

# Belgian Economic Policy Uncertainty Index : Improvement through text mining

Ellen Tobbyack<sup>a</sup>, Walter Daelemans<sup>b</sup>, Enric Junqué de Fortuny<sup>a</sup>,  
Hans Naudts<sup>\*c</sup> and David Martens<sup>a</sup>

<sup>a</sup>Department of Engineering Management, University of Antwerp

<sup>b</sup>Department of Linguistics, University of Antwerp

<sup>c</sup>DG ECFIN, European Commission

March 25, 2014

## Abstract

Economists argue that high levels of policy uncertainty have been the reason for the slow economic recovery from the great recession of 2007-2009. Recently, an attempt has been made to measure economic policy uncertainty using forecast disagreements and news references, resulting in the frequently mentioned ‘Economic Policy Uncertainty-index’. In the original setup, a news article is assumed to address policy uncertainty if it contains certain predefined keywords. In this paper we try to improve the news part of the EPU-index using text mining techniques. We compare the original method to modality annotation and Support Vector Machines classification to create an EPU-index for Belgium. These three methods are compared to consumer and producer uncertainty indices and the ten year OLO-Bund spread. The correlation with the 10-year OLO-Bund spread demonstrates that the more advanced methodology has a higher explanatory power and speaks in favour of using an SVM classification model when constructing a news based policy uncertainty indicator. Since the crisis, it has been clear that negative economic spillovers within the euro area are much stronger than previously anticipated and the exact mechanisms behind these spillovers are still not fully identified. In this context we explore what part of policy uncertainty in Belgium is caused by contagion in the euro area.

---

\*The views expressed are the authors alone and do not necessarily correspond to those of the European Commission.

# 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; and 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 (Kose and Terrone, 2012; Bloom, 2007), dispersion in forecasts by professional forecasters or in expectations of consumers or producers (Bachman et al, 2010), or the prevalence of terms such as economic uncertainty in the media (Baker et al, 2013). In this paper we focus on the third methodology and contribute to the economic literature by using the latest state-of-the-art text mining methods to construct uncertainty indicators. This methodology allows us to identify the main factors with which uncertainty is associated.

Recently, Baker, Bloom and Davis (2013) have constructed an Economic Policy Uncertainty index (EPU) as a proxy for movements in policy related economic uncertainty over time. This index combines the frequency of newspaper references to EPU with the deviation of future inflation expectations. 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 (2013) can thus cause a high rate of both false positives and negatives.

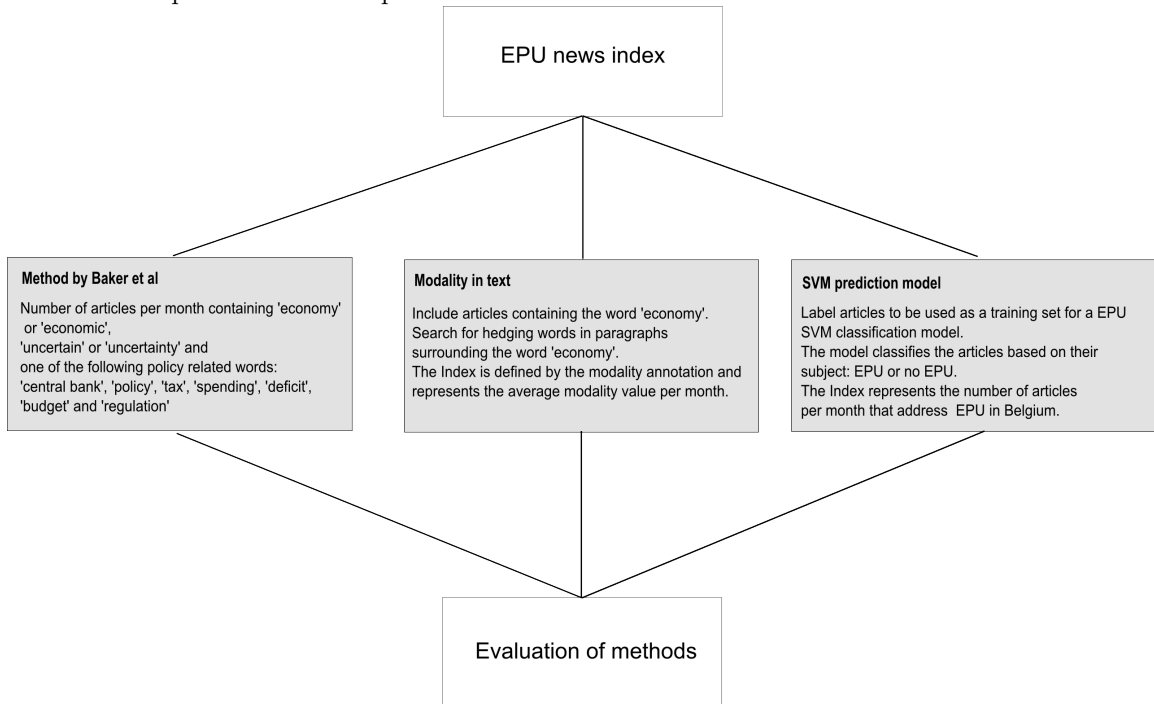
In this paper we attempt to improve this methodology by solving its main issues using text mining. Text mining is the process of deriving high quality information from text documents using techniques from data mining, statistics, information retrieval, machine learning and computational linguistics. (Weiss et

al, 2010) We apply two different text mining algorithms to a data set of approximately 210.000 articles: modality annotation and a Support Vector Machines 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. Figure 1 illustrates the scope of our research. Conform to Baker et al (2013), 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. Therefore it is interesting to explore whether there exist interlinkages and spillovers in economic policy uncertainty within the euro area seen from a Belgian perspective.

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 news articles, using state-of-the art text mining techniques 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 terms are most frequently related to policy uncertainty in the news articles. Finally, this is the first study that estimates an economic policy uncertainty index for Belgium while taking into account spillovers within the euro area.

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 compares the constructed indicators to more traditional uncertainty indicators and discusses spillovers in the eurozone. Finally, Section 5 concludes the paper.

Figure 1: Scope of our research: we improve the economic policy uncertainty index by Baker et al, with two text mining techniques, leading to different samples and different performances.



## 2 Naive method

We will compare our adjustments to the basic technique, as developed by Baker et al(2013). 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 web crawler, we searched for articles containing the said keywords 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'. Per month and per news source, we counted the number of articles containing the aforementioned queries. 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 (2013).<sup>1</sup>

Figure 2 plots the resulting EPU-index for a period of thirteen years, the time period over which articles are found in at least two newspapers. Note the peaks in uncertainty during the dot-com crash, the banking crisis and the fear of a Greek default. According to this index, Belgian uncertainty is influenced by financial uncertainty and U.S. uncertainty, which is a logical consequence of the selection method of articles. 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.

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. Table 1 reports the performance results of the naive method on an out-of-sample test set. We have labeled 100 articles according to their relevance. An article received class 1, if it addresses EPU and 0 otherwise. The method has a low recall, indicating a large false negative rate. 16 out of the 17 positive labeled articles were wrongly predicted as negative. This means that the naive method underestimates the EPU-index. The false positive rate is quite low, 2 out of the 83 negative labeled articles were wrongly predicted as positive. The method predicted only 3 articles as positive and two out of these three predictions were wrong. It is clear that the naive method is prone to overlook relevant articles. According to this method, an article is considered to be relevant if it contains three self-selected keywords. Articles in which the uncertainty is described using different words or in which the uncertainty is implied, though not explicitly mentioned, are not included in the index. We try to solve this major flaw using two different text mining techniques: modality annotation and text classification.

---

<sup>1</sup>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.

Figure 2: Economic Policy Uncertainty Index created using the naive method

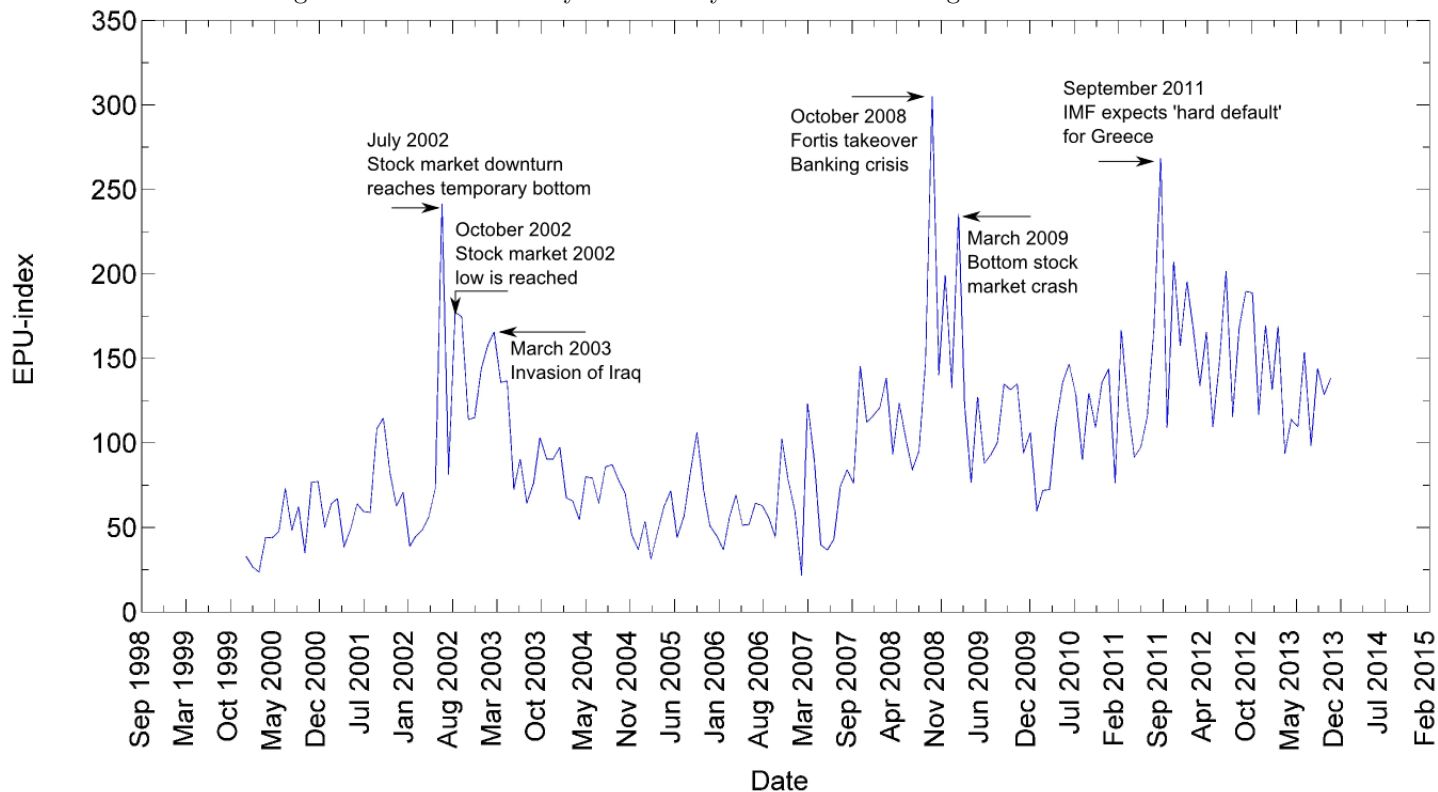


Table 1: Results on the out-of-sample test set for the naive method

Accuracy	Recall	Precision	Specificity	Fall-out	False negative rate
0.82	0.06	0.33	0.98	0.024	0.94

### 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

Table 2: Results on the out-of-sample test set for modality annotation

Accuracy	AUC	Recall	Precision	Specificity	Fall-out	False negatives rate
0.76	0.65	0.24	0.27	0.87	0.13	0.76

et al, 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. 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 eurozone to influence Belgian economic (policy) uncertainty, we believe that U.S. and Asian uncertainty does not directly affect policy uncertainty in Belgium. Uncertainty 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 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 2 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 modal-

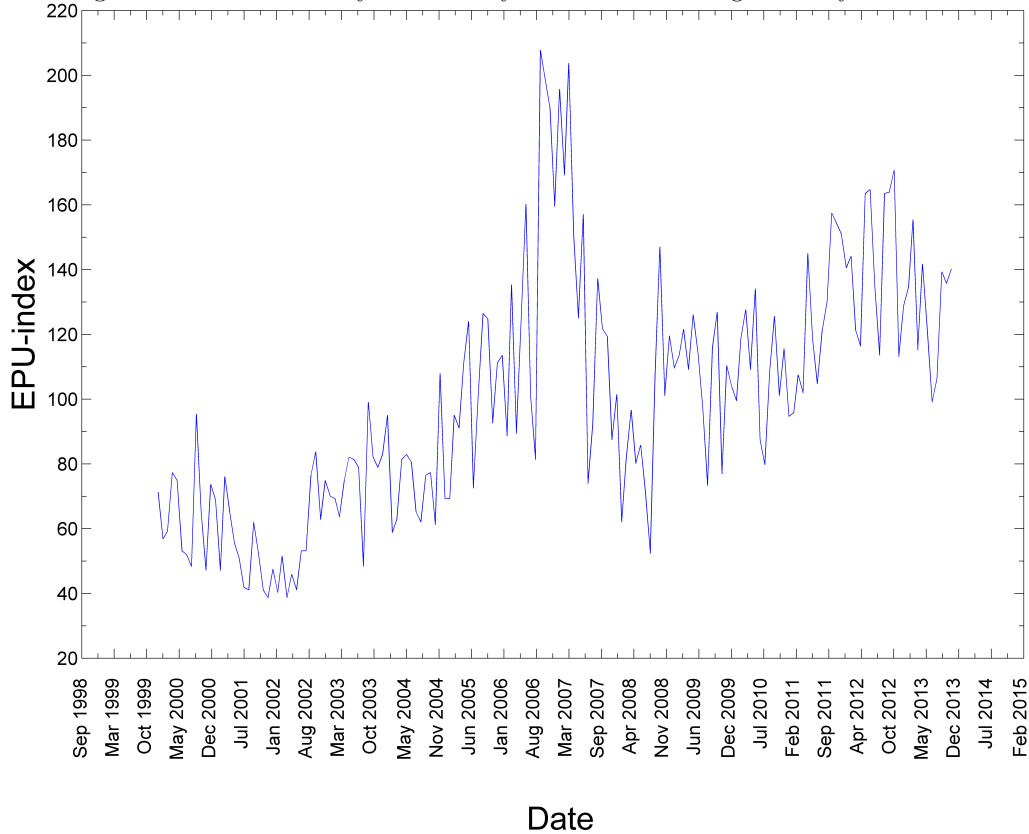
ity as a baseline system. Different improvements are possible that would make the measure more precise. Modality markers have a ‘scope’ (a number of words they apply to). Taking a complete article and counting the uncertainty markers in it, is a course-grained approach that could be improved by using systems that compute the scope of the modality more accurately, based on syntactic analysis (Morante and 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 (+- 150.000 articles) of articles that refer to either Belgium or a European country. 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 authors.

Figure 3 shows the EPU-index created using modality annotation for a period of thirteen years. There appears to be a large peak in uncertainty in 2006-2007, which was caused by the municipal and federal elections taking place during this period. Note that there is an upward trend over the years.



Figure 3: Economic Policy Uncertainty Index created using modality annotation



### 3.2 Text classification

The original method developed by Baker et al (2013) 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) - a state-of-the art classification method - to classify the news articles. This 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 0 (no EPU). (Fan et al, 2008)

In a first attempt, we labeled 500 articles randomly selected from the en-

tire 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 one if the article addresses economic policy uncertainty in Belgium and/or the eurozone and a value of zero otherwise. We define the first group of articles as the relevant articles. When constructing the Belgian EPU-index, we included uncertainty in the eurozone 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. economic policy uncertainty, since U.S. uncertainty is not expected to directly affect Belgian policy.

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  $[t_1 t_2 \dots t_i \dots t_n]$  that contains all  $n$  unique words present in the training set, where  $t_i$  denotes how often the  $i^{th}$  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 words and  $m$  the number of articles. In the term-frequency matrix each cell  $t_{ij}$  indicates the number of times the term  $i$  occurs in article  $j$ . 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)

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

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):

$$\min_w \frac{1}{2} w^T w + C \sum_i^n \max(1 - y_i w^T x_i, 0) \quad (1)$$

With  $w$  the weights of the model, and  $x_i$  and  $y_i$  representing the input vector and label of the  $i$ th observation. An out-of-sample grid search was performed to find the optimal value of  $C$ , the cost parameter.

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 labeling an additional randomly selected set of articles, which is an cumbersome and expensive process, we have opted for a pool-based active learning algorithm with uncertainty sampling. In this procedure, the active learner has access to an unlabeled 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 and Koller, 2001)

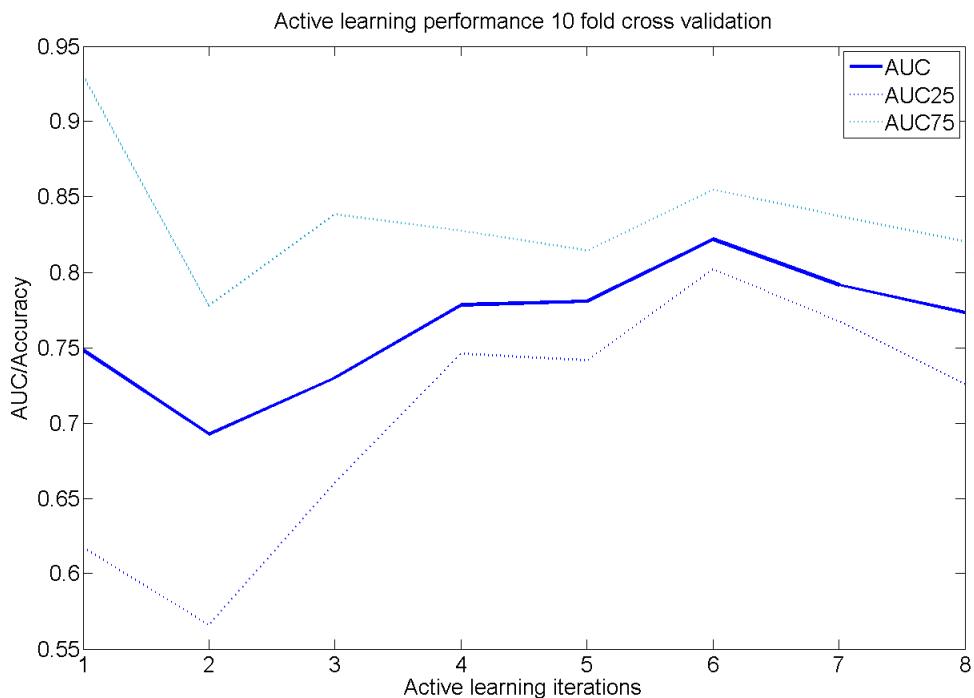
We started with the 400 randomly selected articles as training set to construct an SVM classifier, which defines the decision value for each article. The sign of the resulting decision value is the predicted class the article belongs to. The larger the decision value, the more certain the classifier 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 labeled and added to the training set. The active learner thus requests information about the instances it is the least certain about. By labeling 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 four times, in the end a total of 400 articles was added to the training set in top of the first 400 randomly selected articles. We looked at the AUC<sup>2</sup> of the model calculated using ten-fold cross validation to decide when to stop the active learning process. Figure 4 shows the AUC for each training set. The accuracy curve levels off after the sixth addition of articles. Since no further improvement in AUC was found, we decided to stop at the eighth iteration. Table 4 reports the classification results of all three methods, calculated on the test set. Comparing the different methods, we conclude that the prediction model outperforms the naive method and modality annotation on the test set of 100 labeled articles.

In the naive method by Baker et al (2013), 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. The word *cloud* in Figure 5 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

---

<sup>2</sup>The Area Under the ROC Curve (AUC) is the standard evaluation metric for classification models, and measures to what extent positively labeled observations are ranked higher than negatively labeled observations (Fawcett, 2006).

Figure 4: Evolution of AUC on the test set during the active learning process



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 word cloud 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.

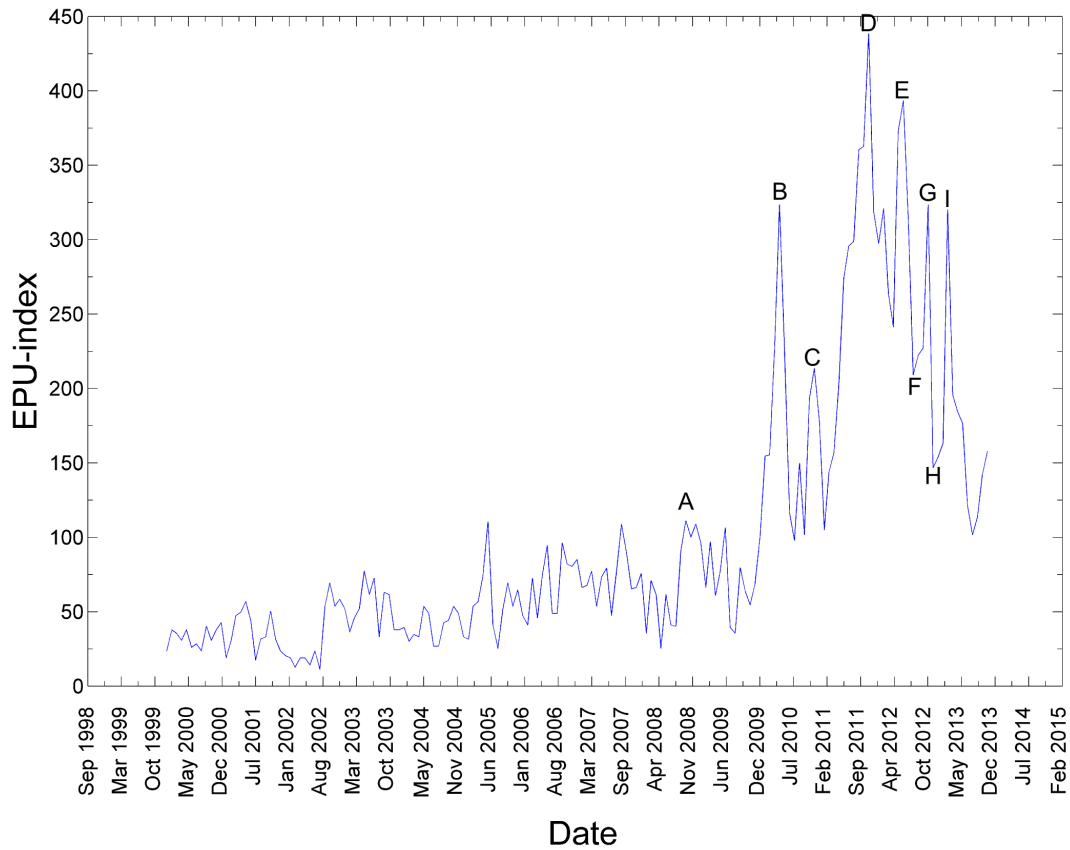
To assess the influence of uncertainty in the euro area on Belgian economic policy, we have investigated the causes of the peaks in the EPU-index created using the classification model. Figure 6 shows the EPU-index for a period of thirteen years and Table 3 lists the events responsible for the most noticeable peaks and troughs in uncertainty. It seems that uncertainty in Belgium is to a large extent influenced by spill-overs.

Figure 5: Most discriminating words to predict the Belgian EPU index according to the trained SVM classifier.



There still remain two issues that cannot be solved by this methodology. First of all, we assume Baker et al (2013) were right in their assumption that a news index can represent policy uncertainty. This can be tested by comparing our indicator with other uncertainty indicators which we do in section 4. Further work could also test whether the uncertainty indicator would be significant in testing economic relationships predicted by economic theory, e.g. in investment equations. Secondly, we start from a data set of articles containing the keyword ‘economy’, thereby assuming that articles without the occurrence of the word ‘economy’ do not address economic policy uncertainty. We had to restrict ourselves to these articles to limit the time spent on labeling 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 labeled before having enough positive examples. Therefore, both with our methodology and the naive methodology, the false negatives rate is presumably larger than reported.

Figure 6: Economic Policy Uncertainty index created using the SVM classification model



## 4 Uncertainty and contagion

### 4.1 Correlation to other uncertainty indicators

Figure 8 plots all three policy uncertainty indicators on the same graph. 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. The naive index spikes during the dot-com crash, the global financial crisis and in 2011 triggered by European default fears. Two

Table 3: Legend to figure 6

Letter	Date	Event
A	October 2008	Fortis Takeover / Banking crisis
B	April/May 2010	Greek bailout
C	November/December 2010	Irish bailout / Debate on EFSF / Greece austerity vote
D	October/November 2011	Referendum Greece / Forced resignation Berlusconi Nationalisation Belfius (former Dexia) End government negotiations
E	June 2012	Bank bailout in Spain / New elections in Greece Cyprus requests eurozone bailout
F	August 2012	Announcement OMT programme
G	November 2012	Renewed worries about Greece's debt crisis Inner cabinet meetings on draft state budget in Belgium Recapitalisation Dexia
H	December 2012	Greek bond buyback / Agreement on banking union
I	March 2013	Cyprus bailout / Italian general election / Belgian budget control

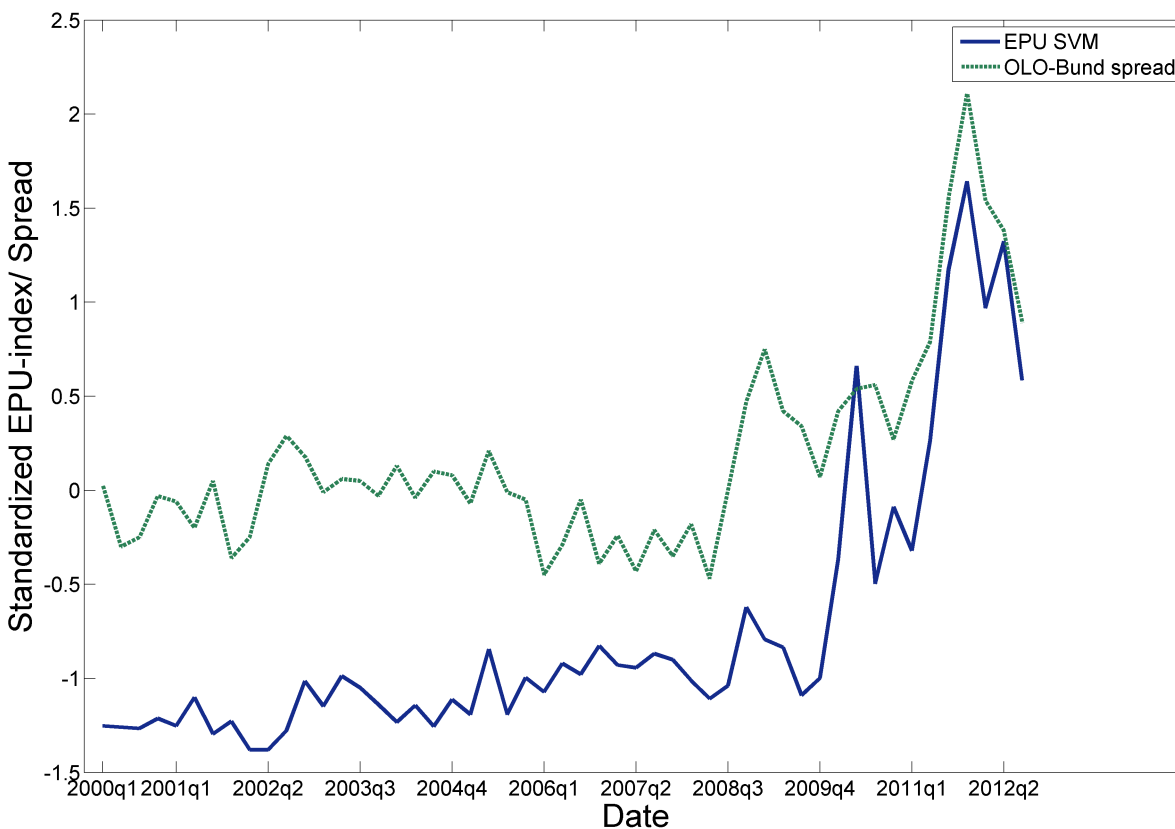
out of the three peaks in this index are caused by events that originated in the United States and lead to a global recession, affecting the Belgian economy along the way. The SVM index shows high volatility during the European and national debt crisis, indicating the alternation of agreement and disagreement inherent to every policy crisis.

While further work is needed to explore the economic uses of the constructed indicators, first results indicate that the SVM uncertainty indicator for Belgium has a correlation of up to 88% with the spread on Belgian 10-year government paper with the German Bund. Using an indicator based on the naive methodology the correlation is only 66%. This seems to demonstrate that the more advanced methodology has a much higher explanatory power. The SVM indicator appears to be better suited for spreads analysis since it is influenced by Belgian and European events and not by the world wide - though not euro area specific- shocks that influence the naive indicator. Figure 7 plots the SVM EPU-index and the 10-year OLO-Bund spread.

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

Method	Accuracy	AUC	Recall	Precision	Specificity	Fall-out	False negatives rate
Naive Method	0.82	\	0.06	0.33	0.98	0.024	0.94
Modality annotation	0.76	0.65	0.24	0.27	0.87	0.13	0.76
SVM classification	0.96	0.97	0.76	1	1	0	0.24

Figure 7: Rescaled EPU-index created using the SVM classification model and the 10-year OLO-Bund spread



To investigate to what extent our EPU-indices (dis)agree with more classic uncertainty indicators, we calculated the correlation between our policy un-



certainty indicators and indicators of uncertainty for industry and consumers derived from surveys conducted by the European Commission.<sup>3</sup> We have selected two consumer uncertainty indicators, which represent the divergence of responses to the following two questions from the survey:

- *Q2 : How do you expect the financial position of your household to change over the next 12 months?*
- *Q4 : How do you expect the general economic situation in this country to develop over the next 12 months?*

Next to the consumer uncertainty indicator, we compare our indicators to an indicator of producer uncertainty. This indicator represents the divergence of the industry responses to the following question:

- *Q5: How do you expect your production to develop over the next 3 months?*

Table 5 reports the correlations for all three EPU indicators. The producer uncertainty indicator shows the highest correlation to the modality index, the Q2 consumer uncertainty indicator to the naive index and the Q4 consumer indicator to the SVM index. Figures 9 to 11 plot the constructed indicators to the survey indicators. The highest peak in the Q2 consumer uncertainty indicator corresponds to the highest peak in the SVM index and seems to be triggered by a convergence of various events that evoked uncertainty, as can be seen in Table 3.

Taking a closer look at our final SVM EPU-index, it appears that the indicator combines parts of consumer and producer uncertainty, while adding new periods of uncertainty. In April/May 2010 the peak in policy uncertainty is accompanied by a peak in producer uncertainty. During the global financial crisis, policy uncertainty was located somewhere in between producer and consumer uncertainty. Producer uncertainty seems to be more pronounced during the dot-com bubble and crash and the transition of the banking crisis into the debt crisis. Being highly correlated to the 10-year OLO-Bund spread and indicating a different evolution in uncertainty than the traditional indicators, our SVM EPU indicator can provide additional knowledge when forecasting or nowcasting macro-economic indicators.

---

<sup>3</sup>More information on these indicators can be found in DG ECFIN (2013).

Figure 8: EPU-index of all three methods, rescaled to average value of 100 after standardisation

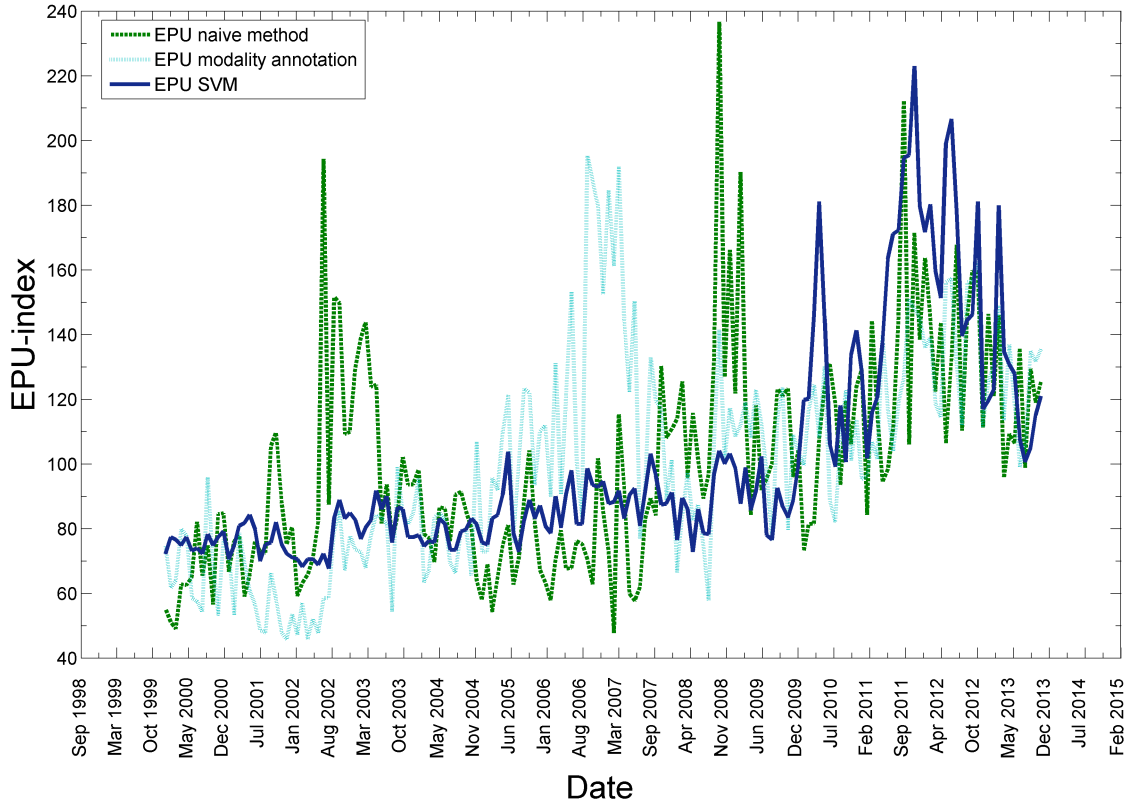


Table 5: Correlations between EPU indicators and consumer and producer uncertainty indicators

	EPU SVM	EPU modality	EPU naive
Producer uncertainty	0.45	-0.033	0.539
Consumer uncertainty (Q2)	0.203	0.051	0.214
Consumer uncertainty (Q4)	0.603	0.267	0.637

Figure 9: Comparison of the EPU-index created using the naive method with the uncertainty indicators

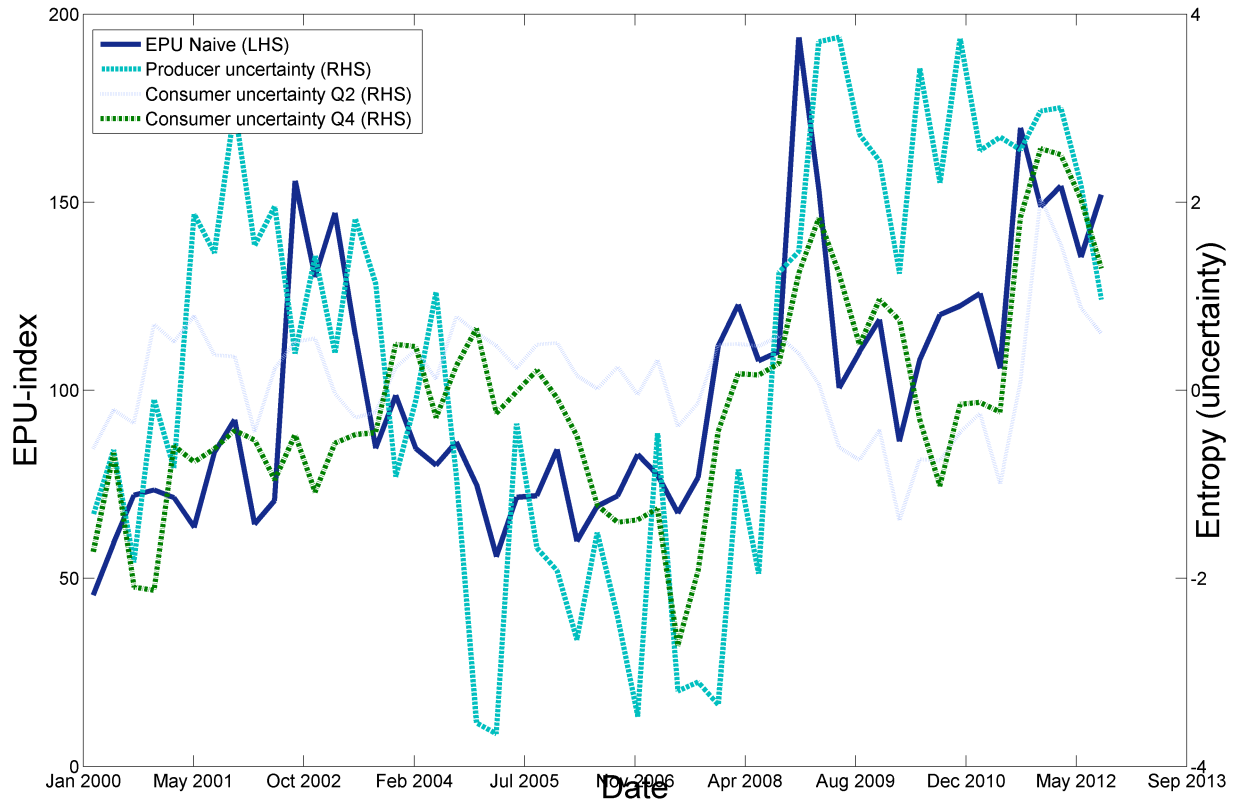


Figure 10: Comparison of the EPU-index created using modality annotation with the uncertainty indicators

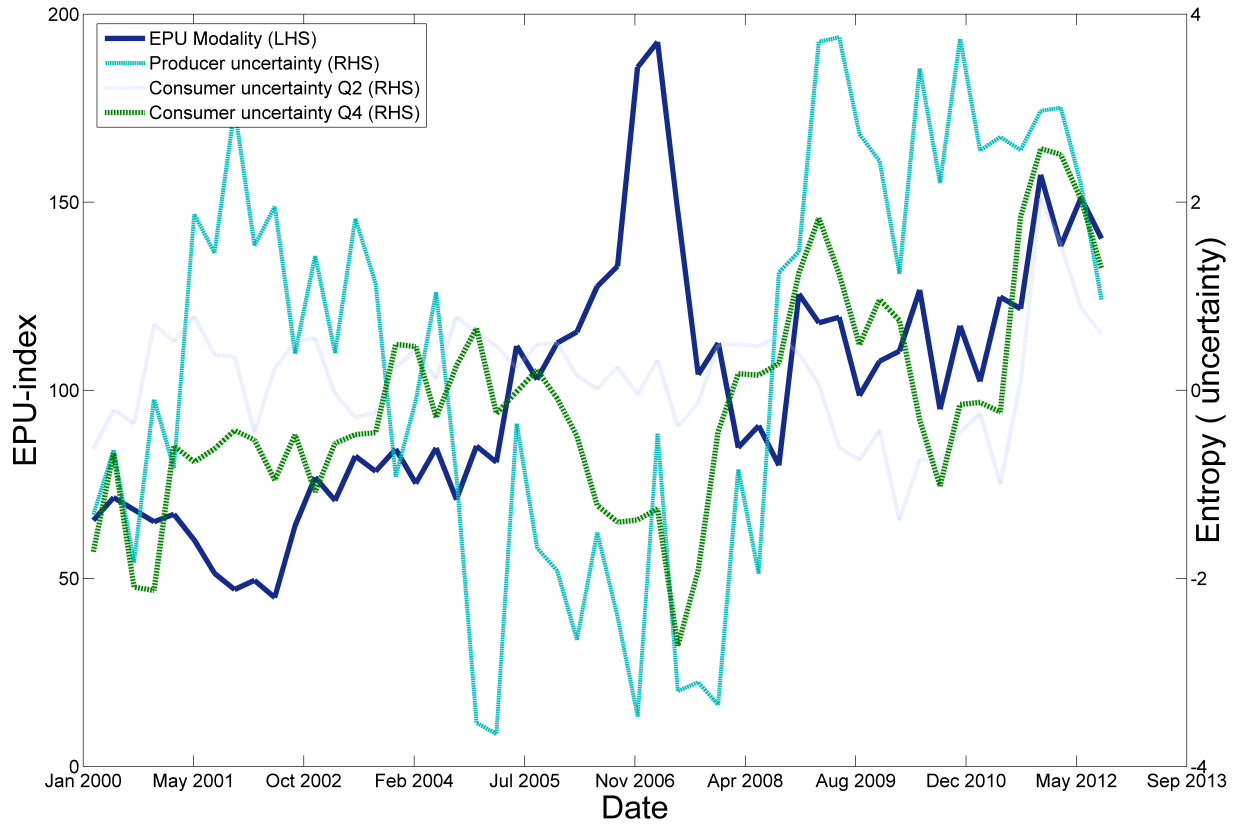
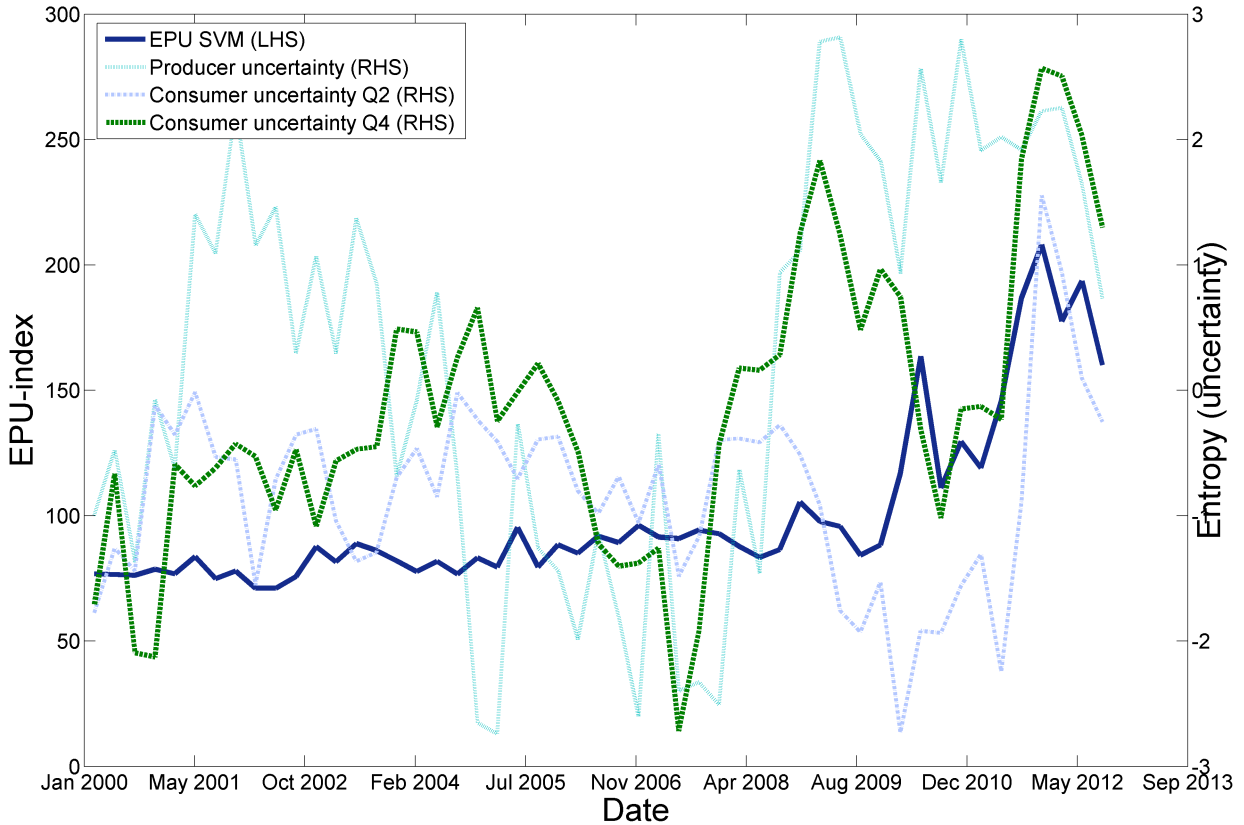


Figure 11: Comparison of the EPU-index created using the SVM classification model with the uncertainty indicators



## 4.2 Contagion

This paper makes a contribution to the literature on spillover effects of uncertainty by identifying which factors are associated with uncertainty. By using an SVM model the words that are most strongly associated with uncertainty in Belgium are identified. These words can be grouped into a category Belgian (with words such as Rupo), Euro area (with words such as ECB, eurozone, Trichet and Rehn), spill-over (with words such as Greek and Cypriot) and other (with words such as debt). From the word-cloud and the peaks in the EPU-index, we can conclude that policy uncertainty in Belgium is to a large extent influenced by euro area and spillover effects. The results confirm the importance of con-

tagion within the euro area where policy uncertainty was not only stemming from national decision making but was - and still is - also determined by policy uncertainty at the euro area level with uncertainty being linked across the region. Firstly, the ECB's monetary policy and the effectiveness of monetary policy transmission can be a source of policy uncertainty at the euro area level. Secondly, uncertainty on the outcome of the EU's decision making process, for example related to firewalls or banking union, has been often cited as a source of policy uncertainty. At the heat of the sovereign debt crisis, the risk of an outright break-up of the euro area fueled this uncertainty and sovereign yields were priced higher to account for the risk of convertibility (Viita, 2012 and ECB, 2012). Even after the OMT decision of the ECB, which is seen by many observers as a turning point in the crisis, euro area related uncertainty continued to linger. For example according to the IMF, uncertainty surrounding prospects and policies for the euro area were weighing significantly on activity in Germany with policy uncertainty at the euro area level being one of the most important reasons for German SMEs to postpone investments and adopt a wait and see attitude in 2013. (IMF, 2013)

There is an emerging empirical literature on spillovers of confidence and uncertainty within the euro area. DG ECFIN (2012) demonstrates that euro area-wide policy uncertainty, as measured by Baker et al (2013), and a common euro area risk factor affect sovereign bond risk premia. D'Auria (2013) demonstrates that consumption and measures of consumer confidence appear to be strongly correlated in the euro area. Moreover, confidence indicators display similar dynamics between the core and the periphery of the euro area. Balta et al (2013) describe the different developments of country uncertainty. For consumer uncertainty, distinct trends between 'core' and 'periphery' countries can be traced back to 2002. Consumer uncertainty increased for both country groups in the very early stages of the crisis, before decreasing again in 2009. Since 2010, uncertainty has been on an increasing trend for both country groups. While uncertainty is still low for Germany, uncertainty in the periphery is at record-high levels and remains high in France from a historical perspective. Uncertainty in Belgium seems to remain quite elevated as well, as can be seen from Figures 9 to 11.

## 5 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 un-

certainty 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. The correlation with the 10-year OLO-Bund spread demonstrates that the more advanced methodology has a higher explanatory power and speaks in favour of using an SVM classification model when constructing a news based policy uncertainty indicator. The correlations between our constructed uncertainty indicators and more traditional consumer and producer uncertainty indicators, imply that the policy uncertainty indicators represent an extra dimension of uncertainty. From our three indicators we can gain more insights on the interaction and sources of uncertainty. Future research is needed to examine the predictive power of the constructed indicators when forecasting or nowcasting the macroeconomy.

## Acknowledgements

Ellen Tobback acknowledges the Flemish Fund for Scientific Research for financial support.

## References

- [1] Bachman, R., S. Esltner and E. Sims (2010). Uncertainty and economic activity: evidence from business survey data. *NBER Working Papers*, 16143.
- [2] Baker, S., N. Bloom and S. Davies (2013). Measuring economic policy uncertainty. *Chicago Booth Research Paper*, 13-02.
- [3] Balta, N, Valds Fernndez I. and Eric Ruscher E. (2013). Assessing the impact of uncertainty on consumption and investment. *Quarterly report on the euro area*, 12, 2.
- [4] Bloom, (2007). The impact of uncertainty shocks. *NBER Working Papers*, 13385.
- [5] D’Auria, F (2013). Cross-border spillovers in confidence. *Quarterly report on the euro area*,12, 3.
- [6] DG ECFIN (2012). Risk and uncertainty in euro area sovereign debt markets and their impact on economic activity. *Quarterly report on the euro area*, 11, 4.
- [7] ECB (2013). Speech by Mario Draghi, President of the European Central Bank at the Global Investment Conference in London. 26 July 2012.

- [8] Fan, C., Chang, K., Hsieh, C., Wang, X. and Lin, C. (2008). LIBLINEAR: A library for large linear classification. *Journal of Machine Learning Research*,9,1871-1874.
- [9] Farkas, R., Vincze, V., Mra, G., Csirik, J., and 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.
- [10] Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*,27(8),861-874.
- [11] Int Veld J (2013). Fiscal consolidations and spillovers in the Euro area periphery and core. *European Economy Economic Papers*, 506.
- [12] Knotek, E.S. and S. Khan (2011). *How do households respond to uncertainty shocks?*. Economic Review, Federal Reserve Bank of Kansas City, issue QII.
- [13] Kose, M., and M. Terrone (2012). How does uncertainty affect economic performance?. *World Economic Outlook*, Box 1.3, pp. 4953.
- [14] Palmer, Frank Robert (2001). *Mood and modality*. Cambridge University Press.
- [15] Roser,M. and 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.
- [16] Tong, S. and Koller,D. (2001). Support Vector Machine Active Learning with Applications to Text Classification. *Journal of Machine Learning Research*,45-66.
- [17] Viita, K. (2012). *Euro Area is running significant risk of breakup: Rehn*. Bloomberg online news article, 1 June 2012.
- [18] Weiss, S.M., Indurkha, N. and Zhang, T. (2010). *Fundamentals of predictive text mining*. London: Springer.