International Journal of Forecasting 34 (2018) 355–365

Contents lists available at ScienceDirect

International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast

Belgian economic policy uncertainty index: Improvement through text mining

Ellen Tobback [a,*], Hans Naudts [b], Walter Daelemans [c], Enric Junqué de Fortuny [d], David Martens [a]

[a] Department of Engineering Management, University of Antwerp, Belgium
[b] DG Economics and Finance, European Commission, Belgium
[c] Department of Linguistics, University of Antwerp, Belgium
[d] Department of Marketing Management, Rotterdam School of Management, Netherlands

ARTICLE INFO

Keywords:
Economic policy
Uncertainty
Text mining
Forecasting

ABSTRACT

Recently, the literature has measured economic policy uncertainty using news references, resulting in the frequently-mentioned 'Economic Policy Uncertainty index' (EPU). In the original setup, a news article is assumed to address policy uncertainty if it contains certain predefined keywords. We argue that the original setup is prone to measurement error, and propose an alternative methodology using text mining techniques. We compare the original method to modality annotation and support vector machines (SVM) classification in order to create an EPU index for Belgium. Validation on an out-of-sample test set speaks in favour of using an SVM classification model for constructing a news-based policy uncertainty indicator. The indicators are then used to forecast 10 macroeconomic and financial variables. The original method of measuring EPU does not have predictive power for any of these 10 variables. The SVM indicator has a higher predictive power and, notably, changes in the level of policy uncertainty during tumultuous periods of high uncertainty and risk can predict changes in the sovereign bond yield and spread, the credit default swap spread, and consumer confidence.
© 2016 International Institute of Forecasters. Published by Elsevier B.V. All rights reserved.

1. Introduction

International institutions indicate that economic policy uncertainty rose to historically high levels after the 2007–2009 recession because of the uncertainty about tax, spending, regulatory, and monetary policies (Balta, Fernández, & Ruscher, 2013; IMF, 2012). This uncertainty has slowed the recovery from the recession by causing businesses and households to cut back or postpone investment, hiring and consumption. For example, in't Veld (2013) models the impact on the GDP of fiscal consolidation under a range of different uncertainty and learning scenarios. In a scenario of uncertainty regarding the credibility of the fiscal consolidation, the short-term negative impact on GDP is up to three times higher than in a scenario of immediate credibility. Balta et al. (2013) find that uncertainty has a significant effect on both investment and consumption in the euro area, with this effect on activity increasing since the crisis and extending beyond traditional cyclical effects. Economic research has come up with several ways of constructing uncertainty measures based on the stock market volatility (Bloom, 2009; Kose & Terrones, 2012),

* Corresponding author.
E-mail address: ellen.tobback@uantwerp.be (E. Tobback).

http://dx.doi.org/10.1016/j.ijforecast.2016.08.006
0169-2070/© 2016 International Institute of Forecasters. Published by Elsevier B.V. All rights reserved.

the dispersion in forecasts by professional forecasters or in the expectations of consumers or producers (Bachmann, Elstner, & Sims, 2010), or the prevalence of terms such as 'economic uncertainty' in the media (Baker, Bloom, & Davis, 2015). This paper focuses on the latter methodology and contributes to the economic literature by using text mining methods to construct uncertainty indicators. This methodology allows us to identify the main factors with which uncertainty is associated.

Recently, Baker et al. (2015) constructed an economic policy uncertainty index (EPU) as a proxy for movements in policy-related economic uncertainty over time. This index represents the frequency of newspaper references to EPU. The authors find that their index peaks around important events such as 9/11 and the bankruptcy of Lehman Brothers. The index has given rise to numerous studies concerning the influence of economic uncertainty on macroeconomic indicators. However, notwithstanding its widespread use and acceptance, there remain some important issues regarding the construction of the index. The method is likely to be prone to both type I and type II errors. First of all, every article that meets the search criteria is added to the EPU index, including articles in which the author states that there is no policy uncertainty. Secondly, articles that address policy uncertainty without explicitly using the word 'uncertain' are not added to the EPU index. Thus, the method suggested by Baker et al. (2015) can cause a high rate of both false positives and negatives.

This paper attempts to improve on this methodology by using text mining to solve its main issues. Text mining is the process of deriving high quality information from text documents using techniques from data mining, statistics, information retrieval, machine learning and computational linguistics (Weiss, Indurkhya, & Zhang, 2010). We apply two different text mining algorithms to a data set of approximately 210,000 articles: modality annotation and a support vector machines (SVM) classification model. The former counts the use of words that express uncertainty, the latter is a trained classifier that predicts whether an article addresses economic policy uncertainty. Following Baker et al. (2015), we define economic policy uncertainty as both uncertainty as to who will make what policy decision when, and uncertainty about the effects of past/present/future policy decisions. We limit Belgian economic policy uncertainty to uncertainty at the Belgian and euro area levels. It is commonly accepted that economic spillovers in the euro area are more important, given the shared currency and the closer interlinkages between euro area member states.

This paper's contributions are three-fold. First and most obviously, we try to improve the existing EPU index by solving some of its most important issues. Second, we demonstrate how data mining techniques, and more specifically text mining techniques, can be used to solve a policy-related problem. In this particular case, the policy-related problem is finding a measure of the economic policy uncertainty. We assess policy uncertainty by detecting patterns automatically in a total of 210,000 news articles, using modality annotation and text classification. By doing so, we add to economic theory; for example, by investigating the coefficients of the trained SVM model, we can see which words in the news articles are related to policy uncertainty most frequently. Moreover, we show that our constructed policy indicator improves the forecasts of the Belgian sovereign bond yield and spread, credit default swap spread, and consumer confidence. Finally, this is the first case study to estimate an economic policy uncertainty index for Belgium by mining all articles about the economy over a period of 13 years from six Belgian online newspapers.

This paper is organised as follows: Section 2 creates an EPU index using the naïve methodology. Next, Section 3 applies text mining techniques to improve the uncertainty indicator. Section 4 evaluates the three final indicators. Section 5 investigates the possible use of the indicators for forecasting macroeconomic and financial variables. Finally, Section 6 concludes the paper.

2. Naïve method

We compare our adjustments with the basic technique, as developed by Baker et al. (2015). Their newspaper index represents the number of articles that contain the words 'economy' or 'economic', 'uncertain' or 'uncertainty' and at least one policy-related word. For Europe, these policy-related words are: 'central bank', 'policy', 'tax', 'spending', 'regulation', 'budget' and 'deficit'. We refer to this as the naïve method, since it adds no weights to the different keywords.

Using a Java-based web crawler that was designed specifically for this study, we searched in the archives of five Flemish newspapers and one online news site for articles containing the keywords 'economy' and 'economic'. The newspapers are 'De Tijd', 'De Standaard', 'Het Nieuwsblad', 'Het Laatste Nieuws' and 'De Morgen', and the news site is 'DeRedactie.be'. Being restricted by the newspaper with the smallest online archive, we collected articles starting from the year 2000. This results in a dataset of approximately 210,000 news items. We automatically counted the number of articles per month and per news source that contained the aforementioned queries, in accordance with the technique of Baker et al. (2015). For each news source, we then rescaled the resulting values to a unit standard deviation. Such standardisation allowed us to sum across the six news sources in each month. The resulting values were divided by the number of news sources that archived articles in the respective month, as this increases with time. Finally, the series was rescaled to an average of 100, in accordance with the method developed by Baker et al. (2015).[1] The introduction mentioned the likelihood of type I and type II errors when using the naïve method to create an EPU index. In the naïve method, all articles that fit the query are added to the index, regardless of the entity that the policy uncertainty in the article is related to. Next to articles about Belgian and European uncertainty, this method also includes articles about Chinese, American and African uncertainty. It is clear that the naïve method is prone to overlooking relevant

[1] Due to the unavailability and/or incompleteness of data on the total number of articles published by certain newspapers, we could not scale the counts by the total number of articles published by the same news source each month.

articles. We try to solve this major flaw using two different text mining techniques: modality annotation and text classification.

3. Improvement through text mining

3.1. Modality annotation

Linguistic modality is a process that allows authors to express beliefs, attitudes and obligation in the sentences they produce (Palmer, 2001). One of the attitudes that can be expressed using modality is uncertainty. The automatic detection of modality is a well-researched topic in natural language processing (Farkas, Vincze, Móra, Csirik, & Szarvas, 2010). Various different lexical items and constructions can be used to express uncertainty, including auxiliary verbs (such as may and might), main verbs (such as hesitate, suggest, wonder, doubt), adjectives (such as uncertain, unclear), adverbs (such as unclearly, possibly) and others. We constructed a list of modal items that express uncertainty in Dutch. A translated version can be found in Table 7. This list was based on textbooks, available lists for English, and introspection. It also includes a number of expressions where negation and modality interact to express uncertainty (such as not certain and no clarity).

Instead of using the entire data set of 210,000 articles as the input for our modality annotation algorithm, we preselected articles based on our scope. While we expect that uncertainty in the euro zone will influence Belgian economic policy uncertainty, we believe that US-specific and Asian-specific policy uncertainty do not affect the policy uncertainty in Belgium directly.[2] In the small country of Belgium, policy uncertainty in America and Asia is more likely to have an impact on the financial and economic uncertainty than on the policy uncertainty. Therefore, we include only articles that refer to Belgium or a European country, where we assume the article refers to these countries if the name of one of the countries is mentioned. For Belgium, we include the names of all political parties and past prime ministers as keywords as well. This leaves us with approximately 150,000 articles for which the modality can be calculated. For each article, the relative frequency of occurrence of each word in the uncertainty list was counted. We then used the resulting modality scores to classify the articles into two classes: class one represents articles that address economic policy uncertainty in Belgium and/or the euro area (relevant articles), and class zero represents articles that address a different subject (irrelevant articles). We ranked the modality scores from high to low and set the classification threshold at 15%, meaning that the articles with scores in the top 15% were classified as relevant. This corresponds to the percentage of relevant articles in a randomly selected training set of 400 articles, as indicated by a human labeller. Finally, for each month, the number of relevant articles was divided by the number of news sources that archived articles in that month, resulting in a monthly EPU index. Table 4 reports the performance results of modality annotation on the out-of-sample test set.

We consider our current system for the detection of uncertainty using modality as a baseline system. Various improvements that would make the measure more precise are possible. Modality markers have a 'scope' (a number of words they apply to). Taking a complete article and counting the number of uncertainty markers in it is a coarse-grained approach that could be improved by using systems that compute the scope of the modality more accurately, based on syntactic analysis (Morante & Daelemans, 2009). In addition, the uncertainty dictionary could be improved by adding part-of-speech information to the words and a part-of-speech tagger to the analysis phase. For example, in English, 'may' is only an uncertainty marker when it is used as a verb, not as a noun. Finally, the dictionary could be improved by adapting it to the domain of discourse (economic texts), as different domains use different lexical items to express uncertainty.

The major issue arising when using this methodology is the fact that there is no selection by the subject of articles. Modality annotation counts the number of occurrences of words that express uncertainty across the entire data set of articles that refer to either Belgium or a European country (roughly 150,000 articles). The index is influenced by articles that address both policy uncertainty and other types of uncertainty. Therefore, this index is more an economic uncertainty index for Belgium than an economic *policy* uncertainty index. The index also depends to a large extent on the journalists' choice of vocabulary.

3.2. Text classification

The original method developed by Baker et al. (2015) assumes that articles that address economic policy uncertainty contain certain predefined keywords. In a human audit, the authors searched for the words that occur most frequently in these articles. However, this method involves self-selection of the discriminative words and cannot guarantee the absence of any predisposition towards certain queries. In order to avoid this bias, we use support vector machines (SVM) to classify the news articles. SVM has become the method of choice in supervised learning approaches to text mining. The technique automatically looks for patterns in the text documents and selects the words with the greatest discriminative power. We use an SVM with a linear kernel and get the output of a linear model where each word is assigned a weight in favour of either class 1 (EPU) or −1 (no EPU) (Fan, Chang, Hsieh, Wang, & Lin, 2008).

In a first attempt, we labelled 500 articles that were selected at random from the entire pool of articles that contain the word 'economy'. Of these, 400 articles were used as a training set, and 100 were set aside as a test set and used to calculate the performance of the classification model. The label obtains a value of 1 if the article addresses economic policy uncertainty in Belgium and/or the euro zone and a value of −1 otherwise. We define the first group of articles as the relevant articles. When constructing the

[2] Please note that we mean policy uncertainty that is specific to these countries, such as uncertainty about fiscal policies. These are articles that make no mention of or reference to either the euro area or Belgium.

Belgian EPU index, we included uncertainty in the euro zone because of the high levels of uncertainty during the European sovereign debt crisis that affected Belgium as well. Speculations about a possible Greek exit, potential bailout schemes and the monetary policy of the European Central Bank increased the policy uncertainty in Belgium due to the direct interaction between Belgian policy and that of the European Union. We chose not to include articles that address US policy uncertainty exclusively, since US policy uncertainty is not expected to affect Belgian policy directly.[3]

An important step in text data mining is the transformation of text to a structured form. Each article can be represented as a 'bag-of-words' vector $[t_1\ t_2 \dots t_j \dots t_m]$ that contains all m unique words that are present in the training set, where t_j denotes how often the jth word occurs in the article. The 'bag-of-words' vector is used to build a term-frequency matrix $\text{tf}(n, m)$, where n is the number of articles and m the number of words. In the term-frequency matrix, each cell$_{ij}$ indicates the number of times that the term j occurs in article i. Each term count is multiplied by the inverse document frequency, in order to diminish the weight of words that occur very frequently in the training set of articles. The inverse document frequency measures the frequency of a term across all documents (Weiss et al., 2010):

$$idf(t, n) = \log \times \frac{\text{Number of articles } n \text{ in the training set}}{\text{Number of articles in the training set in which term } t \text{ occurs}}.$$

The resulting tf-idf matrix is used as an input to the SVM algorithm. SVM searches for the decision boundary that maximizes the margin between the two classes. Linear SVM tries to solve the following optimisation problem (Fan et al., 2008):

$$\min_{\mathbf{w}} \frac{1}{2}\mathbf{w}^T\mathbf{w} + C\sum_i \max(1 - y_i\mathbf{w}^T\mathbf{x}_i, 0)^2, \tag{1}$$

where the vector $\mathbf{w}$ is the weights of the model and $\mathbf{x}_i$ and y_i represent the input vector and the label of the ith observation. $\max(1-y_i\mathbf{w}^T\mathbf{x}_i, 0)^2$ is the squared (L2) hinge-loss function. An out-of-sample grid search was performed to find the optimal value of C, the cost parameter.

The classification model has the following linear form:

$$f(\mathbf{x}_i) = w_0 + w_1x_{i1} + w_2x_{i2} + \cdots + w_jx_{ij} + \cdots + w_nx_{in},$$

where w_j are the weights and x_{ij} the idf-weighted occurrence of the jth unique term of the training set in the ith article. The larger the value of w_j, the more discriminative the term x_j is for an article's classification as relevant. The sign of the resulting decision value $f(\mathbf{x}_i)$ is the class the article is predicted to belong to, with a positive decision value indicating that the article addresses EPU.

Due to the skewed distribution of the complete data set of articles, our training set contained only a small number of relevant articles. In such a situation, it is advisable to expand the training set. Continuing a random selection procedure would require the use of a large selection of articles in order to find enough positive examples. Instead of labelling an additional randomly-selected set of articles, which is a cumbersome and expensive process, we have opted for a pool-based active learning algorithm with uncertainty sampling (Settles, 2010). In this procedure, the active learner has access to an unlabelled pool of articles and requests labels for the articles it is most uncertain about. In an SVM setting, uncertain instances are those that lie close to the decision boundary. Including the uncertain instances in the input vector allows the position of the decision boundary to be optimised, thereby improving the classifier (Tong & Koller, 2002).

We started by using the 400 randomly selected articles as the training set for constructing an SVM classifier, which defines the decision value for each article. The sign of the resulting decision value is the predicted class that the article belongs to. The larger the decision value, the more certain the classifier is about the chosen class. This classifier was used to label each of the news articles in the data set. The 100 articles with the lowest absolute values of the decision value were selected by the active learner to be labelled and added to the training set. Thus, the active learner requests information about the instances it is the least certain about. By labelling these articles and adding them to the new training set, the decision boundary of the classifier comes to be defined better. Next, a second classification model is created starting from this new training set. We repeat this active learning procedure several times. We used the AUC[4] of the model, calculated using ten-fold cross validation, to decide when to stop the active learning process. For each training set, Fig. 1 shows the average AUC, the AUC of the 25th percentile, and the AUC of the 75th percentile, all calculated on the ten folds. The AUC curve levels off after the fifth addition of articles. Since no further improvement in AUC was found, we decided to stop at the seventh iteration. In the end, a total of 500 articles were added to the training set, on top of the first 400 randomly selected articles. These 500 articles correspond to those added in the fifth active learning iteration.

In the naïve method, the discriminating words are defined by the authors themselves. Text mining allows us to find the words that discriminate between a relevant and an irrelevant article automatically, thereby avoiding the inherent bias that occurs when the discriminating words are self-selected. Table 1 lists the top 30 most positively discriminating words. These are the words with the highest positive weights in the classification model, meaning that their occurrence in an article increases the probability of the article being classified as EPU $= 1$. Remarkably, though not unexpectedly, the words that are related to uncertainty most frequently include words

[3] Articles about US monetary policy are generally accompanied by comments about EU monetary policy and therefore are not excluded from our data set. Articles about economic uncertainty that have a global impact are included as well.

[4] The area under the ROC curve (AUC) is the standard evaluation metric for classification models, and measures the extent to which positively labelled observations are ranked higher than negatively labelled observations (Fawcett, 2006).

Table 1
Most discriminating words for predicting the Belgian EPU index according to the trained SVM classifier.

Word	Weight	Word	Weight
ECB	0.0228	Emergency fund	0.0074
Greek	0.0147	Leaders	0.0074
Rompuy	0.0143	Spain	0.0073
Budget	0.0140	Banks	0.0073
Di Rupo	0.0133	Reforms	0.0069
Greece	0.0113	Trichet	0.0068
Eurozone	0.0113	Debt	0.0068
European	0.0112	Money	0.0067
GDP	0.0104	Plan	0.0066
Verhofstadt	0.0095	Scenario	0.0066
VAT	0.0086	Cyprus	0.0064
Irish	0.0082	Budget control	0.0064
Basis points	0.0082	Competitiveness	0.0063
Debt crisis	0.0077	Oil price	0.0062
Interest rate	0.0075	Cypriotic	0.0062

that refer to the eurozone (such as *ECB* and *Trichet*) and to other European countries (such as *Greece* and *Cyprus*). Note though that this does not mean that the largest part of Belgian EPU is due to uncertainty in other European countries, nor does it imply that countries that are not listed in the table did/do not contribute to policy uncertainty. The 30 most discriminating terms are the words that are related to uncertainty most frequently. If the training set contains a hundred articles about Italy of which thirty address EPU and five articles about Cyprus that all address EPU, words related to Cyprus will be listed as more discriminating than words related to Italy.

There still remain two issues that this methodology cannot solve. First of all, we assume that Baker et al. (2015) were right in their assumption that a news index can represent policy uncertainty. Second, we start from a data set of articles that contain the keyword 'economy' or 'economic', thus assuming that articles that do not use these words do not address economic policy uncertainty. We had to restrict ourselves to these articles in order to limit the amount of time spent labelling the articles in the active learning process. Including all articles ever published would lead to a heavily skewed distribution of classes, requiring a large number of articles to be labelled in order to obtain enough positive examples. Therefore, with both our methodology and the naïve methodology, the number of relevant articles missed is presumably larger than we report.

4. Validation

We use two different analyses to evaluate the proposed methodologies: a visual analysis and an analysis of the classification performance on an out-of-sample test set.

4.1. Visual analysis

Fig. 2 plots the three policy uncertainty indicators on the same graph. Table 2 lists major uncertainty-related events that occurred during the months in which uncertainty peaked. The naïve index spikes during the dot-com crash, the global financial crisis and in 2011, the last probably triggered by European default fears. Two of the three peaks in this index correspond to events that originated in the United States and led to global recessions, affecting the Belgian economy along the way. The SVM index shows a high volatility during the European and national debt crisis, indicating the alternation of agreement and disagreement that is inherent to every policy crisis. The modality index follows the same trend as the other two indices, increasing during the European sovereign debt crisis, but with a remarkable peak in uncertainty in 2006–2007, probably due to the municipal and federal elections. Overall, it seems that the uncertainty in Belgium is influenced largely by spillovers. The SVM index is most sensitive to the uncertainty related to the European sovereign debt crisis, followed by the naïve index.

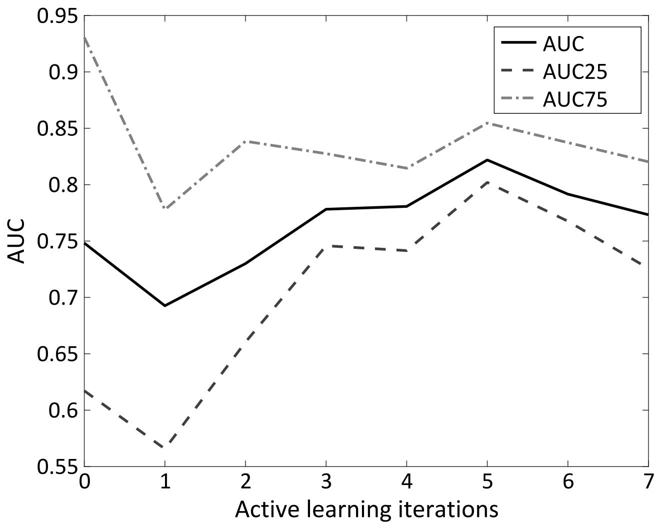

Fig. 1. Evolution of AUC on the test set during the active learning process.

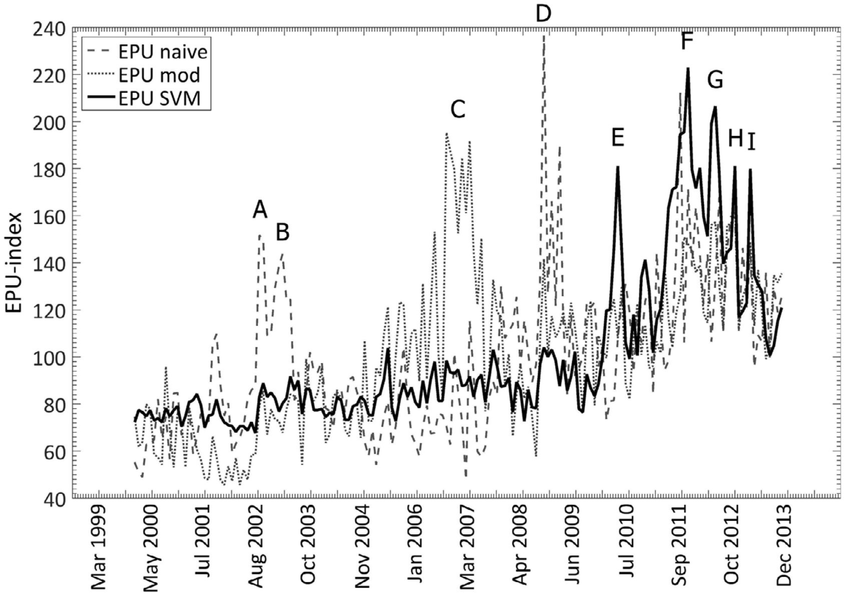

Fig. 2. The three uncertainty indicators. See Table 2 for a list of the major uncertainty-related events that occurred during the months in which the indicators peaked.

Table 2
Legend to Fig. 2.

Letter	Date	Event
A	October 2002	Declaration of Federal Policy Second Irish referendum on the Treaty of Nice
B	March 2003	Invasion of Iraq
C	March 2007	Municipal and federal elections
D	October 2008	Fortis Takeover/Banking crisis
E	April/May 2010	Greek bailout
F	October/November 2011	Referendum Greece/Forced resignation of Berlusconi Nationalisation of Belfius (former Dexia) End of government negotiations
G	June 2012	Bank bailout in Spain/New elections in Greece Cyprus requests eurozone bailout
H	November 2012	Renewed worries about Greece's debt crisis Inner cabinet meetings on draft state budget in Belgium Recapitalisation of Dexia
I	March 2013	Cyprus bailout/Italian general election/Belgian budget control

4.2. Classification performance

We evaluate the performances of the different methodologies for distinguishing between relevant and irrelevant articles, using three performance metrics that are used commonly in document classification: accuracy, specificity and recall (Sokolova & Lapalme, 2009). Table 3 represents the formulas for these metrics, in terms of true (T) or false (F) positive (P) and negative (N) classifications. Accuracy is the percentage of articles that are classified correctly. Specificity represents the fraction of irrelevant articles that are classified correctly, while recall is the fraction of relevant articles that are detected. We classify a test set of 100 articles automatically, using the three methodologies, then compare these labels to manual labels. In order to ensure a higher degree of robustness, this manual classification of the test set was performed by two independent parties: the authors and an Economics Master student. All parties classified the articles according to the assumed definition of economic policy uncertainty, and the disagreement between the labellers is very small. The results in Table 4 are the average classification results of the three methods, calculated on the out-of-sample test set of 100 articles. With an accuracy of 88%, the SVM classification model reports the best performance. The naïve method and the classification method perform equally well at detecting the articles that do not address EPU, but the SVM method outperforms the naïve method when detecting articles about EPU. Modality annotation underperforms at all levels.

Table 3
Performance metrics.

Metric	Formula
Accuracy	$\frac{TP+TN}{TP+FP+TN+FN}$
Specificity	$\frac{TN}{TN+FP}$
Recall (Sensitivity)	$\frac{TP}{TP+FN}$

Table 4
Results on the out-of-sample test set for the three methods.

Method	Accuracy	Specificity	Sensitivity
Naïve method	0.70	0.97	0.21
Modality annotation	0.52	0.67	0.25
SVM classification	<u>0.88</u>	<u>0.99</u>	<u>0.68</u>

5. Forecasting

International institutions claim that high uncertainty levels have slowed the recovery following the financial crisis (IMF, 2012). Several studies have investigated the relationship between uncertainty and macroeconomic time series; see for example ECFIN (2012), Knotek and Khan (2011), and Kose and Terrones (2012). This section investigates the predictive power of the indicators for a variety of macroeconomic and financial variables. The following variables are considered as dependent variables in the forecasting exercise:

1. The 10-year Belgian government bond yield
2. The spread between the Belgian and German 10-year bond yields (i.e., the OLO-Bund spread)
3. The Credit Default Swap (CDS) spread on Belgian senior, unsecured 5-year bonds
4. The Consumer Confidence Indicator (CCI)
5. The Business Survey Indicator
6. The Expected Demand in Construction Indicator
7. The forecast indicator for major purchases of households (over the next 12 months)
8. The Harmonised Consumer Price Index (HICP)
9. The Vehicles Registration Index
10. Bel20 stock returns.

The data were collected from the statistics websites of the National Bank of Belgium and the European Central Bank, and from Thomson Reuter's Datastream. We create several forecast models for each variable. The null (benchmark) model contains only a constant. Each of the alternative models includes one of the uncertainty indicators. We performed rolling forecasts with a fixed window size. Rolling forecasts can deal with parameter and forecast instability to a certain extent, which are recurring problems in financial and macroeconomic time series (Rossi, 2013). Both the indicators and the dependent variables show high levels of volatility over the full sample period, with certain indicators, such as CDS and the spread, peaking after the financial crisis. We compare the predictive powers of the uncertainty indicators in periods of low and high uncertainty by splitting our performance-reporting sample into the three periods listed in Table 5. The first sample covers the period January 2000 to December 2013. The second starts in May 2008, when the financial crisis translated to a higher bond spread, and ends in December 2013. The third and final sample starts in May 2008 and ends in February 2012, thus focusing mainly on the intensification of the debt crisis in 2011. For each sample period S_i with $i = 1, 2, 3$, we examine rolling h-month-ahead forecasts with $h = 1, 2, 3$ at each forecast origin $t = R + 1, \ldots, S_i$, with a fixed window size set to $R = 72$ months (six years) in order to approximate a business cycle.[5] We also consider a smaller window size of $R = 20$ months in order to allow for forecasts of the credit default swap spread, for which only a limited time series was available. The models are generated using ordinary least squares regression. We ensure time series stationarity by working with first differences or growth variables.

Hence, the forecast models are of the general form:

1. $\Delta Y_{S_i,T,t+h} = \alpha_{S_i,T,0} + \xi_{S_i,T,t}$
2. $\Delta Y_{S_i,T,t+h} = \alpha_{S_i,T,0} + \alpha_{S_i,T,1} \cdot \Delta EPUSVM_{S_i,T,t} + \xi_{S_i,T,t}$
3. $\Delta Y_{S_i,T,t+h} = \alpha_{S_i,T,0} + \alpha_{S_i,T,1} \cdot \Delta EPUMOD_{S_i,T,t} + \xi_{S_i,T,t}$
4. $\Delta Y_{S_i,T,t+h} = \alpha_{S_i,T,0} + \alpha_{S_i,T,1} \cdot \Delta EPUNAIVE_{S_i,T,t} + \xi_{S_i,T,t}$.

For each alternative model, we test the null hypothesis that the benchmark and alternative models have equal forecast accuracies. Since the forecasts are generated by nested models, the traditional Diebold and Mariano (1995) MSE test statistic will have a non-standard distribution under the null hypothesis (Diebold, 2015). Therefore, we approximate asymptotically valid critical values using the bootstrap method proposed by Clark and McCracken (2012).

We found four of the 10 variables to have predictive power, namely the spread, yield, CDS spread and CCI. Both the spread and the CDS spread show patterns that are similar to EPU SVM, with remarkable peaks at the beginning of the sovereign debt crisis in 2009 and during its intensification in 2011. The results for these variables are reported in Table 6. For each alternative model, this table provides the ratio of the alternative models' root mean square prediction errors (RMSPE) to that of the benchmark model. Values smaller than one are printed in bold, and indicate that adding the respective uncertainty variable reduces the forecast error. The bootstrapped p-values of the MSE-t test for the pairwise comparisons are listed in parentheses, and p-values that are smaller than 0.10 are underlined. We computed the bootstrapped p-values using 2000 replications.

The OLO-Bund spread surged during the intensification of the sovereign debt crisis, due to concerns about Belgium's credit and liquidity risk. A similar surge can be found in the SVM and naïve EPU indicators. This pattern appears to translate into significant predictive power only for the EPU SVM indicator. For the second and third sample periods, the alternative model including EPU SVM performs significantly better than the null model for forecasting one month ahead. When forecasting further ahead, we find different results for the two window sizes. The SVM index has predictive power for the $R = 20$, $h = 2$ and $R = 72$, $h = 3$ models. Including either EPU MOD or EPU Naïve does not improve the forecast accuracy significantly.

Regarding the Belgian bond yield, the results show that adding EPU SVM to the forecast model improves the accuracy significantly when predicting one month ahead with a small window size. When considering a longer window size of six years, we find the EPU

[5] According to the National Bureau of Economic Research, the business cycle had an average duration of 69 months over the period 1945–2009 (NBER, 2010).

Table 5
The different time periods used for the forecast performance reporting.

Time period	Sample	Abbreviation
01/2000–12/2013	Full sample, forecasts start at $R+1$	S_1
05/2008–12/2013	Financial/debt crisis in Belgium and aftermath	S_2
05/2008–02/2012	Financial/debt crisis in Belgium	S_3

Table 6
Results of the rolling forecasts for spread, yield, CDS spread, and consumer confidence.

Variable	$h=1$	$h=2$	$h=3$	$h=1$	$h=2$	$h=3$	$h=1$	$h=2$	$h=3$
	Spread								
	S_1			S_2			S_3		
	$R=20$			$R=20$			$R=20$		
EPU SVM	1.024	1.010	1.025	**0.982**	**0.996**	1.019	**0.977**	**0.968**	1.000
	(0.613)	(0.125)	(0.798)	(0.022)	(0.116)	(0.588)	(0.045)	(0.022)	(0.252)
EPU Mod	1.060	1.047	1.055	1.041	1.025	1.042	1.041	1.013	1.012
	(0.849)	(0.711)	(0.808)	(0.875)	(0.997)	(0.771)	(0.901)	(0.992)	(0.483)
EPU Naïve	1.021	1.021	1.027	1.062	1.014	1.022	1.055	1.013	1.019
	(0.720)	(0.854)	(0.740)	(0.413)	(0.883)	(0.947)	(0.372)	(0.660)	(0.739)
	S_1			S_2			S_3		
	$R=72$			$R=72$			$R=72$		
EPU SVM	1.001	1.005	1.001	**0.996**	1.011	**0.995**	**0.993**	1.012	**0.997**
	(0.283)	(0.511)	(0.212)	(0.083)	(0.800)	(0.074)	(0.026)	(0.695)	(0.140)
EPU Mod	1.007	1.009	1.008	1.006	1.002	1.002	**0.998**	1.001	1.001
	(0.728)	(0.440)	(0.816)	(0.657)	(0.448)	(0.483)	(0.306)	(0.391)	(0.389)
EPU Naïve	1.019	1.004	1.004	1.003	1.004	1.000	1.003	1.002	1.000
	(0.352)	(0.244)	(0.991)	(0.473)	(0.641)	(0.302)	(0.488)	(0.400)	(0.195)
	Yield								
	S_1			S_2			S_3		
	$R=20$			$R=20$			$R=20$		
EPU SVM	**0.900**	1.009	1.034	**0.972**	1.002	1.032	**0.994**	**0.998**	1.011
	(0.002)	(0.091)	(0.760)	(0.003)	(0.147)	(0.806)	(0.208)	(0.188)	(0.442)
EPU Mod	1.023	1.029	1.038	1.021	1.021	1.036	**0.999**	**0.995**	**0.996**
	(0.845)	(0.485)	(0.510)	(0.848)	(0.302)	(0.559)	(0.286)	(0.206)	(0.207)
EPU Naïve	1.021	1.043	1.021	1.013	1.029	1.025	1.019	1.032	1.025
	(0.201)	(0.963)	(0.568)	(0.223)	(0.961)	(0.953)	(0.419)	(0.835)	(0.728)
	S_1			S_2			S_3		
	$R=72$			$R=72$			$R=72$		
EPU SVM	1.002	1.001	1.005	1.002	**0.999**	1.006	**0.992**	1.004	1.008
	(0.376)	(0.344)	(0.532)	(0.448)	(0.260)	(0.739)	(0.131)	(0.750)	(0.583)
EPU Mod	1.003	**0.999**	1.004	1.003	**0.998**	1.002	**0.996**	1.001	1.000
	(0.486)	(0.075)	(0.708)	(0.586)	(0.097)	(0.681)	(0.200)	(0.638)	(0.344)
EPU Naïve	1.005	1.008	1.026	1.008	1.005	1.006	1.014	1.003	1.007
	(0.392)	(0.690)	(0.654)	(0.747)	(0.802)	(0.668)	(0.700)	(0.433)	(0.535)
	Credit default swap spread								
	S_1			S_2			S_3		
	$R=20$			$R=20$			$R=20$		
EPU SVM	na	na	na	**0.982**	**0.983**	**0.998**	**0.992**	**0.965**	**0.992**
				(0.028)	(0.024)	(0.081)	(0.194)	(0.029)	(0.127)
EPU Mod	na	na	na	1.041	1.051	1.024	1.031	1.050	**0.984**
				(0.879)	(0.614)	(0.292)	(0.964)	(0.554)	(0.110)
EPU Naïve	na	na	na	1.008	1.009	1.039	1.024	1.012	1.048
				(0.175)	(0.147)	(0.716)	(0.364)	(0.169)	(0.697)
	Consumer confidence index								
	S_1			S_2			S_3		
	$R=20$			$R=20$			$R=20$		
EPU SVM	**0.998**	1.137	1.152	**0.998**	1.027	1.011	**0.998**	1.004	1.006
	(0.022)	(0.563)	(0.499)	(0.130)	(0.609)	(0.531)	(0.130)	(0.404)	(0.464)
EPU Mod	1.002	1.036	1.042	1.000	1.076	1.124	1.000	1.015	1.007
	(0.881)	(0.618)	(0.538)	(0.453)	(0.864)	(0.884)	(0.403)	(0.601)	(0.579)
EPU Naïve	1.002	1.013	1.019	1.001	1.035	1.016	1.001	1.036	1.013

(continued on next page)

Table 6 (*continued*)

Variable	$h = 1$	$h = 2$	$h = 3$	$h = 1$	$h = 2$	$h = 3$	$h = 1$	$h = 2$	$h = 3$
	(0.872)	(0.179)	(0.391)	(0.710)	(0.717)	(0.430)	(0.694)	(0.756)	(0.462)
	S_1			S_2			S_3		
	$R = 72$			$R = 72$			$R = 72$		
EPU SVM	1.001 (0.677)	1.007 (0.707)	1.005 (0.459)	**0.999** (0.128)	1.016 (0.804)	1.009 (0.692)	**0.999** (0.122)	1.016 (0.786)	1.008 (0.644)
EPU Mod	1.011 (0.863)	1.019 (0.836)	1.003 (0.216)	1.001 (0.648)	1.004 (0.532)	1.031 (0.637)	1.002 (0.662)	1.003 (0.495)	1.032 (0.689)
EPU Naïve	1.013 (0.180)	1.055 (0.855)	1.021 (0.855)	1.001 (0.460)	1.017 (0.871)	1.009 (0.801)	1.001 (0.444)	1.017 (0.862)	1.008 (0.761)

Notes: The ratio of the alternative RMSPE to the benchmark RMSPE is listed for each uncertainty indicator. The p-values are reported in brackets and are calculated using the bootstrapped distribution of the MSE-t statistic. RMSPE ratios that are smaller than one are printed in bold, and p-values that are smaller than 0.10 are underlined.

Table 7
Dictionary of modal words.

Modality					
apparently	assume	assumption	believe	claim	estimate
apparent	hope	hypothesis	hypothetical	if	imagine
indication	likely	potential	potentially	preferably	sense
show	think	feel	obvious	obviously	can
could	seem	latent	mean	maybe	possibly
possible	perhaps	possibility	obscure	seemingly	ostensibly
evasive	elusive	unstable	unsettled	unclear	vague
uncertain	unsure	unknown	unfamiliar	improbable	improbably
potential	potentially	preferential	preferentially	estimate	questionable
questionably	appear	speculate	speculation	suggest	doubt
doubting	doubtful	dubious	suppose	expect	expecting
suspect	suspecting	presumably	supposedly	pretended	supposed
putative	probably	probable	point at	perchance	might
would	may	no clue	no evidence	uncertainty	no sign
not easily	no clear	no clarity	not easy	no possible	no possibility
no notion	no idea	not plausible	not sure	not certain	not probable
not likely	not credible	not known	not familiar	raise questions	or
either	and/or				

MOD-model to have significant predictive power at the horizon $h = 2$. The forecasting ability of the yield breaks down during the period associated with the financial and debt crisis (S_3). Intuitively, policy uncertainty is related more to the spread between the Belgian and German government bonds than to the yield on Belgian government bonds. Policy uncertainty cannot predict the initial decline in yields between early 2008 and mid-2010 that was due to the relaxation of monetary policy as a reaction to the sharp slowdown of the economy in the euro area. From mid-2010 until the end of our data set, the movement in the Belgian yield was driven by the spread between Belgian and German government bonds, with monetary policy being less effective for steering government bond yields during that period.

The Thomson Reuters dataset on Credit Default Swaps starts in 2008, which is why there are no forecasting results available for $R = 72$ or sample S_1. However, for S_2 and S_3 with $R = 20$, the alternative model that includes EPU SVM outperforms the benchmark model for all three horizons. For S_2, the improvement in forecast accuracy is significant at the 5% level for the first two horizons and at the 10% level for $h = 3$. When focusing only on the intensification of the debt crisis, we find no significance for $h = 1, 3$, while there appears to be a significant performance improvement when predicting two months ahead.

The last variable, the consumer confidence index, is difficult to predict using typical economic indicators, as it appears to be driven by trust and 'animal spirits' (Neisingh & Stokman, 2013). The forecasting results show that the indicator is also driven by economic policy uncertainty to some extent. An improvement in forecast accuracy is found when the EPU SVM indicator is included for predicting one month ahead. However, the significance breaks down when considering only the period after the financial crisis.

Across all permutations of sub-sample, model, horizon and window size, we find only limited evidence of predictive content for any of the three uncertainty indices. The predictive content that we do find is concentrated largely in our SVM index, and is larger when the sample length is relatively small. There are no examples of statistically significant predictive content when using the naïve index. The forecasting exercise shows that EPU SVM can be used to improve the forecast accuracy when predicting changes in both yield-related variables and consumer confidence in the short-term. The fact that a small sample length seems to work better suggests that the policy uncertainty is relevant for these variables only during specific periods.

6. Conclusion

According to international institutions, uncertainty rose to historically high levels after the global recession of 2007–2009, due to uncertainty about the future government policy. Our EPU indicators for Belgium provide strong support for this claim and indicate that the national uncertainty was influenced in part by uncertainty in the euro area. We have applied text mining techniques to a policy-related problem and have tried to improve the original policy uncertainty index by building a classification model that replaces the self-selected keywords of the original methodology. A forecasting exercise on ten macroeconomic and financial variables demonstrates that the more advanced methodology has a greater predictive power and speaks in favour of using an SVM classification model when constructing a news-based policy uncertainty indicator. We find that including the EPU SVM indicator improves the forecast accuracy significantly when predicting the OLO-Bund spread, long term government bond yield, CDS spread and consumer confidence in the short term. On average, the predictive power seems to break down when predicting more than two months ahead, with only the CDS spread being able to be predicted three months ahead. The case study presented in this paper shows a possible application of big data analytics to theory building and risk intelligence in the research area of economics, which is currently still dominated by causal statistical modelling. To encourage further research on the influence of uncertainty on the economy, a daily updated version of the EPU SVM indicator can be downloaded from our website (http://www.applieddatamining.com). The other indicators are available upon request.

Acknowledgments

We thank Michael Schoenmaekers for labelling the test set. Ellen Tobback thanks the Flemish Research Council (FWO) for financial support. We also thank the National Bank of Belgium for their financial support that enabled us to create a web page containing an up-to-date, publicly available version of our EPU index. The views expressed are the authors' alone and do not necessarily correspond to those of the European Central Bank or the European Commission.

References

Bachmann, R., Elstner, S., & Sims, E.R. (2010). Uncertainty and economic activity: Evidence from business survey data. Working Paper No. 16143.
Baker, S.R., Bloom, N., & Davis, S.J. (2015). Measuring economic policy uncertainty. Working Paper No. 21633.
Balta, N., Fernández, I.V., & Ruscher, E. (2013). Assessing the impact of uncertainty on consumption and investment, Section I, Quarterly Report on the Euro Area, 12, 7–6.
Bloom, N. (2009). The impact of uncertainty shocks. *Econometrica*, *77*, 623–685.
Clark, T. E., & McCracken, M. W. (2012). Reality checks and comparisons of nested predictive models. *Journal of Business and Economic Statistics*, *30*, 53–66.
Diebold, F. X. (2015). Comparing predictive accuracy, twenty years later: A personal perspective on the use and abuse of Diebold–Mariano tests. *Journal of Business and Economic Statistics*, *33*, 1–9.
Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics*, *13*, 253–263.
ECFIN, D. (2012). Risk and uncertainty in euro area sovereign debt markets and their impact on economic activity. Quarterly report on the euro area No. 11.
Fan, R.-E., Chang, K.-W., Hsieh, C.-J., Wang, X.-R., & Lin, C.-J. (2008). Liblinear: A library for large linear classification. *The Journal of Machine Learning Research*, *9*, 1871–1874.
Farkas, R., Vincze, V., Móra, G., Csirik, J., & Szarvas, G. (2010). The CoNLL-2010 shared task: learning to detect hedges and their scope in natural language text. In *Proceedings of the fourteenth conference on computational natural language learning: Shared task* (pp. 1–12). Association for Computational Linguistics.
Fawcett, T. (2006). An introduction to roc analysis. *Pattern Recognition Letters*, *27*, 861–874.
IMF. (2012). IMF survey: Global recovery, growth hampered by uncertainty — Lagarde. IMF Survey, http://www.imf.org/external/pubs/ft/survey/so/2012/SurveyartG.htm.
in't Veld, J. (2013). Fiscal consolidations and spillovers in the euro area periphery and core. European Economy Economic Papers 506.
Knotek, E. S., II, & Khan, S. (2011). How do households respond to uncertainty shocks? *Economic Review-Federal Reserve Bank of Kansas City*, 63.
Kose, M.A., & Terrones, M. (2012). How does uncertainty affect economic performance? World Economic Outlook Box 1.3.
Morante, R., & Daelemans, W. (2009). Learning the scope of hedge cues in biomedical texts. In *Proceedings of the workshop on current trends in biomedical natural language processing* (pp. 28–36). Association for Computational Linguistics.
NBER. (2010). US business cycle expansions and contractions. URL http://http://www.nber.org/cycles/cyclesmain.html (Accessed on 17.08.16).
Neisingh, P., & Stokman, A.C. (2013). What drives consumer confidence in times of financial crises? Evidence for the Netherlands. De Nederlandsche Bank Working Paper.
Palmer, F. R. (2001). *Mood and modality*. Cambridge University Press.
Rossi, B. (2013). *Handbook of economic forecasting: 2. Advances in forecasting under instability* (pp. 1203–1324). Elsevier, (chapter 21).
Settles, B. (2010). *Active learning literature survey. Computer Sciences Technical Report 1648*.
Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing and Management*, *45*, 427–437.
Tong, S., & Koller, D. (2002). Support vector machine active learning with applications to text classification. *The Journal of Machine Learning Research*, *2*, 45–66.
Weiss, S. M., Indurkhya, N., & Zhang, T. (2010). *Fundamentals of predictive text mining, Vol. 41*. Springer.

Ellen Tobback is a Ph.D. student at the University of Antwerp, Belgium, where she received her Master's and Bachelor's degrees in Commercial Engineering. Since 2013 she has been a member of the Applied Data Mining research group at the Department of Engineering Management. She received a Ph.D. Fellowship from the Flanders Research Foundation and expects to complete her Ph.D. dissertation during the 2017–2018 academic year. Her general research area is data mining for credit risk prediction.

Hans Naudts is an economic analyst at DG Economics and Finance (DG ECFIN) at the European Commission, and an associated researcher of the Department of Engineering Management at the University of Antwerp. He obtained an MRes in Economics from the London School of Economics and Political Science (LSE) in 2006. During 2009–2011, he served as economic adviser to the Belgian Prime Minister.

Walter Daelemans studied linguistics and psycholinguistics in Antwerp and Leuven and worked as a computational linguist in Nijmegen and the Brussels AI-lab. While teaching computational linguistics at the University of Tilburg, he founded a research group on Machine Learning of Language (ILK) and started investigating machine learning approaches to language processing, among other things. He is currently professor of computational linguistics at the University of Antwerp, where he directs the computational linguistics group of CLiPS. His current research interests are in text mining, computational stylometry, and computational psycholinguistics.

Enric Junqué de Fortuny is an assistant professor at the Rotterdam School of Management, Erasmus University Rotterdam (the Netherlands).

He holds a master's degree in Computer Science Engineering from the University of Ghent (Belgium) and a Ph.D. in Applied Economics from the University of Antwerp (Belgium). During 2014–2016, he was a Senior Research Fellow at INSEAD's eLab for Big Data (France/Singapore). His research interests include the development of machine learning algorithms, with a specific focus on the prediction of human behavior and the improvement of comprehensibility in data science. His findings have been published in well-known journals and top conferences such as *IEEE Transactions on Neural Networks*, MISQ and *Learning Systems and Knowledge Discovery and Data Mining* (KDD).

David Martens is assistant professor at the University of Antwerp, where he heads the Applied Data Mining research group. His research focuses on the development and application of data mining techniques that lead to an improved understanding of human behavior, and their use in marketing and finance. In 2014, David won the "Best EJOR Application Award" (*European Journal of Operational Research*) for his work on churn prediction. In 2008, David was finalist of the prestigious international KDD doctoral dissertation award. Together with prof. Foster Provost, he is the inventor of four pending patent applications of data mining methods.