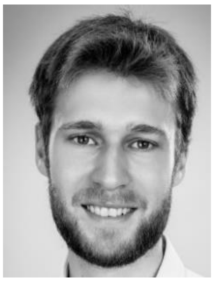
THE ECONOMIC POLICY UNCERTAINTY INDEX FOR FLANDERS, WALLONIA AND BELGIUM[1]

Andreas Algaba
PhD Student
Faculty of Social Sciences & Solvay Business School, VU Brussels
Department Economics, Universiteit Gent



Samuel Borms
PhD Student
Insitute of Financial Analysis, Université de Neuchâtel
Faculty of Social Sciences & Solvay Business School, VU Brussels

Kris Boudt
Professor of Finance & Econometrics
Department of Economics, Universiteit Gent
Faculty of Social Sciences & Solvay Business School, VU Brussels
School of Business & Economics, Vrije Universiteit Amsterdam

Jeroen Van Pelt
PhD Student
Faculty of Social Sciences & Solvay Business School, VU Brussels

ABSTRACT

This research note describes the construction of news-based Economic Policy Uncertainty (EPU) indices for Flanders, Wallonia and Belgium. The indices are computed from January 2001 until May 2020. Important domestic and more global events coincide with spikes in the indices. The COVID-19 pandemic represents the highest point, reflecting very strong consecutive Belgian newspaper attention to economic policy uncertainty. The monthly values of the EPU indices for Flanders, Wallonia and Belgium are published on www.policyuncertainty.com.

[1] We thank David Ardia, Scott Baker, Nick Bloom, Keven Bluteau, Steven Davis, Jolan Stevens, and Jan Wijffels for helpful feedback. We are grateful to the Belga News Agency for providing the news data. This project received financial support from the National Bank of Belgium, the Swiss National Science Foundation (grant #17928), and Innoviris.

1. Introduction

In their seminal work, Baker et al. (2016) propose a novel approach for measuring economic policy uncertainty (EPU) based on newspaper coverage. The coverage is measured by words related to the economy, policy, or uncertainty. The normalized volume of news articles containing these words are considered a good measure of the prevailing underlying uncertainty regarding economic policy. Their focus is on the construction of an index for the U.S., from the country's ten largest newspapers.

Since, many have applied this methodology to construct text based EPU indices for various geographies. For instance, Kroese et al. (2015) create an EPU index for the Netherlands, and Ghirelli et al. (2019) do so for Spain. Some have tried to improve on the methodology. Tobback et al. (2018) attempt to develop an EPU index for Belgium using, among others, support vector machine classification. However, their initiative to regularly publish index updates fell short due to data scraping issues. Azqueta-Gavaldon et al. (2020) extend the original methodology by employing two powerful machine learning methods (word embeddings and topic models).

This note presents three EPU indices constructed for Belgium based on a large archive of news articles. Belgium is situated in the very centre of Europe, split into the Flanders, Wallonia, and Brussels-Capital regions. The respective official languages are Flemish (a variant of Dutch), French, and both in Brussels. The press landscape is divided according to these two languages as well, allowing for a separate construction of two regional indices, before merging the two indices into a Belgium countrywide index.

The monthly EPU indices for Belgium correlate well with other uncertainty indicators. To label the sources of uncertainty, we present a technique to automatically extract non-EPU key terms in news. During 2020 and the COVID-19 pandemic, we reveal a clear daily build-up of news-based economic policy uncertainty.

2. Implementation

The generic methodological steps for the construction of monthly news-based EPU indices as proposed in Baker et al. (2016) are:

1. Select the newspapers of interest.
2. Count the number of newspaper articles containing at least one economic (E) keyword, one policy (P) keyword, and one uncertainty (U) keyword, in the native language of the newspaper in question. These are the raw number of EPU articles.
3. Scale the raw EPU count by a measure of the total number of articles in the same newspaper and month.
4. Standardize each newspaper-level monthly series to unit standard deviation prior to a certain date, and average across newspapers by month.

5. Normalize to a mean of 100 prior to a certain date. Perform another index scaling if deemed useful.
6. If applicable, average across a set of final indices to obtain an aggregated index (e.g. GDP-weighted per country, or, as in our case, per language).

Baker et al. (2016) emphasize that with each monthly update, data from the preceding (two) months may be revised marginally, driven by the fact that some newspapers do not immediately update their online archives with all articles.

The specific implementation requires choices in terms of news data provider, newspaper selection, keywords, and reference period for the standardization. Ghirelli *et al.* (2019) show the sensitivity of the index construction to the amount of newspapers considered, and the number of keywords. We accord with their guidelines by considering more than two newspapers and enlarging the set of keywords used.

2.1 Data

We obtain the news articles for Belgium from the national Belgian News Agency (Belga). Their archive contains over 40 million media news articles in Flemish and French starting from 2001 until now.

We include following eight Flemish newspapers: “De Tijd”, “De Standaard”, “De Morgen”, “Het Laatste Nieuws”, “Het Nieuwsblad”, “Gazet van Antwerpen”, “Het Belang van Limburg”, and “Het Volk.” The newspaper “Het Volk” seized activities in 2008 but is an historically important news source. We include following five Walloon newspapers: “L’Avenir”, “La Dernière Heure”, “La Libre Belgique”, “Le Soir”, and “L’Echo.”

We clean the news data by filtering out exact duplicate articles, and remaining articles of no relevance to economic policy uncertainty (e.g. sports or arts news). Near-duplicate entries are kept, as these are often the same publications but by different newspapers. We also trim too short (up to 450 characters) and too long (from 7500 characters) articles, as these are more sensitive to a biased measurement of EPU relevance. Between January 2001 and May 2020, we select around 109000 articles from the Flemish-speaking press, and around 81800 articles from the French-speaking press.

The database archive we use has no news available during 2006 for none of the Walloon newspapers. We decide to encode the resulting monthly index values as missing and carry this forward to the Belgium index. Alternatively, we could have performed interpolation. Simple linear interpolation across a full year would be uninformative. Proportional interpolation (filling in the missing values in the Wallonia series by mimicking the trend of the Flanders series) would suffer from lookahead bias as one series is based on future known values of another series.

It is also important to highlight that we exclude online newspaper content, for two main reasons. First, to have comparability of our data universe across time, as online publications have become growingly prevalent compared to the earlier years of the time period covered. Second, because web publications typically serve other (monetization) purposes and might therefore be different when it comes to the level of uncertainty expressed. It would also involve a more complex management of possible news duplicates. The differences between offline and online news content in light of uncertainty coverage could be an interesting study in itself.

2.2 Keywords

The entire set of EPU keywords decide on the selection of articles to take into account. The E and P categories are required to trim down the news to the correct topic, in this case reporting about economic policy, and the U category can be seen as the “sentiment” driver.

We start from the Dutch word list provided to us by Kroese et al. (2015), who based themselves on the original paper of Baker et al. (2016). Terms specific for the Netherlands are deleted. We also limit the keywords to single words (called unigrams). We use a pretrained word embedding space on a Flemish corpus to generate candidate words to make the Flemish keywords more comprehensive.[2] The expanded Flemish word list is translated to a French word list. Both lists of keywords are checked manually to omit remaining dubious entries.

2.3 Computation

We follow the computation and normalization approach as explained. We normalize the newspaper-level series to unit standard deviation for a reference period up to 2011 (i.e. divide each index by its standard deviation). As a final normalization, we bring the mean of the averaged series before 2011 to a level of 100 (i.e. divide the index by its mean and multiply by 100). The Belgium EPU index is a simple average of the resulting Flanders and Wallonia EPU indices.

3. Analysis

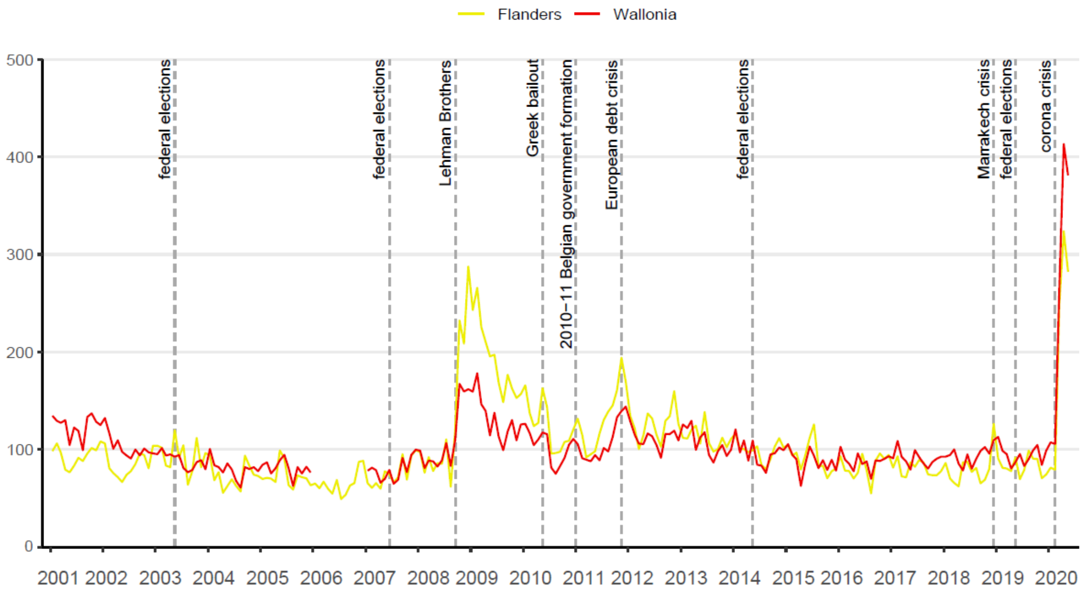
We validate the Belgian EPU indices in four ways. First, we plot the series and label important peaks. Second, we extract frequent terms of EPU articles around some of such peak periods. Third, we gauge the correlation between our indices and other related

[2] A word embedding space maps words into high-dimensional vectors. Similar words have a shorter distance between their vectors. The embedding we use is recycled from another project (hence, pretrained), and was fit with the GloVe technique (Pennington et al. 2014).

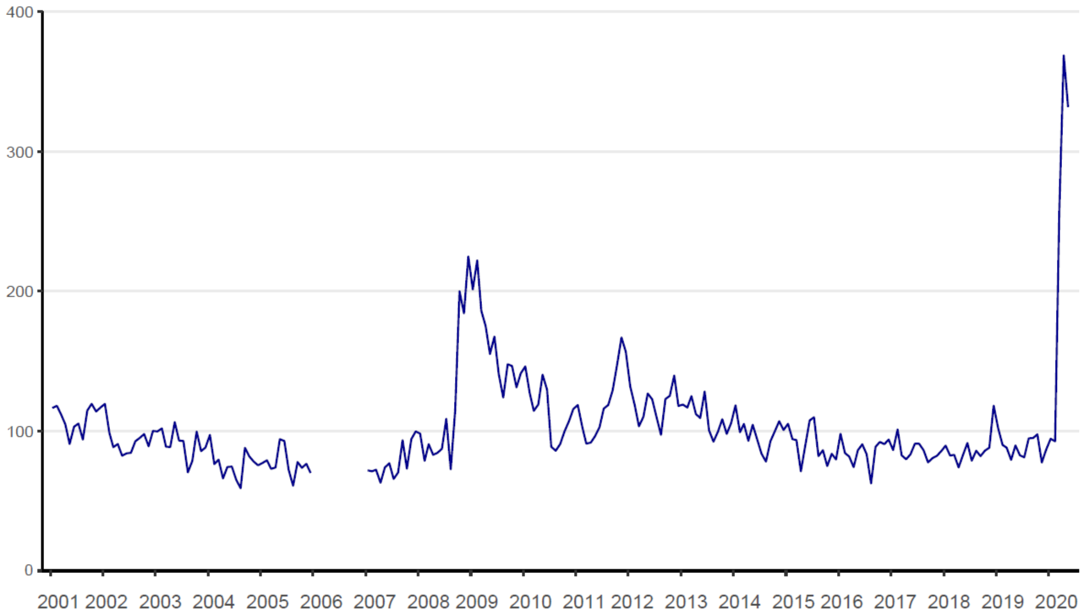
uncertainty time series. Fourth, we discuss alternative index creation methods. Additionally, in the last subsection, we dive deeper into the months covering the COVID-19 pandemic.

3.1 Belgian EPU Time Series and Events

Figure 1 displays the three final EPU indices for Belgium. Figure 1a compares the Flemish to the Walloon EPU index, including an annotation of noteworthy events, whereas Figure 1b plots the resulting index for Belgium. We perceive both peaks related to local events, and peaks related to European and worldwide events.



(a) Annotated chart of EPU index for Flanders and Wallonia.



(b) Aggregated EPU index for Belgium.

Figure 1: EPU indices for Flanders, Wallonia and Belgium.

The global financial crisis starting in 2007 up to 2009 comes with a clear peak in the EPU indices. The period around the European debt crisis as well, with two peak moments (one with the Greek bailout, and one at a later stage of the crisis). Around the first Brexit referendum (mid-2016) and after, the indices bear volatile but not particularly high values. The crisis related to the COVID-19 pandemic reaches a level higher than the financial crisis. The uncertainty also increased from March to April. The surge in uncertainty as a result of the pandemic is also documented by Baker et al. (2020). The situation is unprecedented and not limited to the financial sector but impacts the entire economy. From April to May, there is a small decrease.

There are peaks around federal elections, albeit minimal ones. More apparent is the increase in economic policy uncertainty when Belgium was on its way to obtain the world record of longest government formation during 2010 and 2011. More recently, in December 2018, the Belgian government fell after disagreements about endorsing the United Nations Marrakech migration pact. Economic policy uncertainty raised accordingly.

The proportional evolution of the number of detected keywords per category in observed EPU articles is fairly constant. In the two severest peak times (the global financial crisis, and the corona crisis), the uncertainty category becomes slightly more important, which implies that the indices then capture a relative increase in uncertainty as opposed to increased relative reporting about economic policy matters. However, during the European debt crisis over 2011, this is the opposite.

3.2 Key Terms Extraction for EPU Articles

To give a face to the sources of uncertainty, we apply a three-step technique to automatically summarize the main non-EPU (i.e. excluding the EPU keywords, except for the capitalized abbreviations such as ECB/BCE) terms found in news during a selection of peaks. First, we extract from the relevant corpus subset (here defined as comprising news during the month in which a peak occurs) the nouns and full proper nouns. Second, amongst these terms, we compute the co-occurrence frequency and continue with those sets of terms that co-occur at least five times. The first two steps are performed to maximize the informational content as the basis to form topics. Next, we apply the biterm topic model developed by Yan et al. (2013) with the above as input.

The generative process underlying biterm topic modeling does not consider individual news articles, but a reduction of the whole corpus as an aggregated biterm set. The entire corpus is seen as a mixture of topics. It are the word co-occurrence patterns across the corpus that convey the topics, not single words at the level of documents. Our version models in essence the noun/proper noun biterm (in any direction) co-occurrence relationships.

We infer six topics but drop the first cluster, which is set equal to the empirical word distribution to filter out the most common words. This gives for a given month five topics, which we each define by the ten words most related to it. A peak month is thus explained by at maximum 50 nouns or proper nouns.

We use the R packages **udpipe** (Wijffels 2019) and **BTM** (Wijffels 2020) to do the majority of the calculations. We present the full output of the news summary analysis in Table 1 and Table 3. Terms expected to pop up are indeed cited repeatedly in the news. Sometimes other subjects come up too, for instance, during the Marrakech crisis, news also discussed the ongoing developments about the Brexit and about the trade deal between the U.S. and China. One month after Lehman Brothers' bankruptcy, the U.S. presidential elections were coming up and also a heavy topic of discussion in the Flemish press.

Event	Cluster	Top non-EPU terms
Lehman Brothers (10/2008)	1	banken, dollar, miljard, bank, geld, landen, bedrijven, week, VS, markt
	2	bedrijf, werknemers, directie, bedrijven, productie, miljoen, maanden, vraag, stuk, week
	3	Fortis, bank, Dexia, BNP Paribas, banken, miljard, geld, België, aandeel, week
	4	geld, Leterme, land, miljoen, België, partij, tijd, bedrijven, banken, CD&V
	5	Obama, McCain, president, debat, campagne, Bush, VS, Republikeinen, Palin, Barack Obama
Marrakech crisis (12/2018)	1	Macron, land, Europa, Frankrijk, werk, migratie, president, Antwerpen, België, bedrijven
	2	Trump, China, president, VS, Huawei, land, Europa, wereld, vraag, Congo
	3	rente, bank, VS, Trump, groei, Fed, dollar, miljard, bedrijven, president
	4	N-VA, Michel, partij, CD&V, land, meerderheid, MR, motie, steun, vertrouwen
	5	May, deal, brexit, miljoen, Bpost, land, bedrijf, Europa, stemming, Brussel

Table 1: Automated display through biterm topic modeling of the most recurring non-EPU terms in Flemish press around some peak EPU events (cf. Figure 1a).

A more complex and bottom-up technique to analyze more precisely the uncertainty sources would be to use a regular topic model, as explained and carried out in Azqueta-Gavaldon (2017) and Azqueta-Gavaldon et al. (2020), for instance.

3.3 Belgian EPU Time Series and Related Indices

Table 2 exhibits the contemporaneous correlations between the three constructed EPU indices for Belgium and some other uncertainty indicators. As can also be seen in Figure 1a, the correlation between the Flanders (EPU^{FL}) and Walloon (EPU^{WL}) indices is strong at 79%, but not perfect. Some events are covered more in the Flemish press (e.g. the global financial crisis), others more in the Walloon press (e.g. the COVID-19 pandemic), but the patterns are similar. The keywords are very comparable (due to the translation), yet this does not guarantee ex-ante that the news coverage will be as well. The conclusion is thus meaningful.

	EPU^{FL}	EPU^{WL}	EPU^{BE}	EPU^{NL}	EPU^{DE}	EPU^{EU}	VIX^{NL}	VIX^{EU}	CCI^{BEL}
EPU^{FL}	1	0.79	0.96	0.65	0.37	0.26	0.51	0.54	−0.66
EPU^{WL}	0.79	1	0.94	0.62	0.45	0.28	0.37	0.39	−0.49
EPU^{BE}	0.96	0.94	1	0.64	0.42	0.26	0.47	0.49	−0.62

Table 2: Contemporaneous correlations between the Belgian EPU indices and various other EPU indices and uncertainty indicators.

The correlations with the EPU index of the Netherlands (EPU^{NL}), Belgium's most resembling country, are above 60%. The correlation of the Belgium EPU index (EPU^{BE}) with the German EPU index (EPU^{DE}) is a bit lower at 42%, and the correlation with the index for major European economies (EPU^{EU}) sinks to 26%. Figure 2 plots the benchmark EPU series and the one for Belgium. The most striking divergence took place during the initial Brexit struggles, when the EPU index for Europe (including a U.K. EPU index) and the EPU index for Germany soared severely more than the series for Belgium and the Netherlands did. The differences in the beginning of the time series are because of other normalization reference periods used.

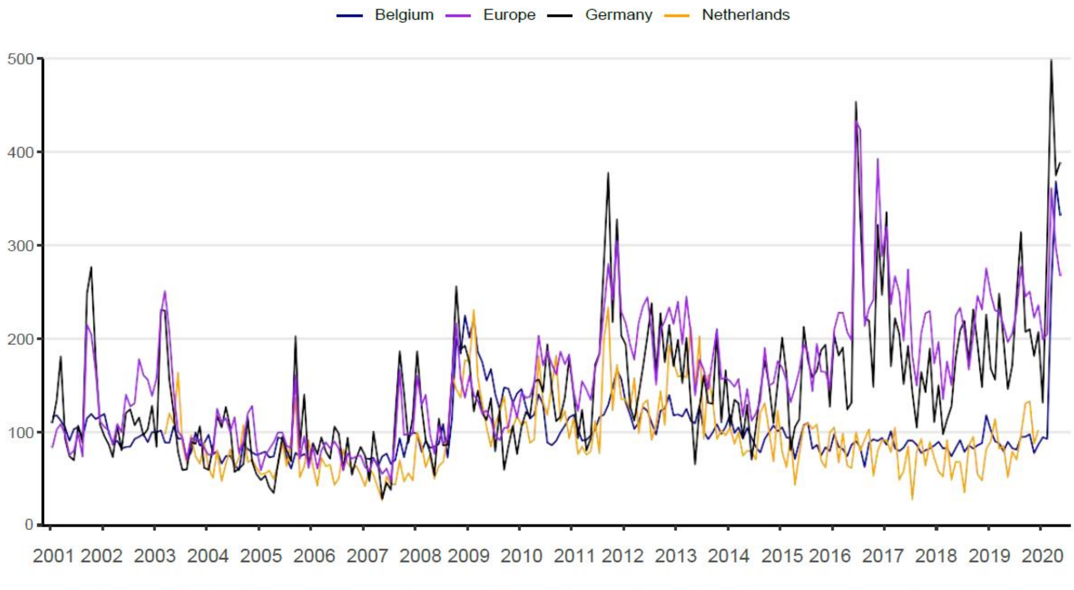


Figure 2: Monthly Belgium EPU index and benchmark EPU indices. *Note*: the EPU from the Netherlands is not available in the beginning and end of the sample.

We also include the VIX index on the Dutch AEX stock market index (VIX^{NL}), and the VIX index on the Euro Stoxx 50 stock market index (VIX^{EU}) in the comparison. The obtained correlations are all in the proximity of 50%. Lastly, we analyse the interaction with the consumer confidence indicator in Belgium (CCI^{BEL}). The correlations are strongly negative up to -66%, as anticipated. If economic policy uncertainty goes up, consumer confidence goes down, and vice versa.

Overall, the sign and strength of the correlations are in line with what other works have reported (such as Kroese et al. 2015), thereby corroborating a correct construction of the Belgian EPU indices. The indices reveal sufficient domestic and foreign affairs.

3.4 Alternative Index Construction Methods

As a robustness check, we tested several alternative index construction approaches. First, in terms of corpus selection. We tried a filtering that keeps a news article only if keywords from all three categories show up within a span of six sentences at least once. This halves the corpus for both languages. Contrary to Kroese et al. (2015)'s additional index confined to the Netherlands only (i.e. their "EBO-NL" index), we stick with the broader EPU index as-is. Adding a complementary filter that ensures news articles discuss Belgium impacted the corpus size too heavily. Second, in terms of computation, for instance, counting the raw number of EPU keywords instead of the normalized number of EPU articles.

The different index versions have a positive correlation with the indices coming from the original indexation approach, but do not result in more qualitatively interpretable indices. They tend to be more volatile and relate less well to the benchmark indices.

This supplementary analysis validates the effectiveness of our keywords and our corpus cleaning procedure. For the Belgian case, more stringent (and more time-consuming to obtain) news corpora, or other index measures, do not result in better EPU indices.

3.5 Economic Policy Uncertainty in Times of COVID-19

In this subsection, we briefly zoom in on the first five months of 2020 during which the COVID-19 crisis unfolded across the world. Table 3 (for Flemish press) and Table 4 (for Walloon press) show the key terms from the topic clusters obtained as above. The focus is on how the news coverage content changed from March to May, along with the high measurement of economic policy uncertainty.

Month	Cluster	Top non-EPU terms
03/2020	1	bedrijven, coronacrisis, aantal, werknemers, werk, banken, miljard, week, weken, miljoen
	2	virus, land, coronavirus, Trump, China, VS, president, tijd, Italië, wereld
	3	bedrijven, banken, miljard, coronavirus, geld, rente, landen, ECB, bank, impact
	4	N-VA, CD&V, PS, MR, noodregering, meerderheid, partij, Dewael, land, Laruelle
	5	Europa, Turkije, EU, Griekenland, grens, Erdogan, landen, Duitsland, president, steun
04/2020	1	bedrijven, miljoen, coronacrisis, bedrijf, miljard, maand, banken, werknemers, week, België
	2	week, N-VA, land, mei, weken, tijd, Veiligheidsraad, coronacrisis, CD&V, leven
	3	virus, land, landen, wereld, aantal, lockdown, China, Trump, leven, coronavirus
	4	landen, miljard, Italië, geld, Nederland, Europa, bedrijven, land, coronacrisis, EU
	5	miljoen, Brussels Airlines, coronacrisis, bedrijven, stad, vraag, Lufthansa, weken, geld, mei
05/2020	1	bedrijven, miljoen, Makhlouf, wereld, situatie, weken, vraag, Facebook, contract, coronacrisis
	2	bedrijven, Brussels Airlines, miljard, coronacrisis, miljoen, geld, Lufthansa, bedrijf, banken, landen
	3	China, land, landen, president, Trump, virus, wereld, coronacrisis, week, Europa
	4	N-VA, partij, PS, Magnette, CD&V, Vlaanderen, VLD, MR, coronacrisis, Vlaams Belang
	5	bedrijven, virus, aantal, weken, werk, lockdown, coronacrisis, maanden, tijd, land

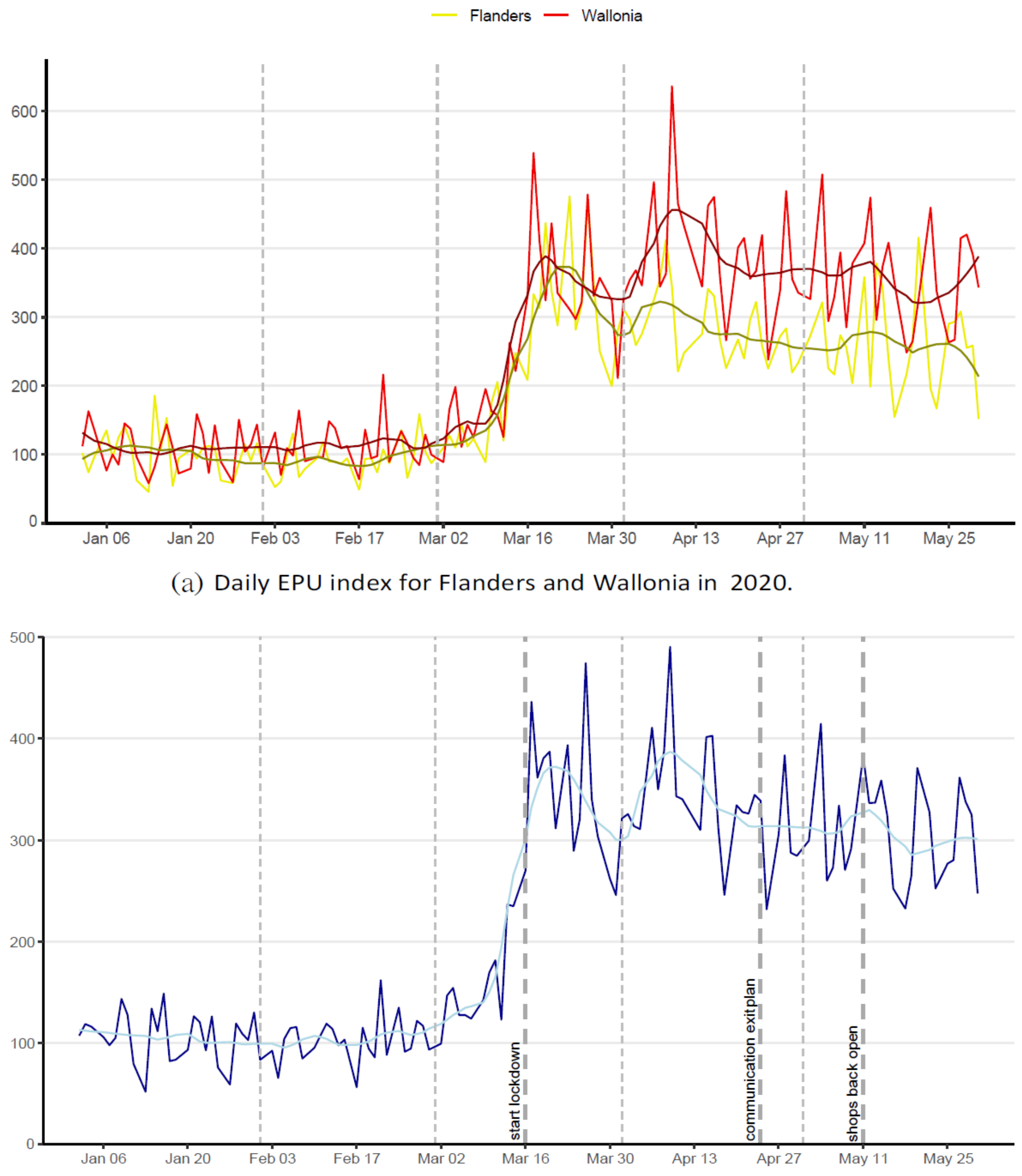
Table 3: Automated display through biterm topic modeling of the most recurring non-EPU terms in *Flemish* press during the COVID-19 pandemic.

In almost all the topic clusters, many terms refer explicitly to the pandemic. Over the months, there is an increasing focus on the U.S. and the possible overall economic and political consequences of the corona virus. The intention of parent company Lufthansa to restructure Brussels Airlines comes up as an important topic in May, in both Flemish and Walloon press.

Month	Cluster	Top non-EPU terms
03/2020	1	taux, BCE, coronavirus, baisse, marché, cours, l'économie, mois, pays, cas
	2	pays, coronavirus, président, États, l'UE, zone, cas, l'épidémie, virus, l'Union
	3	N-VA, PS, président, coronavirus, confiance, mois, parti, temps, Parlement, pays
	4	coronavirus, pays, virus, Chine, cas, monde, temps, santé, pandémie, l'épidémie
	5	cas, coronavirus, secteur, mois, jours, situation, temps, travailleurs, travail, Belgique
04/2020	1	mai, président, mois, fin, pays, temps, monde, saison, N-VA, coronavirus
	2	confinement, secteur, temps, situation, travail, gens, cas, mois, personnel, mai
	3	prix, marché, coronavirus, mois, secteur, baisse, d'euros, groupe, pays, confinement
	4	confinement, monde, virus, pays, temps, santé, population, coronavirus, cas, tests
	5	pays, États, l'UE, d'euros, plan, pandémie, zone, l'Union, l'économie, relance
05/2020	1	pays, États, l'Union, relance, plan, pandémie, l'UE, PIB, zone, BCE
	2	secteur, mois, plan, temps, PS, mai, juin, fin, situation, travail
	3	pays, monde, pandémie, Chine, président, coronavirus, question, fin, Grèce, Donald Trump
	4	confinement, cas, virus, santé, situation, temps, déconfinement, mois, place, jours
	5	d'euros, groupe, mois, Lufthansa, compagnie, secteur, Belgique, plan, pays, Brussels Airlines

Table 4: Automated display through biterm topic modeling of the most recurring non-EPU terms in *Walloon* press during the COVID-19 pandemic.

Figure 3 shows the EPU indices for Flanders, Wallonia and Belgium on a daily scale, from January 2020 to May 2020. It presents the true series and a locally smoothed version (using LOESS regression), with Sundays dropped. The dynamics are interesting. January and February are calm months with a level around 100, which indicates the same degree of uncertainty as on average up to 2011. Thereafter, the uncertainty unequivocally drives up, becomes more volatile transitioning from March to April, and has been decreasing in the last three weeks of April. Most of the uncertainty was accrued before the lockdown in Belgium was officially imposed. The uncertainty remains high in May, but at a fairly constant, and lower, daily level. The daily coverage in Flemish newspapers versus the one in Walloon newspapers is similar, as shown in Figure 3a, though the Wallonia series is consistently higher.



(a) Daily EPU index for Flanders and Wallonia in 2020.

(b) Daily EPU index for Belgium in 2020.

Figure 3: Daily Belgian EPU indices in 2020. The lines on top are LOESS curves.

The increase in the monthly EPU indices for Belgium from March to April thus not stems from a day-to-day increasing trend, but from a sustained high level throughout April. Still, the overall observed daily trend in Belgian newspaper coverage surrounding economic policy uncertainty seems to point toward a soft decline going forward.

4. Conclusion

This paper describes the construction of EPU indices for Flanders, Wallonia and Belgium from press data in the style of Baker et al. (2016). The EPU index is an interesting descriptive measure of the degree to which newspapers are discussing economic policy concerns using terms related to uncertainty. The constructed indices correlate with existing European uncertainty time series but also capture national evolutions. The last part of our analysis focuses on 2020 and the COVID-19 pandemic. News-based economic policy uncertainty reaches unseen levels in March and April, but witnesses a decreasing trend in May. More timely (up to daily) and alternative calculations of the presented indices are available upon request.

The main question for further research is how to use text-based indices, including any version of a news-based EPU index, to improve nowcasts and forecasts about the (Belgian) economy. In an increasingly faster evolving world, nowcasting might have become the hottest practice within departments responsible for economic analysis. Text-based indices have the advantage of being flexible, timely, and are able to uncover latent variables. For an overview of the different steps in creating and researching the added value of textual indices, we refer to the survey of Algaba et al. (2020).

References

Algaba, A., Ardia, D., Bluteau, K., Borms, S., Boudt, K., 2020. Econometrics meets sentiment: An overview of methodology and applications. Journal of Economic Surveys 34, 512–547. doi: 10.1111/ joes.12370.

Azqueta-Gavaldon, A., 2017. Developing news-based economic policy uncertainty index with unsupervised machine learning. Economics Letters 158, 47–50. doi: 10. 1016/j.econlet.2017.06.032.

Azqueta-Gavaldon, A., Hirschbühl, D., Onorante, L., Saiz, L., 2020. Economic policy uncertainty in the Euro area: An unsupervised machine learning approach. URL: https://ssrn.com/abstract=3516756. ECB working paper no. 2359.

Baker, S., Bloom, N., Davis, S., Terry, S., 2020. COVID-induced economic uncertainty. URL: https://www.nber.org/papers/w26983. NBER working paper no. 26983.

Baker, S.R., Bloom, N., Davis, S.J., 2016. Measuring economic policy uncertainty. The Quarterly Journal of Economics 131, 1593–1636. doi: 10.1093/qje/qjw024.

Ghirelli, C., Pérez, J., Urtasun, A., 2019. A new economic policy uncertainty index for Spain. Economics Letters 182, 64–67. doi: 10.1016/j.econlet.2019.05.021.

Kroese, L., Kok, S., Parleviet, J., 2015. Beleidsonzekerheid in Nederland. Economisch-Statistische Berichten (ESB) 100, 4715. URL: https://www.policyuncertainty.com/netherlands.monthly.html.

Pennington, J., Socher, R., Manning, C., 2014. GloVe: Global vectors for word representation, in: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, ACM. pp. 1532–1543. doi: 10.3115/v1/D14-1162.

Tobback, E., Naudts, H., Daelemans, W., Junqué de Fortuny, E., Martens, D., 2018. Belgian economic policy uncertainty index: Improvement through text mining. International Journal of Forecasting 34, 355–365. doi: 10.1016/j.ijforecast.2016.08.006.

Wijffels, J., 2019. **udpipe**: A tokenization, parts of speech tagging, lemmatization and dependency parsing with the UDPipe NLP toolkit. URL: https://CRAN. R-project.org/package=udpipe. R package version 0.8.3.

Wijffels, J., 2020. **BTM**: Biterm topic models for short text. URL: https://CRAN. R-project.org/package=BTM. R package version 0.3.1.

Yan, X., Guo, J., Lan, Y., Cheng, X., 2013. A biterm topic model for short texts, in: Proceedings of the 22nd international conference on World Wide Web, pp. 1445–1456. doi: 10.1145/2488388.2488514.