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

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.

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).

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.