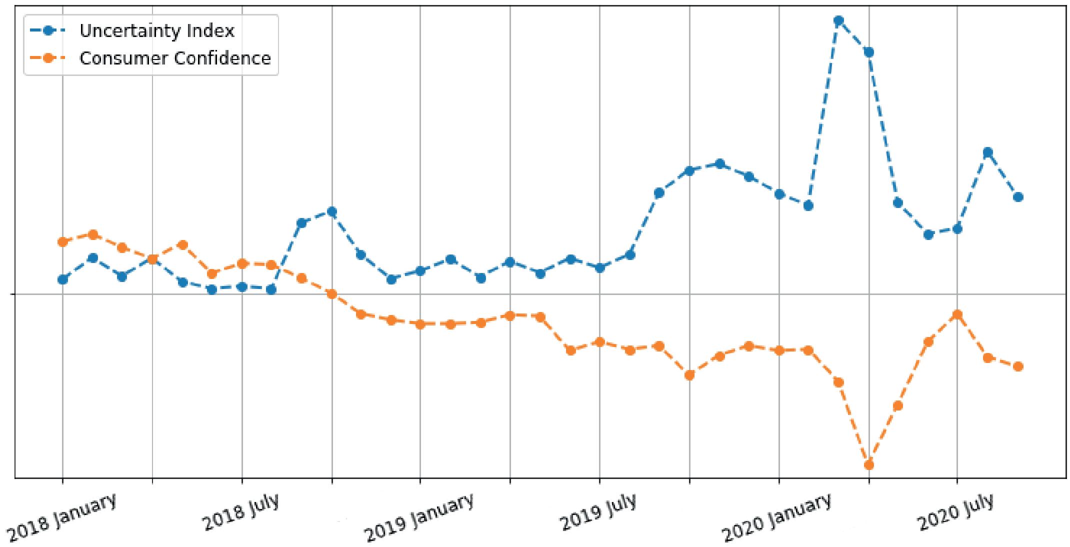


Figure 4: Our uncertainty index and the consumer confidence index by Statistics Finland have significant negative correlation. For example, COVID-19 generates clear peaks in both indices.

According to the results shown in Figure 4, our index correlates negatively with the consumer confidence index as calculated by Statistics Finland. The correlation value between the two series is -0.73. This is important as the uncertainty index is based on the idea of counting news titles whose content has a straight-forward and obvious negative impact on the national economy. Thus, it would counter-intuitive and disappointing if there was no clear negative correlation between these two series.

Of course, one can ask why the uncertainty index starts rising in the autumn of 2018 and more markedly again, in the autumn of 2019, without any such a clear change in the consumer sentiment index. Explanation for the latter could be Brexit- or trade war (mostly between the US and China) related uncertainty whose effects could be reflected in the negative news titles but not in the consumer sentiment. On the other hand, according to the revised national accounts data of Statistics Finland, the Finnish economy started contracting already in the third quarter of 2019, so at least our index seems to be in line with that development.

However, the most remarkable result is that the uncertainty index seems to lead the consumer confidence index, by one month. The correlation value of the lead of the uncertainty index and the consumer confidence index is -0.78. This result is illustrated in Figure 5 where the uncertainty index is shown alongside with the additive inverse of the consumer confidence index by Statistics Finland. Our index for instance recognizes the COVID-19 crisis one month earlier than the consumer confidence index. This also gives support using the uncertainty index developed to assess the current, - and even the future -, state of the Finnish economy.

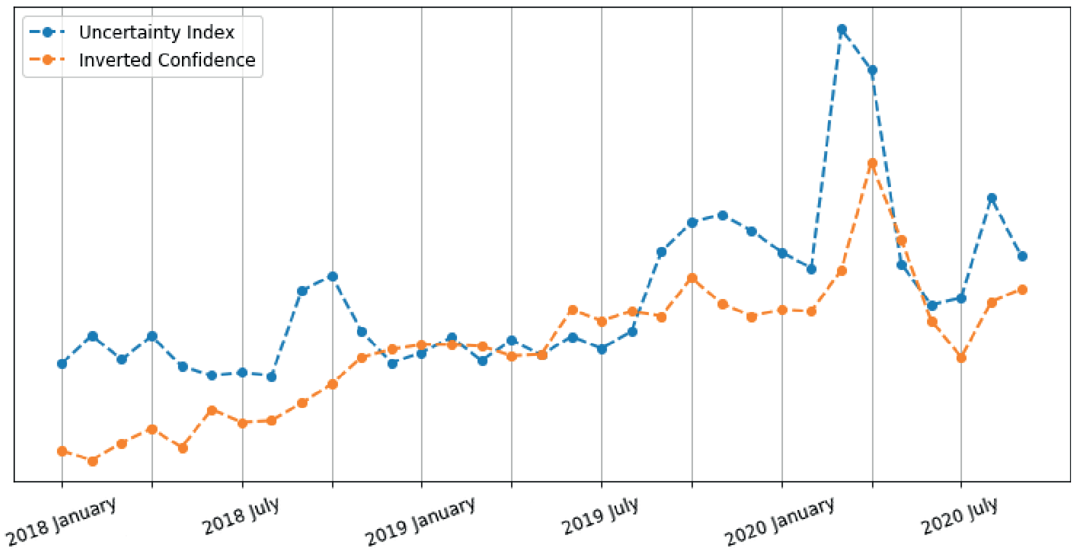


Figure 5: Our uncertainty index and the "inverted consumer confidence index". Leading correlation between the indices is 0.78.

It is also useful to analyze the relationship of our uncertainty index with other macro variables in a simple VAR model. This can be seen as validation of the indicator we have developed, but also, using this tool we can examine how other macro variables respond to increases in uncertainty.

The uncertainty index has been calculated on a monthly basis. Keeping this in mind, 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 (months) for each of the variables. Order of the variables is assumed to be the following: uncertainty, output, employment. Thus, we assume that uncertainty leads to changes in production which then affect employment with a lag.

Even though the uncertainty indicator is built in a manner that it actually consists mostly of negative news about employment, the assumed order of the variables still makes sense as in many cases it takes time, a month or two easily, for negative employment news to materialize in the official employment figure as measured by Statistics Finland. For instance, (temporarily) furloughed persons are still counted as employed for as long as three months after becoming furloughed.

According to standard impulse response functions calculated from the VAR model, an increase in uncertainty is associated with a decline in trend indicator of output as shown in Figure 6, and decline in employment, shown in Figure 7 (times series and all the impulse responses of the model variables are shown in Appendices 1 and 2). The peak negative effect in output comes in three months. The negative effect on employment follows with a slightly longer lag, peaking in six months. The employment effect is also more persistent in time: while the output effect abates close to zero in ten months, the response of employment to uncertainty stays negative even then. This exercise, while based on a limited amount of data and a simple VAR model, supports the use of uncertainty index for further policy analysis.

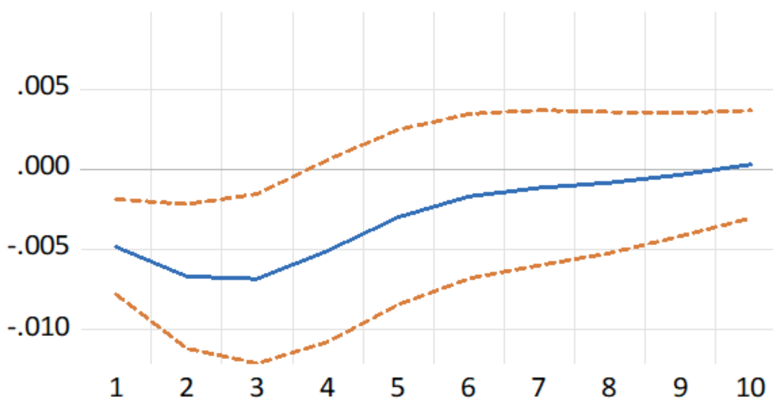


Figure 6: Response of (log) output to uncertainty shock (95% confidence bands).

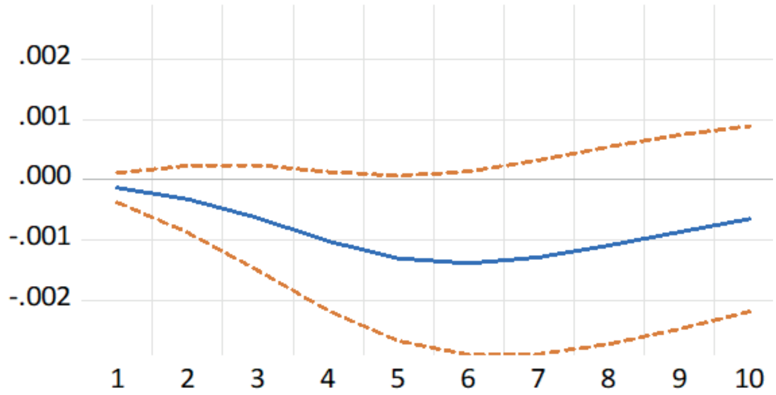


Figure 7: Response of (log) employment to uncertainty shock (95% confidence bands).

4 Conclusions

In this paper, we introduced a method for modeling the economic uncertainty, namely, we proposed that the share of negative economy news titles of the total news feed would describe the level of uncertainty. In order to automate the classification, we developed and trained a transformed naive Bayes text classifier. The model developed is referred to as the Feature-weighted and Transformed Weight-normalized Naive Bayes (FTWNB) classifier. Using this approach, we built three separate classifiers for different classification tasks: whether the title is domestic news, whether it is economy news and whether it has straight negative impact on economy (only considered if the title was classified to be economy news).

We then applied the model on Finnish news titles collected from the website of the Finnish broadcasting company Yle and evaluated the results. We stated that two issues need to be considered when evaluating the results: whether the classifiers themselves produce sufficient results and then whether the uncertainty index is in line with our assessment of the development of economic uncertainty.

First part of the evaluation is rather straight-forward, although we had to consider the unbalanced nature of the data set. Namely, we used a balanced accuracy score instead of traditional accuracy measure. The results of our classifiers are relatively good considering that our training data set is fairly short. Especially the "Domestic News" classifier produced very accurate results.

In the second part of the evaluation, that is, to provide some validation for the proposed uncertainty index, we compared the uncertainty index with the consumer sentiment index as calculated by Statistics Finland. We found a strong negative correlation between these two series. Yet the most significant result was that the uncertainty index seems to lead the consumer sentiment by one month, hence it is possible to forecast reasonably well the next month’s change in the consumer sentiment using the uncertainty index. This is remarkable as the consumer sentiment index is currently one of the best methods for assessing the current state of uncertainty in the national economy and our index could provide information even earlier than it.

Finally, we built a simple VAR model to assess usefulness of the uncertainty index in the policy analysis. This can also be seen as validation for the index we developed. While based on a limited amount of data and a simple VAR model, the results are intuitive and hence support the use of uncertainty index for further policy analysis.

Yet we understand that there is room for improvement with our uncertainty index. The data we used in construction of the index is rather short as it covers less than three years, thus it would be interesting to see how the index would behave with a longer data set. Also, there are naturally possibilities for improvement with the model we use in the text classification. Having said that, together with a contribution to literature by applying natural language processing (NLP) techniques to the Finnish language, we think that our index is helpful in providing a timely assessment of the current state of the Finnish economy.

References

[1] Armelius, H., Hull, I., Stenbacka Köhler, H. (2017), The Timing of Uncertainty Shocks in a Small Open Economy, Economics Letters, 155 (June), 31-34.

[2] Baker, S. R., Bloom, N., and Davis, S. J. (2016), Measuring economic policy uncertainty, The quarterly journal of economics, 131(4), 1593-1636.

[3] Bernanke, B. S. (1983), Irreversibility, Uncertainty and Cyclical Investment, Quarterly Journal of Economics, 97, no. 1 (February), 85–106.

[4] Finland Chamber of Commerce (2020), Kauppakamarin talouskatsaus, Heinäkuu 2020.

[5] Gilchrist, S., Sim, J. W., and Zakrajsek, E. (2014), Uncertainty, Financial Frictions, and Investment Dynamics, NBER Working Paper 20038.

[6] Keeney, M., Kennedy, B., and Liebermann, J. (2012), The value of hard and soft data for short-term forecasting of GDP, Economic Letter Series, Vol 2012, No. 11, Central Bank of Ireland.

[7] Kok, S., Kroese, L., and Parlevliet, J. (2015), Beleidsonzekerheid in Nederland in Economisch Statistische Berichten, 100 (4715), 464-467, August 2015.

[8] Panousi, V. and Papanikolaou, D. (2012), Investment, idiosyncratic risk and ownership, Journal of Finance, 67, no. 3 (June), 1113-1148.

[9] Pastor, L. and Veronesi, P. (2013), Political Uncertainty and Risk Premia, Journal of Financial Economics, 110, no. 3 (December), 520-545.

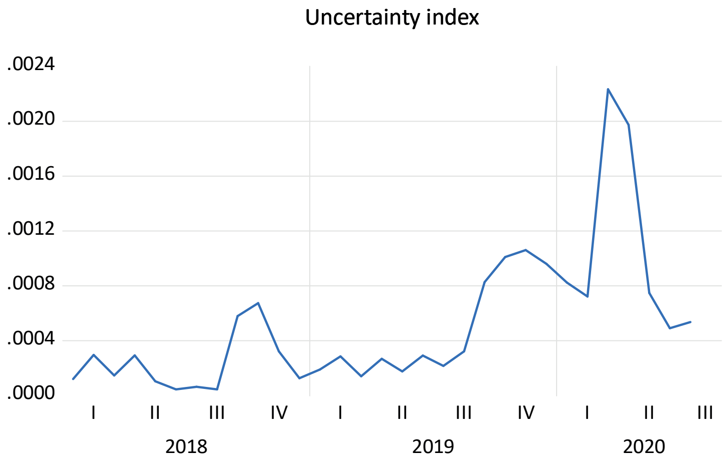
[10] Joulin, A., Grave, E., Bojanowski, P., and Mikolov, T. (2016). Bag of tricks for efficient text classification. arXiv preprint arXiv:1607.01759.

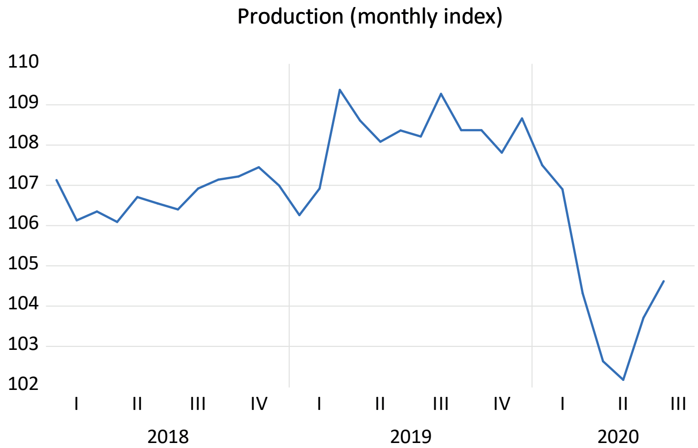
[11] Rennie, J. D., Shih, L., Teevan, J., and Karger, D. R. (2003). Tackling the poor assumptions of naive bayes text classifiers. Proceedings of the 20th international conference on machine learning (ICML-03), 616-623.

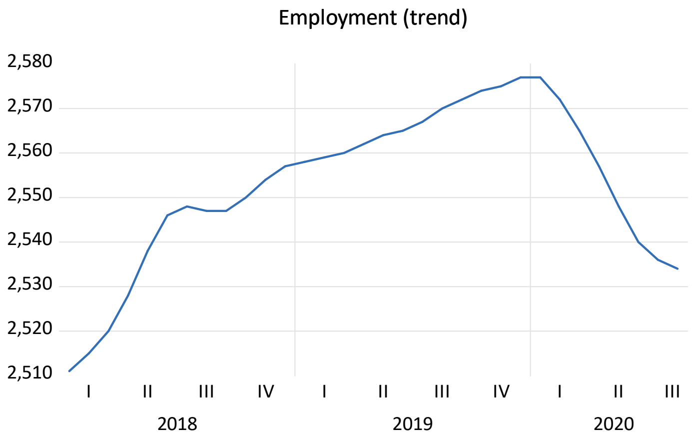
[12] Forman, G. (2003). An extensive empirical study of feature selection metrics for text classification. Journal of machine learning research, 3(Mar), 1289-1305.

[13] Timonen, M. (2012). Categorization of Very Short Documents. KDIR, 5-16.

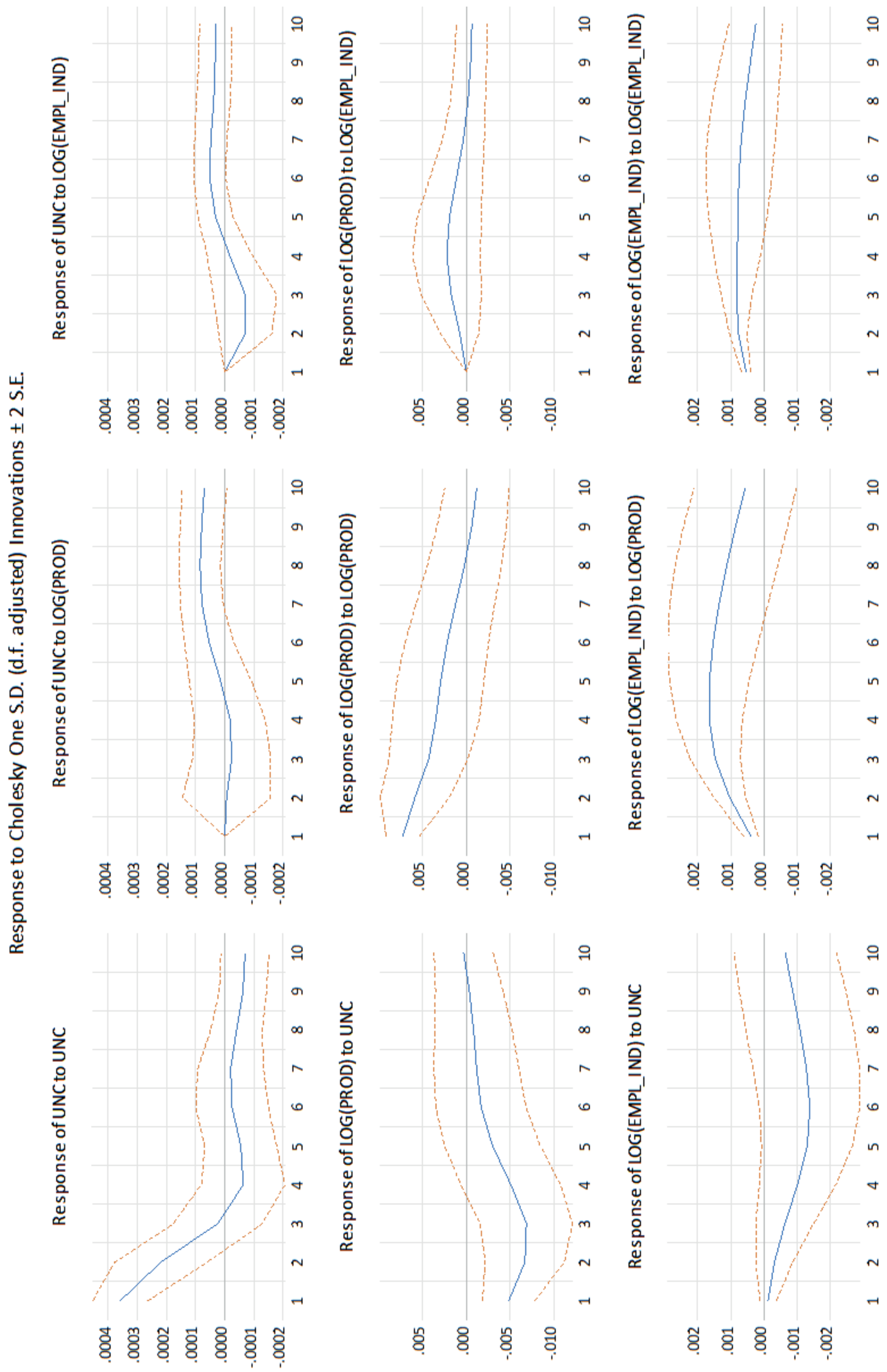
Appendix 1: Uncertainty index, Production, and Employment (trend)







Appendix 2: VAR model responses to Cholesky one S.D. innovations



ETLA

Elinkeinoelämän tutkimuslaitos

ETLA Economic Research

ISSN-L 2323-2420
ISSN 2323-2420 (print)
ISSN 2323-2439 (pdf)

Publisher: Taloustieto Oy

Tel. +358-9-609 900
www.etla.fi
firstname.lastname@etla.fi

Arkadiankatu 23 B
FIN-00100 Helsinki