



**DORTMUND CENTER
FOR DATA-BASED
MEDIA ANALYSIS**

DoCMA Working Paper #5

Economic Policy Uncertainty Index

Extension and optimization of Scott R. Baker, Nicholas Bloom and Steven J. Davis's search term

Richard Brandt*

Cite as:

Brandt, R. (2021). "Economic Policy Uncertainty Index: Extension and optimization of Scott R. Baker, Nicholas Bloom and Steven J. Davis's search term." *DoCMA Working Paper #5, Feb. 2021*. DOI: [10.17877/DE290R-21922](https://doi.org/10.17877/DE290R-21922)

Version 1.0, February 2021

*Richard Brandt studies the Masters's programme "Economics and Journalism" at TU Dortmund University's Institute of Journalism.

Abstract:

The current coronavirus pandemic is a recent example of how an unpredictable event can cause uncertainty and affect people, markets and economies worldwide. Economic indicators like the well-known and frequently cited Economic Policy Uncertainty Index, developed by the three economists Scott R. Baker, Nicholas Bloom and Steven J. Davis, try to capture perceptions of uncertainty and use them to generate predictions for the economic future. The index is based on a search term that is employed to select relevant elements from a population of articles.

However, previous research contributions criticize the composition of the search term for being too broad and restricted to specific policy areas, thereby missing detections of new or unforeseen sources of economic uncertainty. This research note aims to modify the economists' search term in three different ways. The author introduces the term and concept of "risk", which is not considered in the original search term, adds further policy areas that can cause economic uncertainty, and extends the term "uncertainty" by including related terms and synonyms.

In order to evaluate the success of the optimized search terms in selecting relevant articles from a population of articles, all search terms are applied to two randomly drawn samples, derived from an original corpus of 2,723,049 articles and a pre-filtered corpus of 514,297 articles from the German daily newspapers Handelsblatt and Süddeutsche Zeitung. The investigation period ranges from January 1994 to March 2020. For comparison, both samples are also filtered using Baker, Bloom and Davis's original search term. The different selection results are evaluated with the help of the parameters recall and precision.

The results are preliminary, but encouraging. In both samples, only around every tenth relevant article is selected when Baker, Bloom and Davis's search term is applied. By modifying the original search term, the recall could be increased considerably with little disadvantages in terms of precision. The research process shows that economic uncertainties can be related to other concepts and policy areas that are not captured by the economists' original search term.

Keywords: economic policy, uncertainty, risk, search term, optimization, media analyses, recall, precision

1 Introduction

Economic downturns, large amounts of debt and an increase in unemployment and poverty worldwide; this is just a selection of the topics that have dominated the headlines of national and international press since early 2020. The reason being the COVID-19 or coronavirus pandemic. Not only has it been a threat to people's health, but also to their prosperity. This is what media depicts.

Many economic indicators try to capture such moods and use them to derive predictions for the economic future. The Economic Policy Uncertainty Index (EPU Index), created by the three economists Scott R. Baker, Nicholas Bloom and Steven J. Davis from the United States (US), is particularly well-known and frequently cited. The index is based on a search term that is used to select relevant elements from a population of articles.

This research note analyzes the economists' search term and develops approaches to modify the instrument. The exploratory research question is: Can recall and precision of the selection process be increased with an optimized search term? The aim is to increase both the ability of the search term to select relevant articles from an initial corpus, as well as to identify articles as relevant, in case they are already part of a filtered corpus. For this purpose, the research note starts with an overview of the concept and previous applications of the Economic Policy Uncertainty Index in scientific research (section 2). After deriving his own research approach and describing the methodology and data basis (section 3), the author presents the results of the analysis (section 4). Finally, he discusses the results, draws conclusions and suggests approaches for further research projects (section 5). An appendix provides all search terms that are used in this research note, as well as the code book that serves as the central instrument of the coding process.

2 Theory

Baker, Bloom and Davis (2016) focus their study on the analysis of economic uncertainties that are presented in media coverage. However, in such contexts, one does not only speak of uncertainties, but also of risks. This term has not been considered in the search term developed by the researchers yet. Journalists have been using the term "risk" more frequently. In a study of various British and US daily newspapers, Zinn (2010, p. 106) states a "ubiquitous increase in societal 'risk-communication'". He also provides an interesting insight into the use of the terms "risk" and "uncertainty" in the New York Times. While the term "uncertainty" was used in around 1,000 articles in 2006, the term "risk" was used five times as often in almost 5,000 articles (2010, pp. 114–118). It therefore seems sensible and necessary to consider the term "risk" in further investigation.

In recent years, increasingly more research contributions have concentrated on making media data quantifiable for research and interpretable for economic analyses and models (Ammann, Frey & Verhofen, 2014, p. 196; Kholodilin, Kolmer, Thomas & Ulbricht, 2015, p. 2; Larsen & Thorsrud, 2019a, p. 6). Media coverage has a significant impact on which topics appear relevant to a society, and how these topics are thought about through the

selection of information and sources, as well as the tone of reporting. This process is known as “agenda setting effect”. It can shape the public’s perception of economic uncertainties and risks, thereby influencing economic development (Baker et al., 2016, p. 1595; Doms & Morin, 2004, p. 8; Larsen & Thorsrud, 2019b, p. 203; Müller, Nordheim, Boczek, Koppers & Rahnenführer, 2018, p. 561; Ulbricht, Kholodilin & Thomas, 2017, p. 484).

The influencing factors are exogenous, cannot be directly observed and cannot be predicted using traditional economic models or indicators (Moore, 2017, p. 550; Müller et al., 2018, pp. 559, 567–568). These factors are crucial here, as uncertainties can not only result in higher volatility, for example on stock markets, but also impede consumption, investment, the production of goods and the creation of jobs, and thus slow down economic growth (Baker et al., 2016, pp. 1595–1597, 1632–1634; Bloom, Bond & van Reenen, 2007, pp. 391–393; Bloom, 2009, p. 623; Müller et al., 2018, p. 567).

Ten major US daily newspapers serve as the data basis for the Economic Policy Uncertainty Index. Using a defined search term, the researchers select articles from the digital archives of the individual US media from January 1985 on. An article is considered relevant if it includes at least one word from three different sets of terms each. These are

“uncertainty” or “uncertain”; “economic” or “economy”; and one of the following policy terms: “Congress”, “deficit”, “Federal Reserve”, “legislation”, “regulation”, or “White House” [...].” (Baker et al., 2016, p. 1599)

The number of relevant articles determines the course of the Economic Policy Uncertainty Index (Baker, Bloom & Davis, 2012). The researchers compile the EPU Index on a monthly basis. But Baker, Bloom and Davis (2016, p. 1604) have also developed indices for other countries such as Germany, France and Japan, as well as economically emerging nations such as China, India and Russia. The analyses are carried out in the respective language of the examined media, using a translated and slightly adapted search term. For Germany this search term reads as follows:

“wirtschaft OR wirtschaftlich; [...] steuer OR wirtschaftspolitik OR regulierung OR regulierungs OR ausgaben OR bundesbank OR EZB OR zentralbank OR haushalt OR defizit OR haushaltsdefizit; and [...] unsicher OR Unsicherheit.” (Baker et al., 2016, A.2)

The archives of the German daily newspapers Handelsblatt and Frankfurter Allgemeine Zeitung from 1993 onward serve as the basis for the German index. Figure 1 shows the EPU Index for Germany and the United States of America (USA) over the period of January 1994 to December 2020.

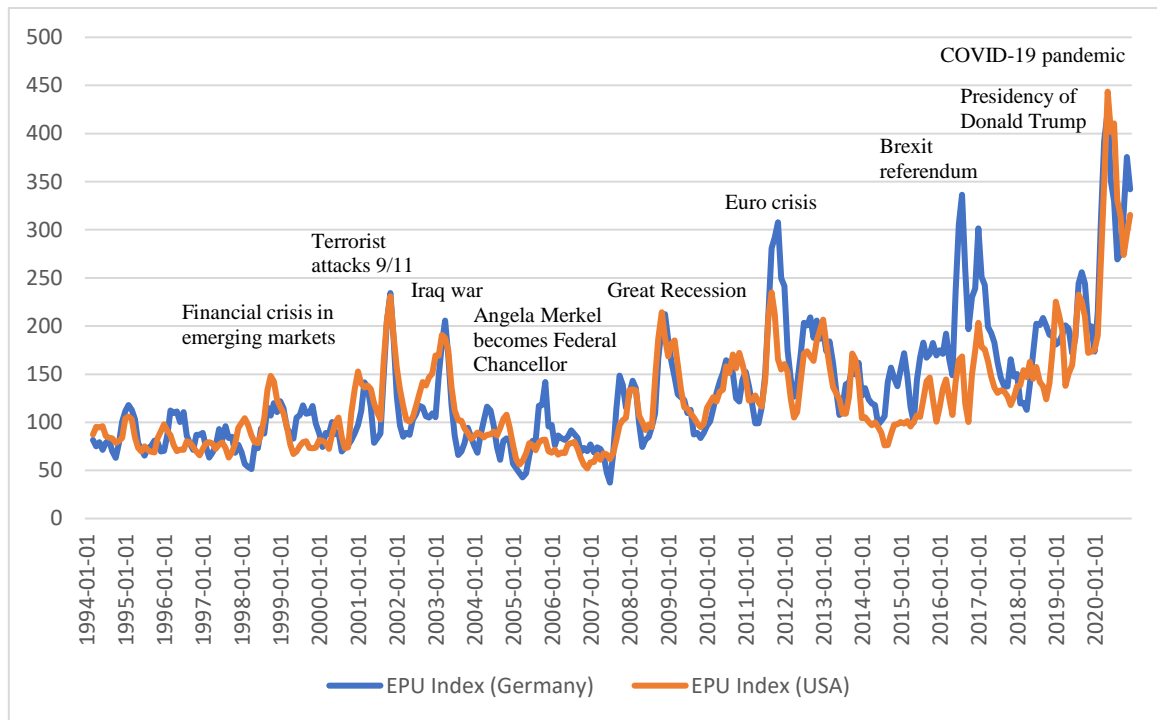


Figure 1: Economic Policy Uncertainty Index for Germany and the United States of America over time, monthly data with a three-month moving average, normalized to starting value 100, selected events (Source: Baker et al., 2012)

More and more research contributions employ Economic Policy Uncertainty Index data. Yu, Fang and Sun (2018) observe that global economic policy uncertainty affects the volatility of the Chinese stock market, while Chen and her colleagues (2019) examine how oil price shocks and global economic policy uncertainty influence the industrial growth in China. Dash and his research team (2019) observe that the stock market liquidity in G7 economies decreases when EPU levels are rising, while Degiannakis and Filis (2019) analyze factors that forecast European economic policy uncertainty, including realized stock market volatilities. Antonakakis, Gabauer and Gupta (2019) examine the mutual effects of uncertainty shocks and spillovers on Greek and European economic policy uncertainty, especially during the Eurozone sovereign debt crisis.

Alqahtani and his research team (2019) calculate that changes in economic policy uncertainty in the United States are not significantly connected with the returns of the Gulf region's stock markets. Caggiano, Castelnuovo and Figueres (2020) find that uncertainty shocks in the US affect unemployment rates in Canada and the United Kingdom. Nguyen and his colleagues (2020) show that higher levels of EPU have negative effects on domestic and global bank credit growth. Other recent investigations that employ EPU data have been conducted by Azzimonti (2018), Caldara and Iacoviello (2018), as well as Baker, Bloom and Davis themselves. The three economists examine the impact of the coronavirus pandemic on global economic uncertainty, stock market volatility and future recovery from the COVID-19 recession (Baker, Bloom, Davis & Terry, 2020; Baker, Bloom, Davis, Kost, et al., 2020; Barrero & Bloom, 2020).

Building on the approach of the US researchers, Müller, Nordheim, Boczek, Koppers and Rahnenführer (2018) conduct their own study and develop a new indicator, the Uncertainty Perception Indicator (UPI). Müller and his colleagues analyze the economic uncertainties represented by the UPI, using the topic modeling method Latent Dirichlet Allocation (LDA), in order to make media narratives measurable and thus interpretable. This probabilistic method allows media reporting patterns to be recorded in large text corpora and converted into numerical time series. The results can then serve as indicators of economic expectations (Larsen & Thorsrud, 2019a, p. 6; Maier et al., 2018, p. 93; Müller et al., 2018, pp. 559–560, 565–566; Müller & Hornig, 2020b, p. 18).

Based on the search term used by Baker, Bloom and Davis for Germany, the researchers generate a similar corpus of articles for the years 1994 to 2017. However, Müller and his colleagues use the German daily newspaper *Süddeutsche Zeitung* as the data basis. In total, the researchers analyze 6,151 articles. In contrast to the Economic Policy Uncertainty Index, Müller and his team can use the Uncertainty Perception Indicator not only to make statements about how often economic uncertainties are depicted in the newspaper, but also identify where the respective uncertainty originates. Since the financial crisis in 2007, political decision-makers and central banks have made a significant contribution to the perception of economic uncertainties (Müller et al., 2018, pp. 559, 572–573; Müller & Hornig, 2020b, p. 12).

But Müller and his colleagues (2018, p. 568) also criticize Baker, Bloom and Davis's approach. The economists' procedure simplifies the economic discourses and developments on which the study is based considerably. Hence, new topics and relations between economic discourses are overlooked by the previously defined search term, and changes in the economic agenda over time are not sufficiently taken into account, considering the broad search term sets. Müller and Hornig (2020b, pp. 4–8, 11–12) point to another inadequacy of the approach. When compiling the EPU Index, the search term restricts the range of possibility of influences that cause economic uncertainty. The search term selects articles that deal with uncertainties in connection with taxation, regulation and monetary and fiscal policy. This procedure has proven itself so far. However, it is not suitable for capturing new sources of economic uncertainty; for instance, those based on variables such as energy policy, climate change, technologies or the COVID-19 pandemic. Therefore, early warning indicators are needed that can not only detect "known unknowns", but also "surprising unknowns" (Müller & Hornig, 2020b, p. 7).

For this reason, the two researchers modify the Uncertainty Perception Indicator by eliminating the part of the search term that specifies the relevant policy areas. 8,295 and 15,077 articles from the two German daily newspapers *Die Welt* and *Handelsblatt* over the period of January 2008 to March 2020 serve as the data basis for the study. The articles are selected using the original and broader search term, respectively. Using LDA analysis, Müller and Hornig receive a new topic, "Energy & Climate Change Mitigation", which includes aspects of exogenous economic uncertainty. With regard to the outbreak of the COVID-19 pandemic, the two researchers find that the modified UPI already shows

increased values in February 2020, while this development occurs a month and a half later for the original UPI, which is based on Baker, Bloom and Davis's original search term (Müller & Hornig, 2020b, pp. 8–11, 16–18).

In following work, the authors extend the scope of research again and present an enhanced version of the UPI by including the German newspaper *Süddeutsche Zeitung*, and extending the time horizon, which already starts in 2000 now. In total, 37,000 articles are selected with the previously developed broader search term and analyzed subsequently. The researchers find that COVID-related issues appear in the previously found topic, “Energy & Climate Change Mitigation”, highlighting the increasing relevance of hardly foreseeable, but possibly impactful “truly exogenous uncertainty” that is present today (Müller & Hornig, 2020a, pp. 10, 18–19).

This research note pursues Müller and Hornig's approach. However, the search term that Baker, Bloom and Davis use to obtain the Economic Policy Uncertainty Index is modified in a different way. Instead of eliminating part of the search term, it is expanded by including additional terms, so that the selection process chooses additional relevant articles. The specific exploratory research question is: Can recall and precision of the selection process be increased with an optimized search term?

3 Methodology and data

Overall, this research note focuses on three extensions of the original search term:

- a. the introduction of the term and concept of “risk”, which does not exist in the original search term, including related terms and synonyms; and
- b. the addition of further policy areas that can cause economic uncertainty; and
- c. the addition of related terms and synonyms to the term “uncertainty”, which is already present in the original search term.

The research process is largely based on the methodology of Stryker, Wary, Hornik and Yanovitzky (2006, pp. 415–419). The procedure is divided into two parts. First, two modified preliminary search terms are designed with the help of various lexicons and previous research contributions (Bibliographisches Institut & F. A. Brockhaus AG, 2007, pp. 717–718; Bibliographisches Institut GmbH, 2011, p. 1458; Frühbrodt, 2010, pp. 61–63; Heinrich, 1989, p. 284; Knapp, 2000, pp. 150, 315, 543, 633, 645; Leonhardt, 2004, pp. 273–274; Mast, 2012, pp. 104, 107–110; Müller, 2017, pp. 30–32; Wahrig, Krämer & Zimmermann, 1983, pp. 399–400; Waite, 2006, p. 710). In the next step, the first sample is drawn randomly and the two preliminary search terms are applied subsequently.

This research takes into account the two German daily newspapers *Handelsblatt* and *Süddeutsche Zeitung*. The investigation period ranges from January 1994 to March 2020. All articles from the two daily newspapers are derived from the media publishers and are taken from the inventory of the Dortmund Center for Data-based Media Analysis (DoCMA). The articles are merged into a corpus of 2,723,049 articles, hereafter referred to as the original corpus, before further analysis. It represents the total population of the

study (Brosius, Haas & Koschel, 2016, p. 59). The programming language “R”, and in particular the package “tosca” (Koppers, Rieger, Boczek & Nordheim, 2020), is used to carry out all research steps. The central instrument of the coding process is the code book (Brosius et al., 2016, p. 157). Using the code book, the coder decides if an article is classified as relevant or not based on specific criteria. It can be found in the appendix A.2 (p. 22). A total of 150 articles from the two daily newspapers serve as analysis units for this research step, as recommended by Stryker and her team (2006, p. 417).

The author then optimizes the two preliminary search terms, using the findings from the first selection process, in order to obtain the two final search terms, referred to by Stryker and her research colleagues as closed search terms (2006, pp. 416–417). For this purpose, the component of the search term which lists the policy areas is expanded again. All search terms that are used in the research process can be found in the appendix A.1 (p. 20). In the next step, the final search terms are applied to a second sample, again consisting of 150 articles from the two examined daily newspapers. But before drawing the second sample, an open search term, excluding the element specifying the policy areas, is used to select all articles from the original corpus of 2,723,049 articles that are considered relevant according to the code book. The resulting corpus, hereafter referred to as the pre-filtered corpus, consists of 514,297 remaining articles. In addition, both samples are also filtered using Baker, Bloom and Davis’s original search term. The success of the single selection processes is evaluated with the help of the parameters recall and precision.

The two measures are defined below - first the recall:

“Recall is an estimate of the conditional probability that a particular text will be retrieved, given that it is relevant, calculated by dividing the number of relevant items returned by a search phrase by the total relevant records in the database.” (Stryker et al., 2006, p. 414)

Precision is defined as follows:

“Precision is an estimate of the conditional probability that a particular text is relevant, given that it is retrieved, calculated by dividing the number of relevant items by the total number of items returned in a specific search.” (Stryker et al., 2006, pp. 414–415)

The measures are standard sizes, as they are particularly suitable for evaluating the performance of a selection process thanks to their easy interpretability. The challenge of earlier studies, as well as the present one, is increasing the values for recall and precision at the same time, starting with the values that are obtained when applying Baker, Bloom and Davis’s original search term. Recall and precision often have the opposite effects. This means, if the recall rises, the precision suffers, and vice versa. However, this does not necessarily have to be the case. High values for both parameters are generally desirable (Chamis, 1991, p. 13; Eastman, 1988, pp. 28–29; Hildreth, 1989, p. 49; Meadow, Boyce, Kraft & Barry, 2006, pp. 328–331). Since the research question of the present work is exploratory, no explicit target values for recall and precision are given.

4 Results

In order to evaluate the success of the research process, the selection results are compared using the two parameters recall and precision. The examination of the original corpus of 2,723,049 articles, using the first preliminary search term, results in a recall value of 0.217, or 21.7 percent, for the analyzed first sample and a precision value of 1, or 100 percent. When using the second preliminary search term, the values are 0.362, or 36.2 percent, for recall and 0.893, or 89.3 percent, for precision. Figure 2 shows the shares of the subcorpora, filtered with the first and second preliminary search term, relative to the original corpus over the period of January 1994 to March 2020.

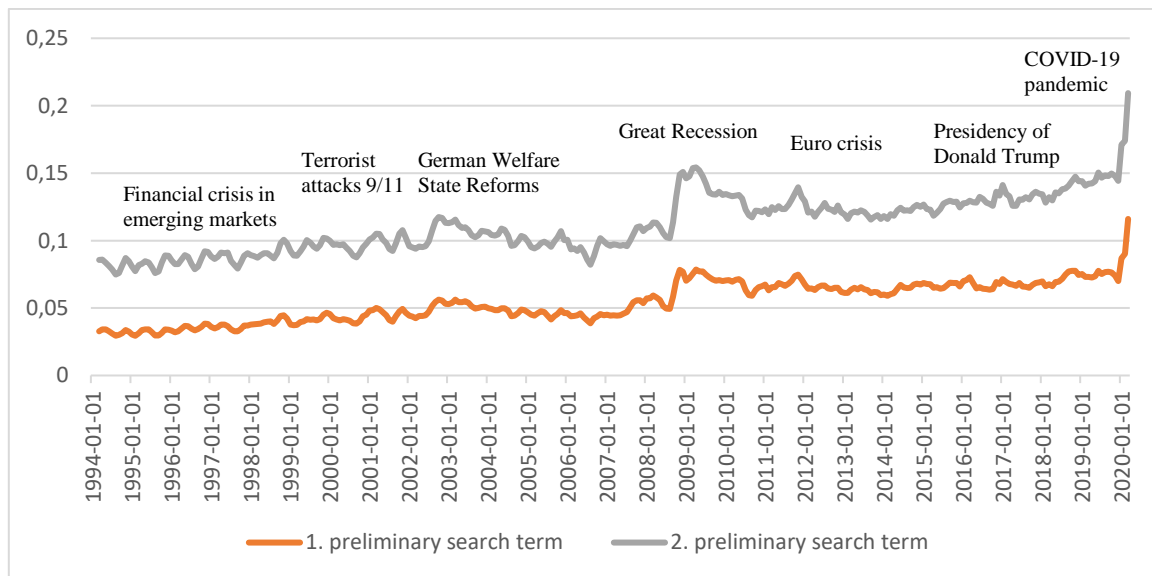


Figure 2: Shares of the subcorpora, filtered with the first and second preliminary search term, relative to the original corpus over time, $1 \triangleq 100\%$, monthly data with a three-month moving average, selected events (Source: author's calculations)

The results imply the following: While the two preliminary search terms are suitable for identifying relevant articles if they are already part of a filtered corpus, they show weaknesses when it comes to selecting relevant articles from the original corpus. For comparison, a selection using Baker, Bloom and Davis's search term achieves a recall value of 0.072, or 7.2 percent, and a precision value of 1, or 100 percent, for the analyzed sample. Consequently, with regard to the selection of relevant articles from the original corpus, it performs significantly worse than the preliminary search terms. Figure 3 shows the share of the subcorpus, filtered with Baker, Bloom and Davis's search term, relative to the original corpus over the observation period.

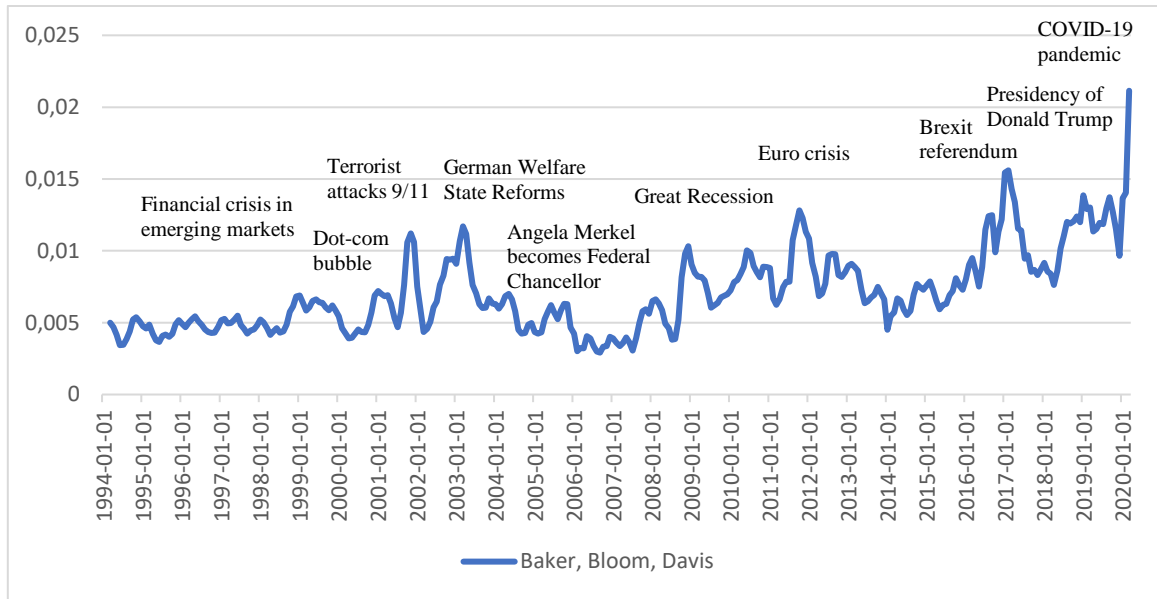


Figure 3: Share of the subcorpus, filtered with Baker, Bloom and Davis's search term, relative to the original corpus over time, $1 \hat{=} 100 \%$, monthly data with a three-month moving average, selected events (Source: author's calculations)

The author then optimizes the two preliminary search terms, using the findings from the first selection process, in order to obtain the two final search terms. If the final search terms are applied to the original corpus, one can observe different developments. The recall and precision values for the first final search term are equivalent to the values the first preliminary search term achieves. However, one can observe changes for the second search term. While the precision of the second final search term is slightly decreasing to a value of 0.875, or 87.5 percent, the recall value is increasing to 0.406, or 40.6 percent, compared to the results achieved with the second preliminary search term. The values of recall and precision for the first sample are summarized in Table 1.

Table 1: Results for the first sample (Source: author's calculations)

Applied search term	Baker, Bloom, Davis	1. prel. search term	2. prel. search term	1. final search term	2. final search term
Recall	0.072	0.217	0.362	0.217	0.406
Precision	1	1	0.893	1	0.875

prel. $\hat{=}$ preliminary

Before drawing the second sample, an open search term, excluding the element specifying the policy areas, is used to select all articles from the original corpus of 2,723,049 texts that are considered relevant according to the code book. The resulting pre-filtered corpus consists of 514,297 remaining articles.

Figure 4 shows the share of the pre-filtered corpus relative to the original corpus.

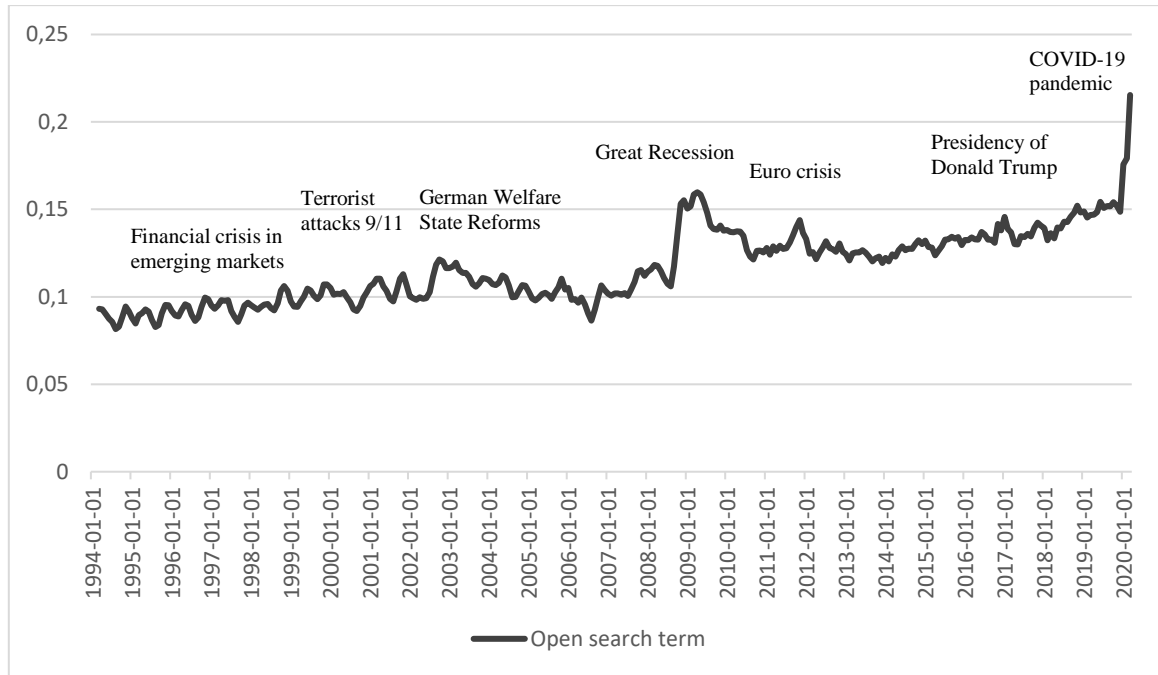


Figure 4: Share of the pre-filtered corpus relative to the original corpus over time, $1 \triangleq 100\%$, monthly data with a three-month moving average, selected events (Source: author's calculations)

The examination of the pre-filtered corpus, using the first final search term, results in a recall value of 0.491, or 49.1 percent, for the analyzed second sample, and a precision value of 0.783, or 78.3 percent. Figure 5 shows the share of the subcorpus, filtered with the first final search term, relative to the pre-filtered corpus over the observation period.

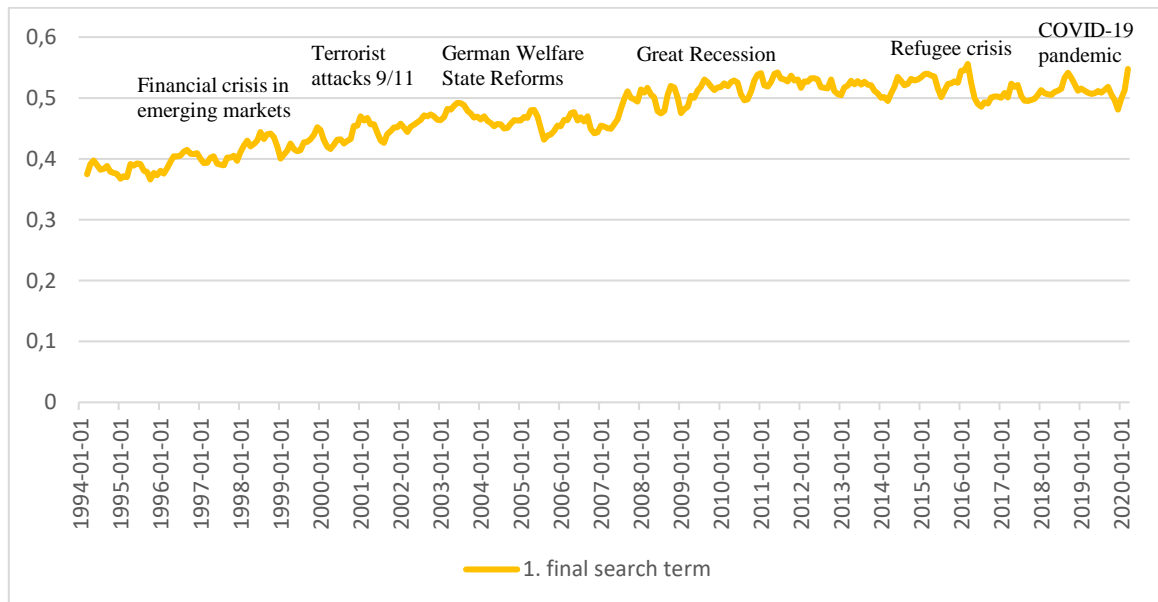


Figure 5: Share of the subcorpus, filtered with the first final search term, relative to the pre-filtered corpus over time, $1 \triangleq 100\%$, monthly data with a three-month moving average, selected events (Source: author's calculations)

Correspondingly, when using the second final search term, the values are 1, or 100 percent, for the recall and 0.733, or 73.3 percent, for the precision. Figure 6 shows the share of the subcorpus, filtered with the second final search term, relative to the pre-filtered corpus over

the period of January 1994 to March 2020.

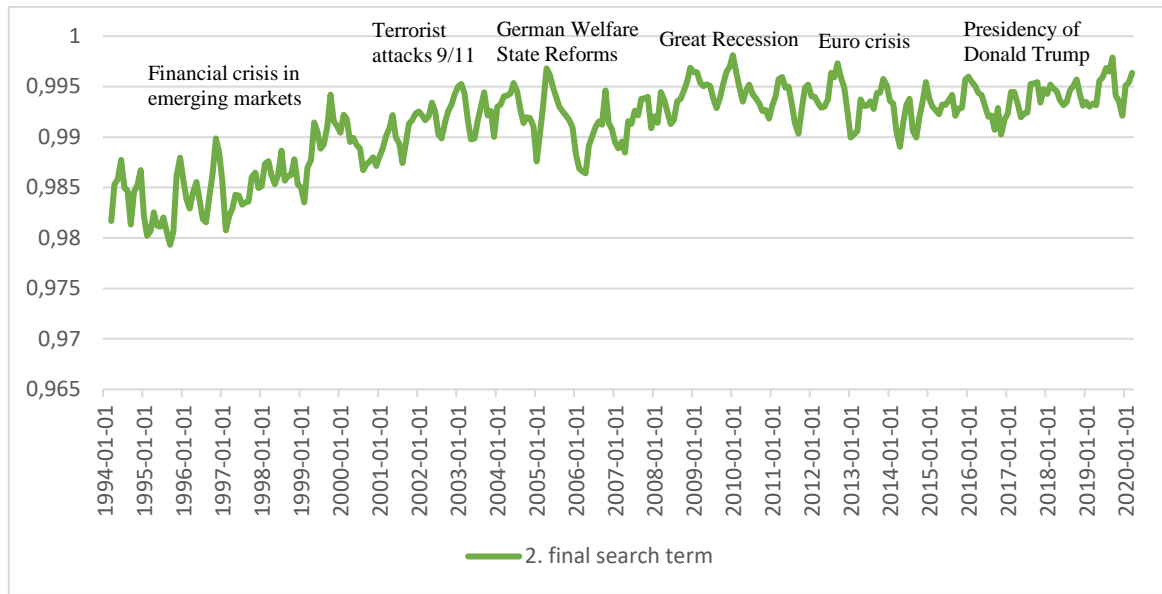


Figure 6: Share of the subcorpus, filtered with the second final search term, relative to the pre-filtered corpus over time, $1 \cong 100\%$, monthly data with a three-month moving average, selected events (Source: author's calculations)

For comparison, a selection using Baker, Bloom and Davis's search term, achieves a recall value of 0.109, or 10.9 percent, and a precision value of 1, or 100 percent. In terms of recall, it performs worse than the final search terms. As for the first sample, only around every tenth relevant article is selected using the economists' search term. In case of precision, however, their search term achieves a better result than the final search terms. Figure 7 shows the share of the subcorpus, filtered with Baker, Bloom and Davis's search term, relative to the pre-filtered corpus.

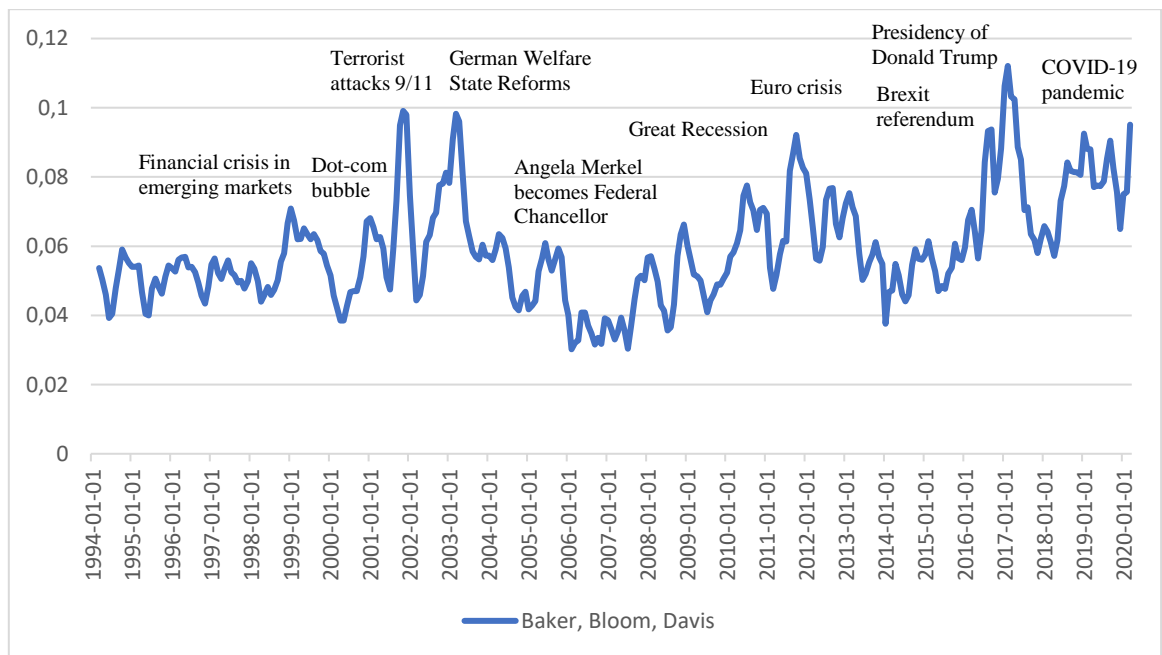


Figure 7: Share of the subcorpus, filtered with Baker, Bloom and Davis's search term, relative to the pre-filtered corpus over time, $1 \cong 100\%$, monthly data with a three-month moving average, selected events (Source: author's calculations)

Figure 8 provides an overview of the shares of the subcorpora relative to the pre-filtered corpus over the period of January 1994 to March 2020.

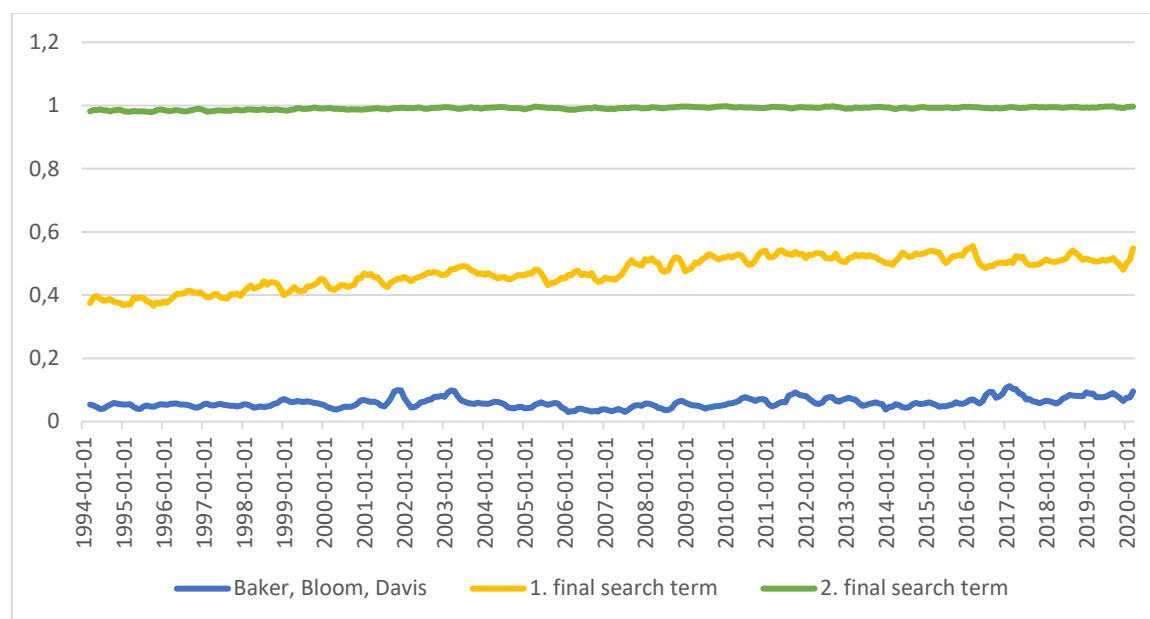


Figure 8: Shares of the subcorpora relative to the pre-filtered corpus over time, $1 \triangleq 100\%$, monthly data with a three-month moving average (Source: author’s calculations)

Comparatively, if the first preliminary search term would have been applied to the pre-filtered corpus, the recall value would be 0.482, or 48.2 percent, and the precision value would be 0.803, or 80.3 percent. The second preliminary search term achieves values of 0.964, or 96.4 percent, for the recall and 0.741, or 74.1 percent, for the precision when applied to the pre-filtered corpus. All results for the second sample are summarized in Table 2.

Table 2: Results for the second sample (Source: author’s calculations)

Applied search term	Baker, Bloom, Davis	1. prel. search term	2. prel. search term	1. final search term	2. final search term
Recall	0.109	0.482	0.964	0.491	1
Precision	1	0.803	0.741	0.783	0.733

prel. \triangleq preliminary

Overall, the research states, between the individual values for recall and precision, only minimal deviations can be observed in the two samples if the preliminary search terms are compared with their final counterparts. However, the optimization of the second preliminary search term to the second final search term has a slightly positive effect on the recall with small compromises in precision.

With the second preliminary and final search term, around 40 percent of all relevant articles in the first sample, based on the unfiltered original corpus, are already selected. The precision values of the search terms used are also higher in the first sample, compared to the second, which is based on the pre-filtered corpus.

5 Discussion and conclusions

This research note has aimed to optimize Baker, Bloom and Davis's original search term regarding recall and precision. For that purpose, modified variants of the search term have been developed and applied to two randomly drawn samples, derived from an original corpus of 2,723,049 articles and a pre-filtered corpus of 514,297 articles from the German daily newspapers *Handelsblatt* and *Süddeutsche Zeitung*. Eventually, the selection results have been compared to the ones of Baker, Bloom and Davis's original search term. The research process provides three important findings.

First, it can be stated that the second final search term turns out to be the term that achieves the best result. Compared to Baker, Bloom and Davis's original search term, it has weaknesses in terms of precision. However, this disadvantage is offset by its improved performance in terms of recall. In both samples, only around every tenth relevant article is selected using Baker, Bloom and Davis's search term. In contrast, using the second final search term results in recall values of 0.406 for the first sample, which is drawn from the original corpus, and 1 for the second sample, which is drawn from the pre-filtered corpus.

Second, it shows that considering both terms "risk" and "uncertainty", including related terms and synonyms, appears to be much more effective with regard to the highest possible result for the two parameters recall and precision than a single consideration of the respective term fields. A combination of the two, as in case of the second preliminary and final search term, is therefore recommended.

Third, it becomes clear, in particular from the analysis of the two random samples, that economic uncertainties can be related to other policy areas that are not captured by Baker, Bloom and Davis's original search term. This research approach is a reaction to the criticism expressed by Müller, Nordheim, Boczek, Koppers and Rahnenführer (2018, p. 568) and Müller and Hornig (2020b, pp. 4–8, 11–12). On the one hand, the expansion of Baker, Bloom and Davis's original search term is intended to better reflect the range of origins of uncertainties and the relationships between possible influences, thus including factors that have not been considered in the investigation yet. On the other hand, the composition of the individual components of the optimized search terms shall ensure that the factors of interest are selected in a targeted manner and not too imprecisely.

The specific research question has been: Can recall and precision of the selection process be increased with an optimized search term? After the investigation, this question can partly be answered with a yes. It is true that the precision value decreases when some optimized search terms are applied. In return, the optimized search terms, especially the second final search term, increase the recall of the selection process considerably.

Still, Baker, Bloom and Davis's search term should not be disregarded at once. Further research and applications of the economists' search term are needed to test its efficacy, especially with regard to the German media landscape. Additionally, the modified search terms that have been developed in this research note need to also be tested for their efficacy and predictive power by applying them to other research contexts; for instance, to further economic or media analyses.

In order to evaluate the success of the selection process more precisely, other research could introduce explicit target values for recall and precision, as suggested by Stryker and her research colleagues (2006, pp. 418–419). Subsequently, confidence intervals for recall and precision could be generated, and a larger sample of the German newspapers *Handelsblatt* and *Süddeutsche Zeitung* could be examined to obtain even more reliable and far-reaching results for optimizing Baker, Bloom and Davis's search term. In addition, other time periods could be chosen for the analysis, in order to consider a broader or narrower time horizon, as well as other media.

Furthermore, additional modifications of the search terms are conceivable. In the optimized search terms, the component that lists the policy areas with which economic uncertainties may be connected is expanded. The search terms are supplemented by elements such as “export”, “globalisierung”, “investition” or “klima”. However, the research note does not claim that the search terms are complete. Further elements could be explicitly included such as “Brexit”, “migration”, “pandemie”, “populismus”, “Trump” or “terrorismus”, which all address current issues and are only implicitly taken into account. However, it must be ensured that the search terms are not too broad or narrow. Balance is important. If the search terms are too narrow, many potentially relevant texts could be excluded. On the contrary, if the search terms are too broad, many articles could be selected that are irrelevant to the specific research interest (Koch, 1991, p. 205).

Finally, research approaches that use Baker, Bloom and Davis's search term for further analyses of economic uncertainties with the help of the Latent Dirichlet Allocation method could make use of the optimized search terms, in order to select more suitable corpora with higher recall values. Slightly lower precision values of the developed search terms, compared to Baker, Bloom and Davis's original search term, might not be a problem, since LDA sorts possibly irrelevant information into separate categories (Maier et al., 2018, pp. 94–96). Generally, applying the optimized search terms developed in this research note could provide new insights, thereby generating more meaningful results (Müller et al., 2018; Müller & Hornig, 2020b).

6 References

- Alqahtani, A. S. S., Ouyang, H. & Ali, A. (2019). The Reaction of Stock Markets in the Gulf Cooperation Council Countries to Economic Policy Uncertainty in the United States. *Economics and Business*, 33(1), 22–34. <https://doi.org/10.2478/eb-2019-0002>
- Ammann, M., Frey, R. & Verhofen, M. (2014). Do Newspaper Articles Predict Aggregate Stock Returns? *Journal of Behavioral Finance*, 15(3), 195–213. <https://doi.org/10.1080/15427560.2014.941061>
- Antonakakis, N., Gabauer, D. & Gupta, R. (2019). Greek economic policy uncertainty: Does it matter for Europe? Evidence from a dynamic connectedness decomposition approach. *Physica A: Statistical Mechanics and its Applications*, 535, 122280. <https://doi.org/10.1016/j.physa.2019.122280>
- Azzimonti, M. (2018). Partisan Conflict and Private Investment. *Journal of Monetary Economics*, 93, 114–131.
- Baker, S. R., Bloom, N. & Davis, S. J. (2012). *Economic Policy Uncertainty*. Retrieved from: <http://policyuncertainty.com/>
- Baker, S. R., Bloom, N. & Davis, S. J. (2016). Measuring Economic Policy Uncertainty. *The Quarterly Journal of Economics*, 131(4), 1593–1636. <https://doi.org/10.1093/qje/qjw024>
- Baker, S. R., Bloom, N., Davis, S. J., Kost, K., Sammon, M. & Viratyosin, T. (2020). The Unprecedented Stock Market Reaction to COVID-19. *Review of Asset Pricing Studies*, 10(4), 742–758. <https://doi.org/10.1093/rapstu/raaa008>
- Baker, S. R., Bloom, N., Davis, S. J. & Terry, S. J. (2020). *COVID-induced economic uncertainty*. NBER Working Paper: 26983.
- Barrero, J. & Bloom, N. (2020). *Economic Uncertainty and the Recovery*, Instituto Tecnológico Autónomo de México; Stanford University.
- Bibliographisches Institut & F. A. Brockhaus AG (Ed.). (2007). *Duden - Das Synonymwörterbuch. Ein Wörterbuch sinnverwandter Wörter* (Der Duden in 12 Bänden, das Standardwerk zur deutschen Sprache = Der Duden in zwölf Bänden; 8, 4th ed.). Mannheim: Dudenverlag.
- Bibliographisches Institut GmbH (Ed.). (2011). *Duden - Deutsches Universalwörterbuch* (7th ed.). Berlin: Dudenverlag.
- Bloom, N. (2009). The Impact of Uncertainty Shocks. *Econometrica*, 77(3), 623–685.
- Bloom, N., Bond, S. & van Reenen, J. (2007). Uncertainty and Investment Dynamics. *The Review of Economic Studies*, 74(2), 391–415.

- Brosius, H.-B., Haas, A. & Koschel, F. (2016). *Methoden der empirischen Kommunikationsforschung. Eine Einführung* (Studienbücher zur Kommunikations- und Medienwissenschaft, 7th ed.). Wiesbaden: Springer VS.
- Caggiano, G., Castelnuovo, E. & Figueres, J. M. (2020). Economic Policy Uncertainty Spillovers in Booms and Busts. *Oxford Bulletin of Economics and Statistics*, 82(1), 125–155. <https://doi.org/10.1111/obes.12323>
- Caldara, D. & Iacoviello, M. (2018). *Measuring Geopolitical Risk*. International Finance Discussion Paper: 1222.
- Chamis, A. Y. (1991). *Vocabulary control and search strategies in online searching* (New directions in information management, vol. 27). New York, N.Y.: Greenwood Press.
- Chen, J., Jin, F., Ouyang, G., Ouyang, J. & Wen, F. (2019). Oil price shocks, economic policy uncertainty and industrial economic growth in China. *PLOS ONE*, 14(5), 1-19. <https://doi.org/10.1371/journal.pone.0215397>
- Dash, S. R., Maitra, D., Debata, B. & Mahakud, J. (2019). Economic policy uncertainty and stock market liquidity: Evidence from G7 countries. *International Review of Finance*, 1–16. <https://doi.org/10.1111/irfi.12277>
- Degiannakis, S. & Filis, G. (2019). Forecasting European economic policy uncertainty. *Scottish Journal of Political Economy*, 66(1), 94–114. <https://doi.org/10.1111/sjpe.12174>
- Doms, M. & Morin, N. (2004). *Consumer Sentiment, the Economy, and the News Media*. Finance and Economics Discussion Series: 51.
- Eastman, C. M. (1988). An approach to the evaluation of catalog selection systems. *Information Processing & Management*, 24(1), 23–30. [https://doi.org/10.1016/0306-4573\(88\)90074-X](https://doi.org/10.1016/0306-4573(88)90074-X)
- Frühbrodt, L. (2010). Wirtschaftsjournalismus. In S. Quandt & Deutscher Fachjournalisten-Verband (Ed.), *Fachjournalismus. Expertenwissen professionell vermitteln* (Praktischer Journalismus, vol. 58, 2nd ed., pp. 55–80). Constance: UVK Verlagsgesellschaft.
- Heinrich, J. (1989). Wirtschaftsjournalismus: zur Fundierung einer rezipientenorientierten Wirtschaftsberichterstattung. *Publizistik*, 34(3), 284-296.
- Hildreth, C. R. (1989). *Intelligent interfaces and retrieval methods for subject searching in bibliographic retrieval systems* (Advances in library information technology, vol. 2). Washington, D.C.: Cataloging Distribution Service Library of Congress.
- Kholodilin, K., Kolmer, C., Thomas, T. & Ulbricht, D. (2015). *Asymmetric Perceptions of the Economy: Media, Firms, Consumers, and Experts*. DICE Discussion Paper: 188.

- Knapp, S. D. (2000). *The Contemporary Thesaurus of Search Terms and Synonyms. A Guide for Natural Language Computer Searching* (2nd ed.). Phoenix, Arizona: Oryx Press.
- Koch, T. (1991). *Journalism in the 21st Century. Online Information, Electronic Databases and the News* (Adamantine studies on the 21st century). Twickenham: Adamantine.
- Koppers, L., Rieger, J., Boczek, K. & Nordheim, G. von. (2020). *tosca: Tools for Statistical Content Analysis*. Retrieved from: <https://cran.r-project.org/package=tosca>
- Larsen, V. H. & Thorsrud, L. A. (2019a). *Business Cycle Narratives*. CESifo Working Paper: 7468.
- Larsen, V. H. & Thorsrud, L. A. (2019b). The value of news for economic developments. *Journal of Econometrics*, 210(1), 203–218. <https://doi.org/10.1016/j.jeconom.2018.11.013>
- Leonhardt, R. (2004). *Scharf gedacht - geschliffen formuliert. Das Praxiswörterbuch für die Wirtschaft: Synonyme - Zitate - Abkürzungen*. Munich, Vienna: Hanser Verlag.
- Maier, D., Waldherr, A., Miltner, P., Wiedemann, G., Niekler, A., Keinert, A. et al. (2018). Applying LDA Topic Modeling in Communication Research: Toward a Valid and Reliable Methodology. *Communication Methods and Measures*, 12(2-3), 93–118. <https://doi.org/10.1080/19312458.2018.1430754>
- Mast, C. (2012). Die Wirtschaft und ihr Journalismus. Zur theoretischen Grundlage des Forschungsprogramms. In C. Mast (Ed.), *Neuorientierung im Wirtschaftsjournalismus. Redaktionelle Strategien und Publikumserwartungen* (pp. 73–123). Wiesbaden: VS Verlag für Sozialwissenschaften / Springer Fachmedien.
- Meadow, C. T., Boyce, B. R., Kraft, D. H. & Barry, C. (2006). *Text Information Retrieval Systems* (Library and information science, 3rd ed.). Amsterdam: Elsevier.
- Moore, A. (2017). Measuring Economic Uncertainty and Its Effects. *Economic Record*, 93(303), 550–575.
- Müller, H. (2017). Funktionen und Selbstverständnis des wirtschaftspolitischen Journalismus. In K. Otto & A. Köhler (Ed.), *Qualität im wirtschaftspolitischen Journalismus* (pp. 27–48). Wiesbaden: Springer Fachmedien.
- Müller, H. & Hornig, N. (2020a). "I heard the News today, oh Boy". An updated Version of our Uncertainty Perception Indicator (UPI) – and some general thoughts on news-based economic indicators. DoCMA Working Paper: 2. <http://dx.doi.org/10.17877/DE290R-21669>

- Müller, H. & Hornig, N. (2020b). *Expecting the Unexpected. A new Uncertainty Perception Indicator (UPI) – concept and first results*. DoCMA Working Paper: 1. <http://dx.doi.org/10.17877/DE290R-21089>
- Müller, H., Nordheim, G. von, Boczek, K., Koppers, L. & Rahnenführer, J. (2018). Der Wert der Worte – Wie digitale Methoden helfen, Kommunikations- und Wirtschaftswissenschaft zu verknüpfen. *Publizistik*, 63(4), 557–582. <https://doi.org/10.1007/s11616-018-0461-x>
- Nguyen, C. P., Le, T.-H. & Su, T. D. (2020). Economic policy uncertainty and credit growth: Evidence from a global sample. *Research in International Business and Finance*, 51, 101118. <https://doi.org/10.1016/j.ribaf.2019.101118>
- Stryker, J. E., Wray, R. J., Hornik, R. C. & Yanovitzky, I. (2006). Validation of Database Search Terms for Content Analysis: The Case of Cancer News Coverage. *Journalism & Mass Communication Quarterly*, 83(2), 413–430. <https://doi.org/10.1177/107769900608300212>
- Ulbricht, D., Kholodilin, K. A. & Thomas, T. (2017). Do Media Data Help to Predict German Industrial Production? *Journal of Forecasting*, 36(5), 483–496.
- Wahrig, G., Krämer, H. & Zimmermann, H. (Ed.). (1983). *Brockhaus Wahrig. Deutsches Wörterbuch*. Wiesbaden: F. A. Brockhaus & Deutsche Verlags-Anstalt GmbH.
- Waite, M. (2006). *Oxford Paperback Thesaurus* (3rd ed.). Oxford: Oxford University Press.
- Yu, H., Fang, L. & Sun, W. (2018). Forecasting performance of global economic policy uncertainty for volatility of Chinese stock market. *Physica A: Statistical Mechanics and its Applications*, 505, 931–940. <https://doi.org/10.1016/j.physa.2018.03.083>
- Zinn, J. O. (2010). Risk as Discourse: Interdisciplinary Perspectives. *Critical Approaches to Discourse Analysis Across Disciplines*, 4(2), 106–124.

7 Appendix

The appendix presents all search terms that are used in this research note, as well as the code book, the central instrument of the coding process.

A.1: Search terms

For clarity, the Boolean AND operators are highlighted in bold in all search terms.

The first preliminary search term consists of Baker, Bloom and Davis's original search term, excluding the component (unsicher OR unsicherheit OR unsicherheiten), but supplemented by the term "risk" and an extension of the policy areas related to economic uncertainties. The first preliminary search term reads:

(wirtschaft OR wirtschaftlich) **AND** (steuer OR wirtschaftspolitik OR regulierung OR regulierungs OR ausgaben OR bundesbank OR ezb OR zentralbank OR haushalt OR defizit OR haushaltsdefizit OR konjunktur OR regierung OR politik OR politisch OR boerse OR inflation OR handel OR schulden OR waehrung OR arbeitsmarkt OR finanz OR euro OR europa OR europaeisch) **AND** (risiko OR risiken OR riskant OR gefahr OR gefaehrlich OR gefaehrung OR bedrohung OR bedrohlich OR wagnis OR gewagt).

The second preliminary search term includes all components of Baker, Bloom and Davis's original search term, supplemented by the term "risk" and an extension of the policy areas, as well as additional related terms and synonyms of the term "uncertainty", which is already present in the original search term. The second preliminary search term reads:

(wirtschaft OR wirtschaftlich) **AND** (steuer OR wirtschaftspolitik OR regulierung OR regulierungs OR ausgaben OR bundesbank OR ezb OR zentralbank OR haushalt OR defizit OR haushaltsdefizit OR konjunktur OR regierung OR politik OR politisch OR boerse OR inflation OR handel OR schulden OR waehrung OR arbeitsmarkt OR finanz OR euro OR europa OR europaeisch) **AND** (unsicher OR unsicherheit OR unsicherheiten OR ungewiss OR unberechenbar OR unvorhersehbar OR instabil OR moeglich OR unklar OR unkalkulierbar OR unabsehbar OR risiko OR risiken OR riskant OR gefahr OR gefaehrlich OR gefaehrung OR bedrohung OR bedrohlich OR wagnis OR gewagt).

The author then optimizes the two preliminary search terms, using the findings from the first selection process, in order to obtain the two final search terms. For this purpose, the component of the search term which lists the policy areas is expanded again. The first final search term reads:

(wirtschaft OR wirtschaftlich) **AND** (steuer OR wirtschaftspolitik OR regulierung OR regulierungs OR ausgaben OR bundesbank OR ezb OR zentralbank OR haushalt OR defizit OR haushaltsdefizit OR konjunktur OR regierung OR politik OR politisch OR boerse OR inflation OR handel OR schulden OR waehrung OR arbeitsmarkt OR finanz OR euro OR europa OR europaeisch OR unternehmen OR industrie OR konzern OR klima OR globalisierung OR umwelt OR ressourcen OR technologie OR rohstoffe OR wettbewerb OR wachstum OR aktie OR bank OR prognose OR firmen OR branche OR rezession OR oel OR zinsen OR geldpolitik OR korruption OR export OR dollar OR nachfrage OR konsum OR produktion OR umsatz OR gewinn OR investition OR verlust OR krise) **AND** (risiko OR risiken OR riskant OR gefahr OR gefaehrlich OR gefaehrdung OR bedrohung OR bedrohlich OR wagnis OR gewagt).

The second final search term reads:

(wirtschaft OR wirtschaftlich) **AND** (steuer OR wirtschaftspolitik OR regulierung OR regulierungs OR ausgaben OR bundesbank OR ezb OR zentralbank OR haushalt OR defizit OR haushaltsdefizit OR konjunktur OR regierung OR politik OR politisch OR boerse OR inflation OR handel OR schulden OR waehrung OR arbeitsmarkt OR finanz OR euro OR europa OR europaeisch OR unternehmen OR industrie OR konzern OR klima OR globalisierung OR umwelt OR ressourcen OR technologie OR rohstoffe OR wettbewerb OR wachstum OR aktie OR bank OR prognose OR firmen OR branche OR rezession OR oel OR zinsen OR geldpolitik OR korruption OR export OR dollar OR nachfrage OR konsum OR produktion OR umsatz OR gewinn OR investition OR verlust OR krise) **AND** (unsicher OR unsicherheit OR unsicherheiten OR ungewiss OR unberechenbar OR unvorhersehbar OR instabil OR moeglich OR unklar OR unkalkulierbar OR unabsehbar OR risiko OR risiken OR riskant OR gefahr OR gefaehrlich OR gefaehrdung OR bedrohung OR bedrohlich OR wagnis OR gewagt).

Before the second sample is drawn, an open search term is used to select all articles from the original corpus of 2,723,049 texts that are relevant according to the code book. The open search term reads:

(wirtschaft OR wirtschaftlich) **AND** (unsicher OR unsicherheit OR unsicherheiten OR ungewiss OR unberechenbar OR unvorhersehbar OR instabil OR moeglich OR unklar OR unkalkulierbar OR unabsehbar OR risiko OR risiken OR riskant OR gefahr OR gefaehrlich OR gefaehrdung OR bedrohung OR bedrohlich OR wagnis OR gewagt).

A.2: Code book

Category: relevance of the article (RELEVANCE)

Table 3: Relevance of the article (Source: author's work)

Variable ID	Variable name	Code	Variable manifestations	Explanations
RELEVANCE	relevance of the article	1 0	relevant not relevant	<p>This category indicates whether an article is relevant; this means, whether an article depicts a (possible) economic uncertainty or a (possible) economic risk, or includes statements regarding a (possible) economic uncertainty or a (possible) economic risk.</p> <p>Note: An article is considered relevant if it contains at least one depiction of or statement about a (possible) economic uncertainty or a (possible) economic risk. These can be connected with topics such as taxes, economic policy, regulatory measures, spending, banks, budgetary measures, the economy, government, politics, stock market, inflation, trade, debt, currency, labor market, finance, Europe etc.</p>