Belgian economic policy uncertainty index : improvement through text mining

Reference:
Tobback Ellen, Naudts Hans, Daelemans Walter, Junqué de Fortuny Enric, Martens David.- Belgian economic policy uncertainty index : improvement through text mining
International journal of forecasting - ISSN 0169-2070 - 34:2(2018), p. 355-365
Full text (Publisher's DOI): https://doi.org/10.1016/J.IJFORECAST.2016.08.006
To cite this reference: https://hdl.handle.net/10067/1363340151162165141

Belgian Economic Policy Uncertainty Index : Improvement through text mining

Ellen Tobback[1], Hans Naudts[11], Walter Daelemans[1], Enric Junqué de Fortuny[1], David Martens[1]

Abstract

The recent literature measures economic policy uncertainty using news references, resulting in the frequently mentioned 'Economic Policy Uncertainty-index' (EPU). In the original set-up, a news article is assumed to address policy uncertainty if it contains certain predefined keywords. We argue that the original set-up 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 to create an EPU-index for Belgium. Validation on an out-of-sample test set speaks in favour of using an SVM classification model when constructing a news based policy uncertainty indicator. The indicators are then used to forecast 10 macroeconomic and financial variables. The original method for measuring EPU does not have predictive power on any of these 10 variables. The SVM indicator has a higher predictive power and, notably, during tumultuous periods of high uncertainty and risk, changes in the level of policy uncertainty can predict changes in sovereign bond yield and spread, credit default swap spread, and consumer confidence.

1. Introduction

According to international institutions economic policy uncertainty rose to historically high levels after the 2007-2009 recession because of uncertainty about tax, spending, regulatory, and monetary policies (Balta, Fernández & Ruscher, 2013; IMF, 2012). This uncertainty has slowed recovery from the recession by causing businesses and households to cutback or postpone investment, hiring and consumption. For example, in't Veld (2013) models the impact on GDP of fiscal consolidation under different uncertainty and learning scenarios. In a scenario of uncertainty on the credibility of the fiscal consolidation, the short term negative impact on GDP is up to 3 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 the effect of uncertainty on activity increasing since the crisis and going beyond traditional cyclical effects. Economic research has come up with several ways of constructing uncertainty measures based on stock market volatility (Bloom, 2009; Kose & Terrones, 2012), dispersion in forecasts by professional forecasters or in 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). In this paper we focus on the latter methodology and contribute 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) have 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 near 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. Notwithstanding its widespread use and acceptance, there remain some important issues regarding the construction of the index. The method is likely 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. The method suggested by Baker et al. (2015) can thus cause a high rate of both false positives and negatives.

In this paper we attempt to improve this methodology by solving its main issues using text mining. 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 expressing uncertainty, the latter is a trained classifier that predicts whether an article addresses economic policy uncertainty. Conform to Baker et al. (2015), we define economic policy uncertainty as uncertainty about who will make what policy decision when and as uncertainty about the effect of past/present/future policy decisions. We limit Belgian economic policy uncertainty to uncertainty at Belgian and euro area level. 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.

The contribution of this paper is 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 applied to solve a policy related problem. In this particular case, the policy related problem is finding a measure for economic policy uncertainty. We assess policy uncertainty by automatically detecting patterns in a total of 210,000 news articles, using modality annotation and text classification on a large data set of news articles. By doing so, we add to the economic theory, for example, by investigating the coefficients of the trained SVM model, we can see which words are most frequently related to policy uncertainty in the news articles. 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 that estimates an economic policy uncertainty index for Belgium by mining all articles about the economy over a period of 13 years from 6 Belgian online news papers.

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

2. Naive method

We will compare our adjustments to 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 it as the naive method since it adds no weights to the different keywords.

Using a Java-based web crawler specifically designed for this study, we searched for articles containing the keywords ‘economy’ and ‘economic’ in the archives of five Flemish newspapers and one online news site. The newspapers are ‘De Tijd’, ‘De Standaard’, ‘Het Nieuwsblad’, ‘Het Laatste Nieuws’ and ‘De Morgen’, the news site is ‘DeRedactie.be’. Being restricted by the newspaper with the smallest online archive, we collected all articles starting from the year 2000. This results in a dataset of approximately 210,000 news items. Per month and per news source, we automatically counted the number of articles containing the aforementioned queries, in accordance with the technique of Baker et al. (2015). For each news source, we rescaled the resulting values to unit standard deviation. 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). [2]

In the introduction, we have mentioned the likelihood of type I and type II errors when applying the naive method to create an EPU-index. In the naive method, all articles that fit the query are added to the index, regardless of the entity the policy uncertainty in the article is related to. Next to articles about Belgian and European uncertainty, this method includes articles about Chinese, American and African uncertainty as well. It is clear that the naive method is prone to overlook relevant 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 belief, attitude and obligation in the sentences they produce (Palmer, 2001). One of the attitudes that can be expressed with modality is uncertainty. Detecting modality automatically is a well-researched topic in Natural Language Processing (Farkas, Vincze, Móra, Csirik & Szarvas, 2010). 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 expressing uncertainty in Dutch. A translated version can be found in the Appendix. This dictionary 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 input to our modality annotation algorithm, we preselected articles according to our scope. While we consider uncertainty in the euro zone to influence Belgian economic policy uncertainty, we believe that U.S. specific and Asian specific policy uncertainty does not directly affect policy uncertainty in Belgium.[3] Policy Uncertainty

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

[3] Please note that we mean policy uncertainty specific for these countries, such as uncertainty about fiscal policies. These are articles without any mention of or reference to the euro area or Belgium.

in America and Asia is more likely to have an impact on financial and economic uncertainty than on policy uncertainty in the small country of Belgium. Therefore, we include only articles that refer to Belgium or a European country, where we assume the article refers to the considered countries if one of the countries' name 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 of which the modality can be calculated. For each word in an article, the relative frequency of words occurring in the uncertainty list was counted. We used the resulting modality scores to classify the articles into two classes: class one represents the 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 the 15% highest scores were classified as relevant. This percentage corresponds to the percentage of relevant articles, as indicated by a human labeller, in a randomly selected training set of 400 articles. 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. Different improvements are possible that would make the measure more precise. Modality markers have a 'scope' (a number of words they apply to). Taking a complete article and counting the uncertainty markers in it, is a course-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 adding a part of speech tagger in the analysis phase. For example, 'may' in English is only an uncertainty marker when used as a verb, not as a noun. A final way of improving the dictionary would be to adapt 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 on the subject of articles. Modality annotation counts the occurrence of words expressing uncertainty in the entire data set of articles that refer to either Belgium or a European country (roughly 150,000 articles). Both articles addressing policy uncertainty and articles addressing a different type of uncertainty have an impact on the index. Therefore, this index is more an economic uncertainty index for Belgium than an economic *policy* uncertainty index. The index also depends to a larger extent on the choice of vocabulary by the journalists.

3.2. Text classification

The original method developed by Baker et al. (2015) assumes that articles addressing economic policy uncertainty contain certain predefined keywords. During a human audit, the authors have searched for the words that occur most frequently in these articles. This method involves self-selection of the discriminative words and cannot guarantee the absence of a predisposed inclination 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 largest discriminative power. We use an SVM with a linear kernel and as output we get 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 randomly selected from the entire pool of articles that contain the word 'economy'. 400 articles were used as training set, 100 articles were set aside as 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 Belgian EPU-index, we included uncertainty in the euro zone due to 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 increase policy uncertainty in Belgium due to the direct interaction between Belgian policy and the policy of the European Union. We chose not to include articles that exclusively address U.S. policy uncertainty, since U.S. policy uncertainty is not expected to directly affect Belgian policy.[4]

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 $[t_1 t_2 \ldots t_j \ldots t_m]$ that contains all m unique words present in the training set, where t_j denotes how often the j^{th} word occurs in the article. The 'bag-of-words' vector is used to build a term-frequency matrix tf(n, m) with n the number of articles and m the number of words. In the term-frequency matrix each cell$_{ij}$ indicates the number of times the term j occurs in article i. Each term count is multiplied by the inverse document frequency to diminish the weight of the 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).

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

The resulting tf-idf matrix is used as 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):

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

With vector $\mathbf{w}$ the weights of the model and $\mathbf{x}_i$ and y_i representing the input vector and the label of the i_{th} 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:

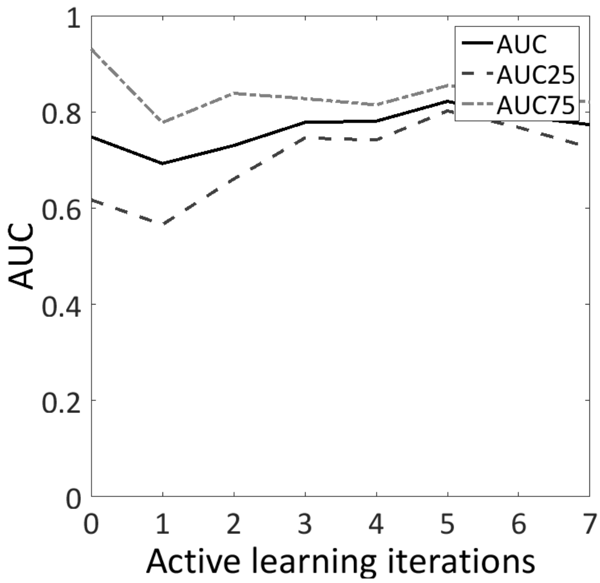
$$f(\mathbf{x}_i) = w_0 + w_1 x_{i1} + w_2 x_{i2} + ... + w_j x_{ij} + ... + w_n x_{in}$$

With w_j the weights and x_{ij} the idf-weighted occurrence of the j_{th} unique term of the training set in the i_{th} 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 predicted class the article belongs to, with a positive decision value indicating that the article addresses EPU.

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

Due to the skewed distribution of the complete data set of articles, our training set contained only a small amount of relative articles. In such a situation it is advisable to expand the training set. Continuing a random selection procedure would require a large selection of articles 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 the labels for the articles it is most uncertain about. In an SVM-setting, uncertain instances are those that lie close to the decision boundary. By including the uncertain instances in the input vector, the position of the decision boundary can be optimised, thereby improving the classifier (Tong & Koller, 2002).
We started with the 400 randomly selected articles as training set to construct an SVM classifier, which defines the decision value for each article. The sign of the resulting decision value is the predicted class 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 all the news articles in the data set. The 100 articles with the lowest decision value in absolute value were selected by the active learner to be labelled and added to the training set. The active learner thus requests information about the instances it is the least certain about. By labelling and adding these articles to the new training set, the decision boundary of the classifier becomes better defined. Starting from this new training set, a second classification model is created. We repeat this active learning procedure several times. We looked at the AUC [5] of the model calculated using ten-fold cross validation to decide when to stop the active learning process. Figure 1 shows, for each training set, the average AUC, the

Figure 1: Evolution of AUC on the test set during the active learning process



AUC of the 25th percentile, and the AUC of the 75th percentile calculated on the ten folds. The

[5]The Area Under the ROC Curve (AUC) is the standard evaluation metric for classification models, and measures to what extent positively labelled observations are ranked higher than negatively labelled observations (Fawcett, 2006).

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 was 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 naive method, the discriminating words were defined by the authors themselves. Text mining allows us to automatically find the words that discriminate between a relevant and an irrelevant article, thereby avoiding the inherent bias that occurs when the discriminating words are self-selected. Table 1 illustrates 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 unexpected, amongst the words most frequently related to uncertainty are words referring to the eurozone (such as *ECB* and *Trichet*) and to other European countries (such as *Greece* and *Cyprus*). Note that this does not mean that the largest part of Belgian EPU is due to uncertainty in European countries, neither does it imply that countries not listed in the Table did/do not contribute to policy uncertainty. The 30 most discriminating terms are the words that are most frequently related to uncertainty. If the training set contains a hundred articles about Italy with thirty addressing 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.

Table 1: Most discriminating words to predict 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

There still remain two issues that cannot be solved by this methodology. First of all, we assume Baker et al. (2015) were right in their assumption that a news index can represent policy uncertainty. Secondly, we start from a data set of articles containing the keywords ‘economy’ or ‘economic’, thereby assuming that articles without the occurrence of these words do not address economic policy uncertainty. We had to restrict ourselves to these articles to limit the time spent on labelling the articles in an 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 before having enough positive examples. Therefore, both with our methodology and the naive methodology, the number of relevant articles missed is presumably larger than reported.

4. Validation

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

4.1. Visual analysis

Figure 2 plots all three policy uncertainty indicators on the same graph. Table 2 lists major uncertainty related events that occurred during the months where uncertainty peaked. The naive index spikes during the dot-com crash, the global financial crisis and in 2011, likely triggered by European default fears. Two out of the three peaks in this index coincide with events that originated in the United States and lead to a global recession, affecting the Belgian economy along the way. The SVM index shows high volatility during the European and national debt crisis, indicating the alternation of agreement and disagreement inherent to every policy crisis. The modality index follows the same trend as the other two indices, going upwards during the European sovereign debt crisis, though showing a remarkable peak in uncertainty in 2006-2007, probably caused by the municipal and federal elections. Overall, it seems that uncertainty in Belgium is to a large extent influenced by spill-overs. The SVM index is most sensitive to the uncertainty related to the European sovereign debt crisis, followed by the naive index.

Figure 2: The three uncertainty indicators. Accompanying Table 2 lists the major uncertainty related events that occurred during the months where the indicators peaked

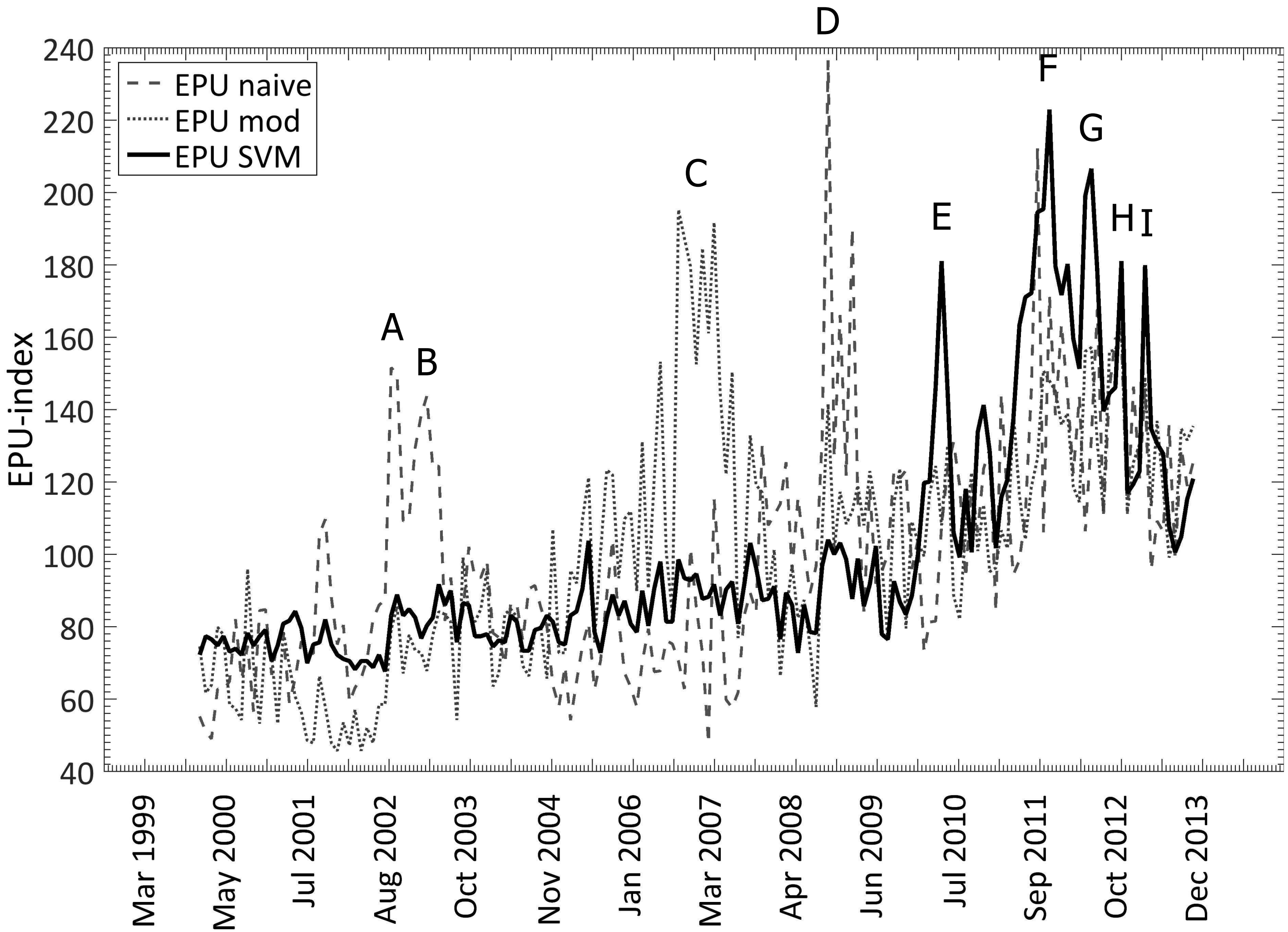


Table 2: Legend to Figure 2

Letter	Date	Event
A	October 2002	Declaration of Federal Policy Second Irish referendum on the Treaty of Nice
B	March 2003	Invasion in 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 Berlusconi Nationalisation Belfius (former Dexia) End 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 Dexia
I	March 2013	Cyprus bailout / Italian general election / Belgian budget control

4.2. Classification performance

We evaluate how the different methodologies perform at distinguishing relevant from irrelevant articles, using three performance metrics commonly used 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 correctly classified, while recall is the fraction of relevant articles that are detected. We automatically classify a test set of 100 articles, using all three methodologies. These labels are then compared to manual labels. In order to assure a higher degree of robustness, the test set was classified 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. The disagreement between the labellers is very small. The results in Table 4 are the average classification results of all 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 naive method and the classification method perform equally well at detecting the articles that do not address EPU, however, the SVM method outperforms the naive method when detecting articles about EPU. Modality annotation underperforms on all levels.

Table 3: Performance metrics

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

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

Method	Accuracy	Specificity	Sensitivity
Naive 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 after the financial crisis (IMF, 2012). Several studies have investigated the relationship between uncertainty and macroeconomic time series (ECFIN, 2012; Knotek II & Khan, 2011; Kose & 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. Business Survey Indicator
6. Expected Demand in Construction Indicator
7. Forecast indicator for major purchases of households (over the next 12 months)
8. Harmonised Consumer Price Index (HICP)
9. Vehicles Registration Index
10. Bel20 stock returns

The data was collected from the National Bank of Belgium's and European Central Bank's statistics websites and Thomson Reuter's Datastream. For each variable, we create several forecast models. The null (benchmark) model contains only a constant. The alternative models include one of the uncertainty indicators. We performed rolling forecasts with a fixed window size. Rolling forecasts handle to a certain extent parameter and forecast instability, which is an occurring problem in financial and macroeconomic time series (Rossi, 2013). Both the indicators and the dependent variables show high volatility over the full sample periods, with certain indicators, such as CDS and spread, peaking after the financial crisis. To compare the predictive power of the uncertainty indicators between periods of low and high uncertainty, we split our performance reporting sample into three periods as listed in Table 5. The first sample covers the period January 2000 to December 2013. The second sample 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, thereby 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 (6 years) in order to approximate a business cycle[6]. We also consider a smaller window size of $R = 20$ months to allow forecasts for the credit default swap spread, of which only a limited time series was available. The models are generated using Ordinary Least Squares regression. To ensure time series stationarity, we work with first differences or growth variables.

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

Hence, the forecast models are of the following 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 model have equal forecast accuracy. Since the forecasts are generated by nested models, the traditionally used Diebold & Mariano (1995) MSE test statistic will have a non-standard distribution under the null hypothesis (Diebold, 2015). Therefore, we use the bootstrap-method proposed by Clark & McCracken (2012) to approximate the asymptotically valid critical values.

We found predictive power for 4 out of the 10 variables, i.e. spread, yield, CDS spread and CCI. Both the spread and CDS spread show a pattern similar to EPU SVM, with remarkable peaks at the beginning of the sovereign debt crisis in 2009 and during the intensification thereof in 2011. The results for these variables are reported in Table 6. This Table provides, for each alternative model, the ratio of the alternative models' Root Mean Square Prediction Error (RMSPE) to the benchmark model's RMSPE . Values smaller than one are printed in bold. This indicates 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 between brackets. P-values smaller than 0.10 are underlined. We used 2000 replications in computing the bootstrap p-values.

[6]According to the National Bureau of Economic Research the business cycles in the period 1945-2009 had an average duration of 69 months (NBER, 2010).

During the intensification of the sovereign debt crisis, the OLO-Bund spread surged due to concerns about Belgium's credit and liquidity risk. A similar surge can be found in the SVM and naive EPU indicators. This similar pattern appears to be translated into significant predictive power for the EPU SVM indicator only. For the second and third sample period, the alternative model including EPU SVM performs significantly better than the null model when forecasting 1-month ahead. When forecasting further ahead, we find different results for the two window sizes. The SVM index has predictive power in the $R = 20$,$h = 2$- and the $R = 72$,$h = 3$-models. Including either EPU MOD or EPU Naive does not significantly improve the forecast accuracy.

Regarding the Belgian bond yield, the results show that adding EPU SVM to the forecast model, significantly improves accuracy when predicting 1-month ahead with a small window size. When considering a longer window size of 6 years, we find significant predictive power for the EPU MOD-model 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 more related 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 from early 2008 to mid-2010 which 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 during that period to steer the government bond yields.

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

The last variable, Consumer Confidence Index, is difficult to predict using the typical economic indicators as it appears to be an indicator driven by trust and 'animal spirits' (Neisingh & Stokman, 2013). The forecasting results show that the indicator is to some extent also driven by economic policy uncertainty. An improvement in forecast accuracy is found when including the EPU SVM indicator to predict one month ahead. However, 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 limited evidence of predictive content for any of the three uncertainty indices. The predictive content that we do find is largely concentrated 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 naive index. The forecasting exercise shows that EPU SVM can be used to improve forecast accuracy when predicting changes in yield-related variables as well as consumer confidence in the short-term. The fact that a small sample length seems to work better suggests that policy uncertainty is only relevant for these variables during specific periods in time.

Table 6: Results of the rolling forecasts for spread, yield, CDS spread, and consumer confidence. For each uncertainty indicator the ratio of the alternative RMSPE to the benchmark RMSPE is listed. The p-values are reported between brackets and are calculated using the bootstrapped distribution of the MSE-t statistic. RMSPE ratios smaller than 1 are printed in bold, p-values smaller than 0.10 are underlined.

					Spread				
		S_1			S_2			S_3	
		$R = 20$			$R = 20$			$R = 20$	
Variable	$h = 1$	$h = 2$	$h = 3$	$h = 1$	$h = 2$	$h = 3$	$h = 1$	$h = 2$	$h = 3$
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 Naive	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	
		$R = 72$			$R = 72$			$R = 72$	
Variable	$h = 1$	$h = 2$	$h = 3$	$h = 1$	$h = 2$	$h = 3$	$h = 1$	$h = 2$	$h = 3$
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 Naive	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	
		$R = 20$			$R = 20$			$R = 20$	
Variable	$h = 1$	$h = 2$	$h = 3$	$h = 1$	$h = 2$	$h = 3$	$h = 1$	$h = 2$	$h = 3$
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 Naive	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	
		$R = 72$			$R = 72$			$R = 72$	
Variable	$h = 1$	$h = 2$	$h = 3$	$h = 1$	$h = 2$	$h = 3$	$h = 1$	$h = 2$	$h = 3$
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 Naive	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	
		$R = 20$			$R = 20$			$R = 20$	
Variable	$h = 1$	$h = 2$	$h = 3$	$h = 1$	$h = 2$	$h = 3$	$h = 1$	$h = 2$	$h = 3$
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 Naive	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	
		$R = 20$			$R = 20$			$R = 20$	
Variable	$h = 1$	$h = 2$	$h = 3$	$h = 1$	$h = 2$	$h = 3$	$h = 1$	$h = 2$	$h = 3$
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 Naive	1.002	1.013	1.019	1.001	1.035	1.016	1.001	1.036	1.013
	(0.872)	(0.179)	(0.391)	(0.710)	(0.717)	(0.430)	(0.694)	(0.756)	(0.462)
		S_1			S_2			S_3	
		$R = 72$			$R = 72$			$R = 72$	
Variable	$h = 1$	$h = 2$	$h = 3$	$h = 1$	$h = 2$	$h = 3$	$h = 1$	$h = 2$	$h = 3$
EPU SVM	1.001	1.007	1.005	**0.999**	1.016	1.009	**0.999**	1.016	1.008
	(0.677)	(0.707)	(0.459)	(0.128)	(0.804)	(0.692)	(0.122)	(0.786)	(0.644)
EPU Mod	1.011	1.019	1.003	1.001	1.004	1.031	1.002	1.003	1.032
	(0.863)	(0.836)	(0.216)	(0.648)	(0.532)	(0.637)	(0.662)	(0.495)	(0.689)
EPU Naive	1.013	1.055	1.021	1.001	1.017	1.009	1.001	1.017	1.008
	(0.180)	(0.855)	(0.855)	(0.460)	(0.871)	(0.801)	(0.444)	(0.862)	(0.761)

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 give strong support for this claim and indicate that national uncertainty was partly influenced 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 higher 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, significantly improves the forecast accuracy when predicting the OLO-Bund spread, long term government bond yield, CDS spread and consumer confidence in the short term. On average, predictive power seems to break down when predicting for more than two months ahead, with only the CDS spread that can be predicted three months ahead. The case study presented in this paper shows a possible application of big data analytics in 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.

7. Acknowledgements

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.

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*, 7–16.

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 & 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 & Economic Statistics*, *33*, 1–9.

Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & 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.

Knotek II, E. S., & Khan, S. (2011). How do households respond to uncertainty shocks? *Economic Review-Federal Reserve Bank of Kansas City*, (p. 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. `http://http://www.nber.org/cycles/cyclesmain.html`. Accessed on 2016-08-17.

Neisingh, P., & Stokman, A. C. (2013). What drives consumer confidence in times of financial crises? evidence for the netherlands.

Palmer, F. R. (2001). *Mood and modality*. Cambridge University Press.

Rossi, B. (2013). Chapter 21 - advances in forecasting under instability. In *Handbook of Economic Forecasting* (pp. 1203 – 1324). Elsevier volume 2, Part B.

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 & 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.

in't Veld, J. (2013). Fiscal consolidations and spillovers in the euro area periphery and core. European Economy Economic Papers 506.

Weiss, S. M., Indurkhya, N., & Zhang, T. (2010). *Fundamentals of predictive text mining* volume 41. Springer.

Appendix A

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				