Larsen, V.H. (2021), COMPONENTS OF UNCERTAINTY. International Economic Review, 62: 769-788. https://doi.org/10.1111/iere.12499

Components of Uncertainty*

Vegard Høghaug Larsen[†]

September 10, 2020

Abstract

Uncertainty is acknowledged to be a source of economic fluctuations. But, does the type of uncertainty matter for the economy's response to an uncertainty shock? This paper offers a novel identification strategy to disentangle different types of uncertainty. It uses machine learning techniques to classify different types of news instead of specifying a set of keywords. The paper finds that, depending on its source, the effects of uncertainty on a macroeconomic variable may differ. I find that both good (expansionary effect) and bad (contractionary effect) types of uncertainty exist.

JEL-codes: D80, E32, E66

Keywords: Newspaper, Topic model, Uncertainty, Business cycles, Machine learning

*This working paper should not be reported as representing the views of Norges Bank. The views expressed are those of the author and do not necessarily reflect those of Norges Bank. The research reported in this paper was mostly carried out while the author was a graduate student at the Centre for Applied Macro and Petroleum economics (CAMP) at the BI Norwegian Business School. I am grateful for research assistance from Helene Olsen. I would like to thank two anonymous referees, Farooq Akram, Hilde C. Bjørnland, Gernot Doppelhofer, Ragnar E. Juelsrud, Jon Fiva, Gisle Natvik, Juan Rubio-Ramírez, Tommy Sveen, and Leif Anders Thorsrud, as well as conference participants at the SITE Summer Workshop 2017, for valuable comments.

[†]Norges Bank. Email: `vegard-hoghaug.larsen@norges-bank.no`

1 Introduction

A large and growing literature investigates the effect of elevated uncertainty on aggregate macroeconomic fluctuations. Most uncertainty measures tend to be countercyclical, and several studies document that an increase in uncertainty is followed by worsening economic conditions, see e.g., Bloom (2009), Jurado et al. (2015), and Baker et al. (2016). Common to these studies is the construction of uncertainty measures that capture similar types of events related to episodes of financial and economic distress. However, measuring uncertainty that consistently rises in bad times makes it difficult to study potential alternative effects of uncertainty.

This paper offers an identification strategy to disentangle different types of uncertainty. It relies on machine learning techniques to uncover the content of a large set of news articles published in a daily business newspaper. Using a structural VAR model, this paper shows that depending on the source, uncertainty may have different effects on the same macroeconomic variables.

Specifically, I create measures of uncertainty by first classifying news articles according to theme, and then quantifying uncertainty by the count of uncertainty terms within the different types of news. This method belongs to the field of topic modeling, where the objective is to identify hidden patterns in textual data. I estimate the content of news articles using Latent Dirichlet Allocation (LDA), introduced by Blei et al. (2003). The method is an unsupervised learning algorithm, meaning that there is no pre-training of the model or labeling of the news articles before the classification. I identify well-defined uncertainty measures related to categories of high economic relevance such as *Oil price*, *Monetary policy*, *Politics*, and *Stock market*.

Using textual data to extract uncertainty has become popular:[1] For instance Alexopoulos and Cohen (2009) create an uncertainty measure based on the number of New York Times articles about both uncertainty and economic activity. Baker et al. (2016) create Economic Policy Uncertainty (EPU) indices for various countries by counting articles about uncertainty, the economy, and policy. Common to these papers is that they classify articles by a set of pre-determined keywords, and if an article contains words from all categories, it contributes to the index.[2] I propose to use a topic model to classify different

[1] Using text as data has exploded in recent years, see Gentzkow et al. (2019) for an overview.

[2] Baker et al. (2016) also identify narrower category-specific uncertainty measures by counting articles with

types of news instead of relying a set of keywords for the classification. An advantage of the topic model is that the classification does not rely on the article containing a particular set of words. Instead, the mixture of all the words in an article determines the theme of that article.[3] I use news articles published over more than 30 years in Norway's largest business newspaper, *Dagens Næringsliv*. Few papers in economics use a topic model to extract information from textual data. A related paper is Larsen and Thorsrud (2019), who create a topic-based news index to study the impact of news and noise shocks on the business cycle in Norway. Another example is Hansen et al. (2017), who study how transparency affects monetary policymakers, deliberations using a topic model to classify textual data from the Fed.

I investigate the validity of this topic-based approach by evaluating the uncertainty measures in two ways: First, I do a narrative exercise evaluating whether the uncertainty measures capture known historical events where we expect uncertainty to be high. Second, I compare the uncertainty measures with other proxies for uncertainty such as the US VIX, realized stock market volatility in Norway, and some of the economic policy uncertainty measures created by Baker et al. (2016). Overall, the topic-based measures capture well-known historical events and there is a tendency for positive correlations between the topic-based measures and the alternative ones.

Uncertainty shocks can have real and substantial negative effects on firm investment and hiring, because firms delay taking action. This is often referred to as "wait and see" behavior, see e.g., Bernanke (1983), McDonald and Siegel (1986) and Bloom (2009). Uncertainty also affects households: Elevated uncertainty can increase precautionary savings and thereby deflate aggregate demand in the economy, see, e.g., Basu and Bundick (2017), Leduc and Liu (2015) and Fernandez-Villaverde et al. (2011). Uncertainty can affect financial markets, where higher firm risk leads to increased cost of capital and more cautionary behavior by investors, see, e.g., Gilchrist et al. (2014) and Arellano et al. (2010). On the other hand, some papers argue for a positive effect of uncertainty, so called "growth options" theories, where the willingness to invest can increase due to an improved upside in the economy, see e.g. Segal et al. (2015) and Kraft et al. (2013). Another mechanism for a positive effect of increased uncertainty is the "Oi-Hartman-Abel" effect, originating from Oi (1961), Hartman (1972) and Abel (1983). In this scenario firms can easily reduce

words from specific categories such as *National security* and *Health care*.

[3] A related paper using machine learning techniques to extract uncertainty is Manela and Moreira (2016).

costs if a bad outcome from uncertainty materializes, while in the case of a good outcome has the potential to bring large rewards.

There is a large literature estimating the economic response to an uncertainty shock. There are papers analyzing different, but correlated, types such as macroeconomic uncertainty (Bloom (2009) and Jurado et al. (2015)), economic policy uncertainty (Baker et al. (2016)), and fiscal policy uncertainty (Fernandez-Villaverde et al. (2015)).

I analyze the impact of shocks to different uncertainty measures, on aggregate economic fluctuations by estimating a structural VAR model using narrative sign restrictions as proposed by Antolín-Díaz and Rubio-Ramírez (2018).[4] This approach is well suited to my framework since it is easy to check the narratives in different episodes by reading the news. I find that different types of uncertainty have different implications for the economy. A shock to uncertainty related to *Macroeconomics* foreshadows declines in investment and output in line with previous studies. The effect is sizable and economically important. On the other hand, a shock to uncertainty related to *Mergers & acquisitions*, leads to a significant increase in investment and output. The finding that uncertainty can have both positive and negative effects indicates that both good and bad types of uncertainty exist.

The rest of the paper is organized as follows: Section 2 describes the newspaper data, the topic model and how the uncertainty measures are constructed. Section 3 discusses and evaluates the uncertainty measures. Section 4 discusses some potential channels through which uncertainty can affect the economy. In Section 5, investigates the effect of uncertainty shocks on aggregate macroeconomic variables. Section 6 concludes.

2 Measuring category-specific uncertainty

This section describes the newspaper data and how the articles are classified according to their underlying content. I describe how the uncertainty of the articles is quantified as well as how these measures of uncertainty are combined with the classification of the news articles to create topic-based measures of uncertainty.

[4] For Norway, Gudmundsson and Natvik (2012) create an uncertainty measure in the same way as Alexopoulos and Cohen (2009) and find negative effects of uncertainty shocks on consumption.

2.1 The newspaper data

The raw data used are articles from *Dagens Næringsliv*, which is Norway's largest business newspaper and also the fourth largest newspaper overall. The news data has been generously provided by the company Retriever through their "Atekst" database, and collected manually for the latter part of the sample. The paper was founded under the name *Norway's Trade and Seafaring Times* in 1889 by Magnus Andersen and has a right-wing and neoliberal political stance. I use all articles published in the paper version of the newspaper from May 2 1988 to December 31 2018. During this period, there were two editors-in-chief: Kåre Valebrokk (1985–1999) and Amund Djuve (2000-current). The data consist of close to 500 000 articles, spread over a period of more than 8000 days. This is a large amount of data that are highly unstructured and, in line with the literature on modeling text, several steps are performed to clean and reduce the data to a more manageable form. First, I remove words that would not convey any important meaning for the underlying theme of a news story, examples of such words are *the*, *is*, and *are*. I also remove common Norwegian surnames and given names. Next, each word is reduced to its word stem.[5] Lastly, I calculate a corpus measure called the *tf–idf* score which stands for term frequency–inverse document frequency.[6] This is a way of scoring a word in the corpus based on how frequently it occurs in a single document, relative to how frequently it occurs in the whole text corpus. I select a cutoff for this *tf–idf* score and discard the words with the lowest relative importance for informing us of the content in single documents.[7] I keep around 250 000 of the stems with the highest *tf–idf* score and move on to the classification using an LDA model.[8]

[5] The word stem is the part of a word that is common to all the word's inflections, an example is the word *production*, which has the word stem *produc*.

[6] See Gentzkow et al. (2019) for how to compute the *tf–idf* score.

[7] An example of an informative word is "barrel" which is not used much in the whole corpus of documents, but it occurs in news about oil, and therefore the word "barrel" is informative for identifying documents about oil. Calculating the *tf-idf* score is not absolutely necessary since the LDA does a similar job when selecting the relevant words for the various topics. The main reason for doing this is to reduce the number of words in the corpus, to ease the computational burden when estimating the LDA.

[8] The corpus reduction and cleaning are standard in the natural language processing literature, maybe with the exception of removing the surnames and given names. This choice is made because many persons share the same names, and names often occur in the newspaper; including them will only pollute the underlying meaning of the article since the algorithm gives the same "meaning" to all unique names.

2.2 Latent Dirichlet Allocation

The LDA is a model that allows sets of observed documents to be explained by latent structures that determine why documents belong together. It is an unsupervised learning algorithm, meaning that there is no labeling of the articles or training of the model before the articles are classified. It is assumed that all documents are constructed by combining a given set of themes or topics and then drawing words from these topics. Each article is a random mixture of all the topics. The word "topic" is used frequently in this paper and it refers to a distribution over a fixed vocabulary. All the observed words in the newspaper have a positive probability of occurring in all the topics, and all the topics occur with a positive probability in all of the documents. To apply some structure to the model, I follow the exposition in Heinrich (2005) and assume that an article can be represented as a mixture of latent topics. The latent topic is referred to as z, and we have a fixed number, K, of these topics. A document is referred to as d and we have M of them. The objective when estimating the LDA is twofold: First, for the observed words, w, we want to find the word distribution $p(w|z = k) = \boldsymbol{\theta}_k$ for all topics k. Second, we want to find the topic distribution $p(z|d = m) = \boldsymbol{\varphi}_m$ for all documents m. The LDA is a generative model that works as follows:

(i) For all topics $k \in [1, K]$ sample word mixtures as $\boldsymbol{\theta}_k \sim \text{Dir}(\beta)$

(ii) For all documents $m \in [1, M]$ sample topic mixtures as $\boldsymbol{\varphi}_m \sim \text{Dir}(\alpha)$

   (a) For all words $n \in [1, N_m]$:

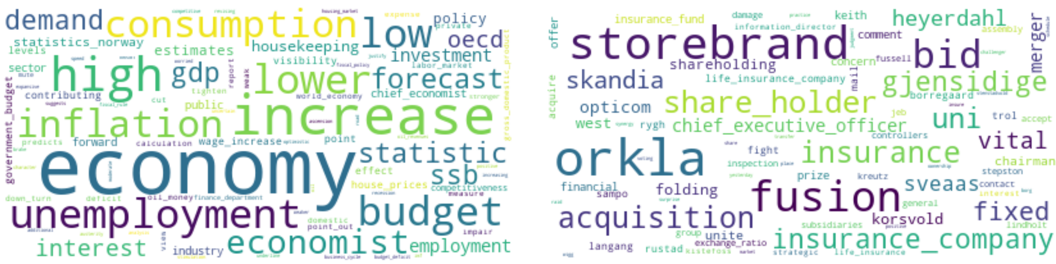
      (a-1) Sample topic index $z_{m,n} \sim \text{Multinomial}(\boldsymbol{\varphi}_m)$

      (a-2) Sample word $w_{m,n} \sim \text{Multinomial}(\boldsymbol{\theta}_{z_{m,n}})$

where α and β are priors on the Dirichlet distributions for the topic mixtures and the word mixtures respectively. We assume that the documents were generated this way, but in reality, we only observe the outcomes, the published news articles. We use the generative model for how the articles were created together with the realized articles to infer the underlying structure of the $\boldsymbol{\theta}_k$s and the $\boldsymbol{\varphi}_{\boldsymbol{m}}$s. The topics are estimated by starting out with a given set of word distributions where the probabilities of the different words occurring are random. Then we improve these distributions by changing the probabilities and evaluating how well they describe the documents. I use a Bayesian approach to estimate the topic

Figure 1. Examples of word mixtures, $\boldsymbol{\theta}_k$

(a) Macroeconomics **(b)** Mergers & Acquisitions



Note: The 150 words with the highest probabilities are shown. The size of the words corresponds to the probability of that word occurring in the topic distribution. Words that are connected by an underscore represent single words that become phrases when translated from Norwegian to English. All the word clouds are available at http://www.vegardlarsen.com/Word_clouds/.

model using Gibbs simulations. The estimation procedure follows the algorithm described in Griffiths and Steyvers (2004). The topic model is estimated on data up until 2015, and the last four years of data are classified using the previously estimated topics. [9]

Before estimating the topic model, I need to specify the number of topics to be identified, and I set $K = 80$. The choice is made based on a subjective evaluation of the topics. I find that 80 topics are preferable to fewer topics, where I observe that different concepts are grouped together in one topic distribution. The goal is not to find the topic model that best describes the documents, but rather a model that delivers topics that give a reasonable description of the newspaper and the Norwegian economy. I find that 80 topics give a good result, where the topics are neither too broad nor too narrow. Chang et al. (2009) show that improving the document fit of a topic model by e.g. increasing the number of topics can lead to semantically less meaningful topics. Increasing the number of topics is also problematic computationally.

The output from the topic model is two sets of distributions: one set of distributions over words, $\boldsymbol{\theta}_k$, for all topics $k \in \{1, \dots, K\}$, and one set of distributions over topics, $\boldsymbol{\varphi}_m$, for all articles in $m \in \{1, \dots, M\}$.

Figure 1 shows two examples where the word distributions are represented as word clouds. The size of the word in the word cloud corresponds to the probability of that

[9] Ke et al. (2019) show that the parameters in the LDA model are set-identified and the choice of priors will affect the model's output. I leave it for further research to investigate how this affect the results when using output from the LDA in structural modeling.

word occurring in the given topic. The topics are given by the word distributions, and are not given any label by the topic model. Since referring only to topic numbers gives very little meaning, and since I want an economic interpretation of the different topics, I label the topics. The labeling is done by visual inspection of the word distributions and then picking a word that gives a reasonable description of the distribution. Most topics convey a clear theme or category. A list of all the 80 topics and their labels, together with a list of the 10 most frequent words occurring in each topic, is given in Table 2 in Appendix B. I get topics related to the aggregate economy such as *Macroeconomics* and *Monetary policy*, topics related to financial markets such as *Banking* and *Funding*, topics related to politics such as *Politics* and *Elections*, and international topics such as *USA* and *Asia*.

An alternative approach to classifying an article is by identifying specific keywords that are linked to specific categories. By searching through all the articles and looking up these keywords, we can classify the articles according to some pre-specified categories. This is the approach taken by Baker et al. (2016), and I follow this approach in creating an index for Norway to compare with the category-specific uncertainty measures that is the focus of my paper.[10]

2.3 Quantifying uncertainty

I create measures of uncertainty by combining the article classification discussed above with a number representing the level of uncertainty calculated for each article. To quantify the extent to which a news article signals uncertainty, I count the terms related to uncertainty within that article.

I start out by counting the term *uncertain* and its inflections for all the articles.[11] The count of uncertainty terms in article m is given by

$$\upsilon_m = \text{number of uncertainty terms in article } m. \tag{1}$$

To control for a varying amount of news coverage over time, I keep track of the total

[10]The details of this Norwegian version of the Baker et al. (2016) index can be found in Appendix D.

[11]The words that are counted (given in Norwegian): *usikker, usikre, usikkert, usikkerhet, usikkerheter, usikkerheten, usikkerhetene*. I have also experimented with using a broader list of words including terms such as *risk* and *unpredictability*, and this gives indices that lie close to the ones created in this paper. An advantage of using a broader list is that more articles get a non-zero uncertainty term count, which gives us a richer measure. I choose to use only the terms directly affiliated with uncertainty to create a measure that is clean and easy to interpret.

number of words in article m given by:

$$\omega_m = \text{number of total words in article } m. \tag{2}$$

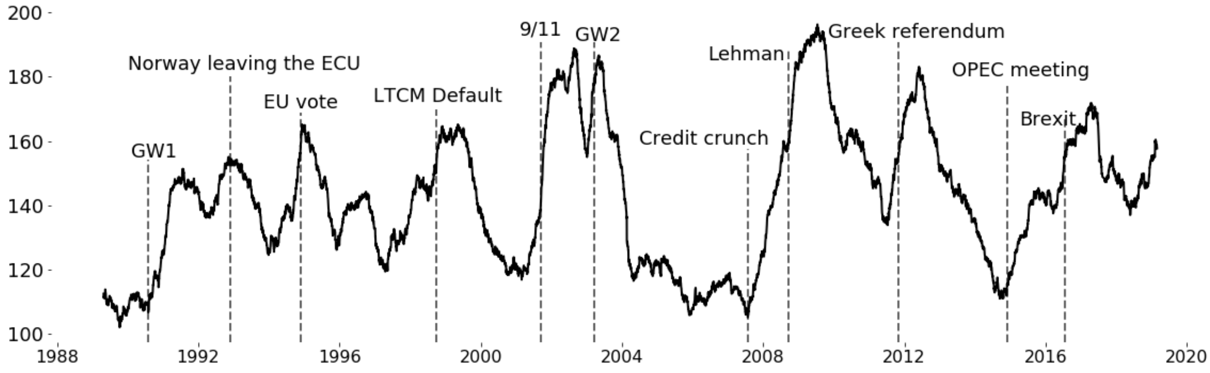
As a first approach, I calculate an aggregate daily measure, that is the overall uncertainty count in the newspaper each day. Calculating an aggregate count reflects uncertainty about many different concepts, such as sports, the economy, political elections etc. Even if the interpretation of this aggregate uncertainty measure is unclear, it is a point of departure, before looking at the more disaggregated measures. I calculate the aggregate uncertainty measure as follows:

$$\Upsilon_t^{\text{Agg}} = \frac{\sum_{m \in \text{ day } t} \upsilon_m}{\sum_{m \in \text{ day } t} \omega_m}. \tag{3}$$

On each day, the total count of the uncertainty terms is divided by the total word count that day. Figure 2 plots this aggregate measure as the 300 days backward-looking mean.[12] Over the sample the total daily count of the word *uncertain* and *uncertainty* varies approximately between 100 and 200 out of one million words. From the figure we see that there are large variations in the uncertainty measure and that there are clear episodes where aggregate uncertainty is high. I plot some events that coincided with significant increases in uncertainty. Based on these events it appears that the uncertainty count in *Dagens Næringsliv* is driven mostly by foreign crises such as wars and international financial crises. The episodes that are displayed are the first and the second gulf war (GW1 and GW2), the Long-Term Capital Management (LTCM) default, the 9/11 terrorist attacks, the credit crunch (often considered as the start of the financial crisis), the collapse of Lehman Brothers, the Greek proposed referendum related to a bailout of the Greek government, the OPEC meeting in fall 2014 after the large drop in oil prices, and the UK deciding to leave the EU (Brexit). The only Norwegian events displayed in the figure are the referendum on joining the European Union, and Norway depegging its currency from the European Currency Unit (ECU). Of course, many of the episodes where uncertainty is high in the figure coincide with other Norwegian events such as the banking crisis in the early 1990s and a short recession in 2002–2003 and 2008–2009.

[12] The backward-looking mean is used because it reduces noise and makes it easy to identify episodes of high uncertainty. This is done for visual clarity, and all empirical results presented are based on the measures at a daily, monthly or quarterly frequency.

Figure 2. Aggregate newspaper uncertainty



Note: The black line plots the 300 day backward-looking rolling mean. The series gives the share of uncertainty terms per 1 million words in the newspaper.

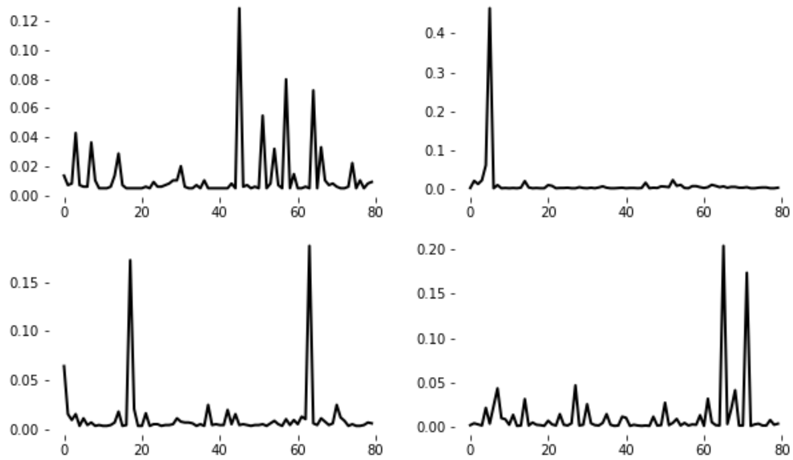
2.4 Topic-based measures of uncertainty

The category-specific uncertainty measures are created based on the uncertainty count within the categorized news articles. The topic model delivers the classification of all news articles. The classification is a probability distribution over all topics reflecting content in the articles that relates to several topics at once. I calculate an uncertainty measure for all the different news topics by weighing the uncertainty counts using the relative contribution of all articles to the different topics. That is, article m has an uncertainty count given by v_m, which contributes by $\boldsymbol{\varphi}_m(\text{topic} = k)$ to topic k. To see what these topic distributions, $\boldsymbol{\varphi}_m$, look like, Figure 3 plots such topic distributions for four news articles. These distributions tell us how much the uncertainty count from the articles they represent contributes to the uncertainty indexes for the various topics. We see that for some articles there is one or a few topics that describe the content of the article, while others are a broader mix of topics. *Dagens Næringsliv* is a business newspaper, as seen by the large share of topics that relate to business and economics.[13] Thus, an article about the economy is likely to be a mix of economy-related topics. On the other hand, there are very few topics related to sports, so a sports-related article is more likely to be described by few topics.

The total amount of content in a newspaper varies over time, as does the coverage of an individual news topic. To control for this, I normalize with respect to the amount of

[13]The topics are listed in Table 2 in Appendix B.

Figure 3. Document mixtures, φ_m, for four random articles



Note: The document distributions, φ_m, for four randomly drawn articles, the numbers on the x-axis represent the topics and the corresponding label is found in Table 2 in Appendix B.

news content on any given day. The more articles and words we observe in one day, the more uncertainty terms we expect to observe in total. For the baseline normalization, I divide the topic-specific uncertainty term count within one day by the total number of words that day.[14] This uncertainty measure is given by:

$$\Upsilon_{k,t} = \frac{\sum_{m \in \text{ day } t} \upsilon_m \varphi_m(\text{topic} = k)}{\sum_{m \in \text{ day } t} \omega_m}. \tag{4}$$

One alternative specification is to divide by the total number of words used within a specific news category.[15] The denominator is important to consider because it in itself causes fluctuations in the uncertainty measure. I choose to use the normalization in Equation 4 as the baseline, because fluctuations in the coverage of a given news topic can vary substantially. Daily fluctuation in topic coverage can have large effects on the alternative uncertainty measure, and this variation is not driven by the uncertainty count.

What types of news categories use the uncertainty terms the most? Table 1 reports

[14]Dividing by the total daily count is in line with the literature, see e.g. Baker et al. (2016).

[15]This alternative measure is calculated as

$$\tilde{\Upsilon}_{k,t} = \Big(\sum_{m \in \text{ day } t} \upsilon_m \varphi_m(\text{topic} = k) \Big) / \Big(\sum_{m \in \text{ day } t} \omega_m \varphi_m(\text{topic} = k) \Big).$$

I have also computed these measures, and for most topics they give very similar results. The average correlation between the two measures is 0.86.

Table 1. Uncertainty share in different news categories

Top 10	Mean	Std. dev.	Bottom 10	Mean	Std. dev.
Monetary policy	5.7	11.2	*Drinks*	0.9	1.7
Stock market	4.6	8.0	*Movies/Theater*	0.9	1.8
Macroeconomics	4.4	9.1	*Food*	1.0	1.5
Fear	3.7	5.5	*Literature*	1.0	2.3
Oil price	3.2	6.8	*Art*	1.0	2.4
Debate	3.0	5.1	*Music*	1.0	2.1
Negotiation	2.4	2.9	*Watercraft*	1.1	2.1
Elections	2.4	6.8	*Family business*	1.1	1.6
Oil production	2.3	5.5	*Tourism*	1.1	2.4
Results	2.3	3.7	*Sports*	1.1	2.2

Note: The mean and standard deviation of uncertainty terms used in the different types of news. The number of uncertainty terms per 1 million words in the newspaper.

the 10 news categories with the largest number of uncertainty terms, and also the 10 news categories with the lowest count. The news category where the newspaper writes the most about uncertainty is *Monetary policy*. During the period studied, Norway had five different monetary policy regimes, and this may have led to increased uncertainty. The news category with the second highest uncertainty count is *Stock market*, followed by *Macroeconomics*, *Fear*, and *Oil price*. The *Fear* topic is a news topic where the word *uncertainty* is one of the words with the highest probability and the frequency of the *Fear* topic itself is a possible proxy for uncertainty. The type of news where the uncertainty terms are least frequent are *Drinks*, *Movies/Theater* and *Food*.

3 Evaluating the topic-based uncertainty measures

The topics are identified by an unsupervised learning algorithm, and uncertainty is identified as the frequency of uncertainty terms within news related to various topics. There is no subjectivity involved other than the labeling of the topics, which is a way of referring to the word distributions. I evaluate whether the uncertainty measures capture what they are supposed to, which is the underlying uncertainty related to various themes or categories. I do a narrative exercise where I plot some of the uncertainty measures together with episodes where it is reasonable to think that uncertainty is high. To conserve space, I discuss eight of the 80 measures.

The first four measures are selected based on the type of news that uses the uncertainty terms the most. The remaining four measures are based on news categories that can easily be linked to well-known historical events. An example is oil price uncertainty, which we expect to be high during episodes with, say, conflicts in oil-producing regions. Finally, I evaluate the full set of measures by comparing them with other proxies for uncertainty.

3.1 Narrative evaluation

The first four examples of category-specific uncertainty are chosen by selecting the news topics where the uncertainty terms are used with the highest frequency. These topics are *Monetary policy*, *Stock market*, *Macroeconomics*, and *Fear*. The top four measures are plotted in Figure 4 together with some notable events where uncertainty was high. The exact dates and a short description of the events can be found in Table 3 in Appendix B.

In Panel (a) in Figure 4, I plot the measure for *Monetary policy* uncertainty. It is plotted together with the dates when the monetary policy regime changed, as well as when a new central bank governor assumed office. We see that uncertainty tends to be elevated around these events. Uncertainty was especially high during the second part of the 1990s. This was a period when Norway had a debate on which monetary policy regime should be implemented. The monetary policy regime in Norway changed four times during the sample studied here. Uncertainty also increased during global events such as the Lehman Brothers bankruptcy and the Greek government-debt crisis. In Panel (b) Figure 4, the series for *Stock market* uncertainty is plotted. This measure captures well-known events of heightened uncertainty, such as the debate in Norway on whether or not to join the EU, the LTCM default, the short Norwegian recession in the early 2000s, and the Global Financial Crisis. *Stock market* uncertainty tends to increase when the stock market is in decline. In, Panel (c) in Figure 4, I plot the *Macroeconomic* uncertainty measure. This series captures many of the same events as *Stock market* uncertainty, but there are a few periods where the two measures diverge: First, the *Macroeconomics* measure captures more uncertainty in the early 1990s during both the Norwegian banking crisis and the episodes of changing monetary policy regimes. Second, we see a large surge in *Macroeconomics* uncertainty after the oil price fall that started in the summer of 2014. The *Macroeconomics* measure is countercyclical and has a negative correlation with the business cycle. Panel (d) in Figure 4 plots the frequency of uncertainty terms within news classified as *Fear*. The *Fear* topic is a type of news that gets considerable coverage during a crisis. The measure is especially high during the Global Financial Crisis and the Greek government-debt crisis.

The first four examples often capture related events. The average correlation between the four measures at a daily (quarterly) frequency is 0.28 (0.43). Turning to measures capturing more distinct types of uncertainty, I plot in Figure 5 uncertainty related to *Oil price*, *Telecommunication*, *International conflicts*, and *Politics* as examples. The average correlation between these four measures is 0.13 (0.20) at a daily (quarterly) frequency.

Panel (a) in Figure 5 displays the series for *Oil price* uncertainty, which is important for Norway as a large oil exporter, cf. Bjørnland and Thorsrud (2016). By inspecting the spikes in *Oil price* uncertainty, they seem to be driven mostly by foreign events, often related to unrest in the Middle East or global financial crises. Hamilton (2013) identifies historical oil shocks and notes that all his shocks coincide with elevated *Oil price* uncertainty. In Panel (b) of Figure 5, I plot uncertainty related to *Mergers & Acquisitions*. The series is plotted together with some important events: The first large rise in this uncertainty measure is related to the banking crisis in Norway in the early 1990s when there were discussions on merging banks to increase the resiliency of the banking sector. The second large increase in uncertainty is related to an eventually unsuccessful merger between the insurance company Gjensidige and the bank Kredittkassen. We also see a large increase in uncertainty in the late 1990s and early 2000s, coinciding with the dot-com boom in this period. Lastly, in late 2006 it was proposed that the two largest oil companies in Norway should merge, and this proved to be a success.

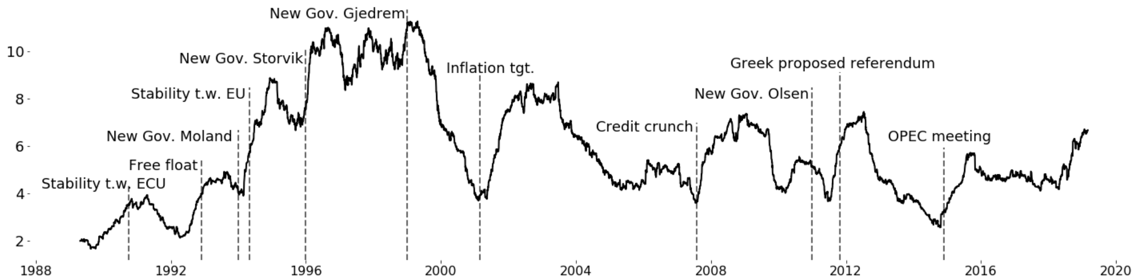
Panel (c) in Figure 5 shows uncertainty related to *International conflicts*. The series picks up well-known conflicts such as the first and second Gulf War, and several episodes during the Arab spring. The uncertainty measure is especially high during the first and second Gulf Wars, which were likely given wide coverage in the business newspaper due to the effect on the oil price.

Lastly, in Panel (d) in Figure 5, I show uncertainty related to *Politics*. The dates for the parliamentary elections and local elections in Norway are indicated by the red and green dashed lines respectively. I also indicate whether there is a left-leaning (red) or right-leaning (blue) central government in office. The uncertainty tends to increase around the parliamentary elections.[16]

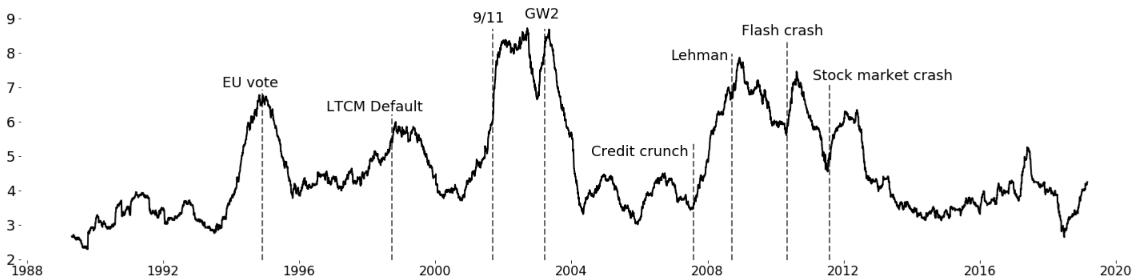
[16] We also see a large surge in uncertainty around the time of the local election in 2011. However, this also likely captures an increase in uncertainty related to the terrorist attacks on the government headquarters

Figure 4. Examples of uncertainty measures I

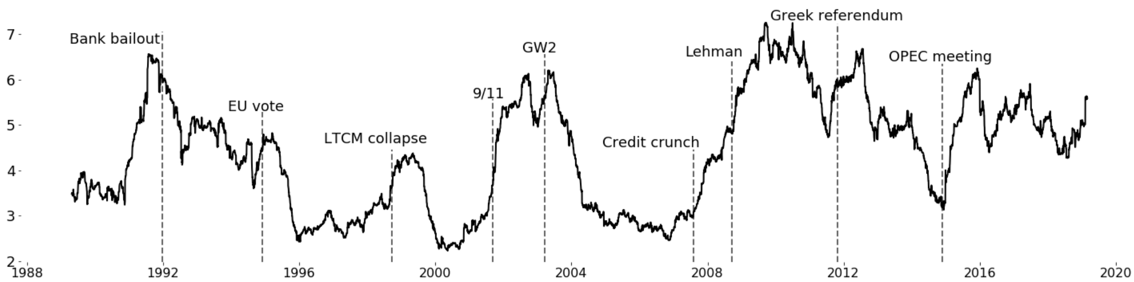
(a) ***Monetary policy*** uncertainty



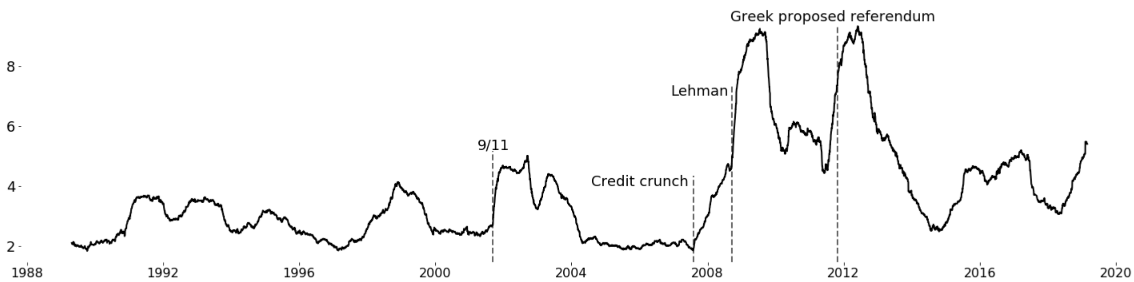
(b) ***Stock market*** uncertainty



(c) ***Macroeconomics*** uncertainty



(d) ***Fear*** uncertainty



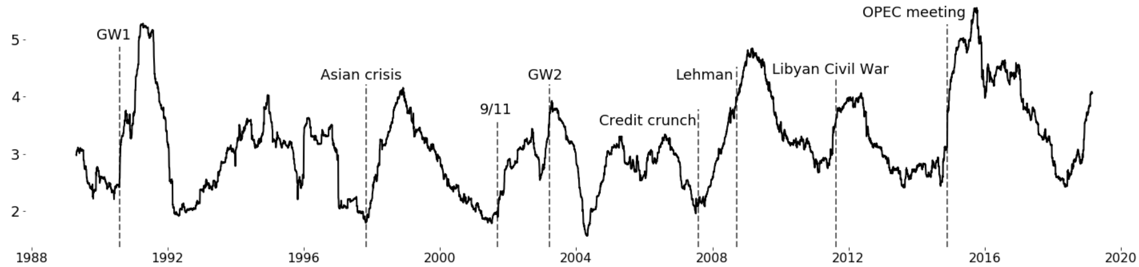
Note: The black line plots the 300 day backward-looking mean. The uncertainty count is the number of uncertainty terms per 1 million words in the full newspaper. 1.

---

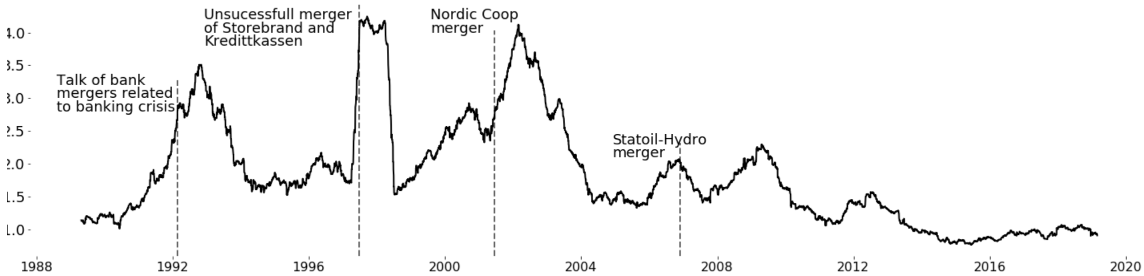
and on the Workers' Youth League summer camp on July 22.

Figure 5. Examples of uncertainty measures II

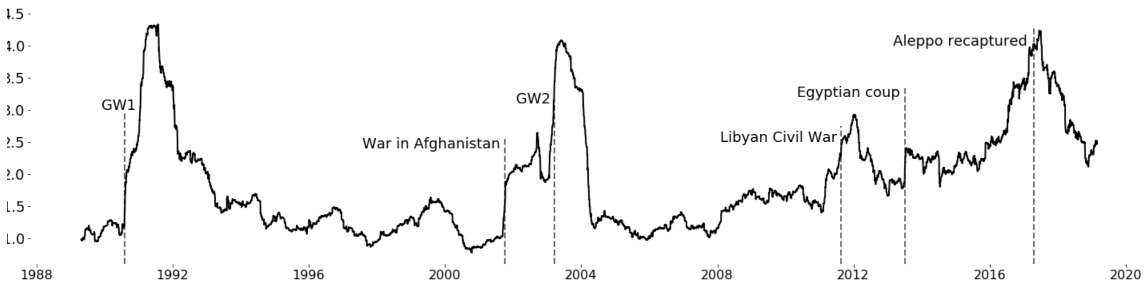
(a) *Oil price* uncertainty



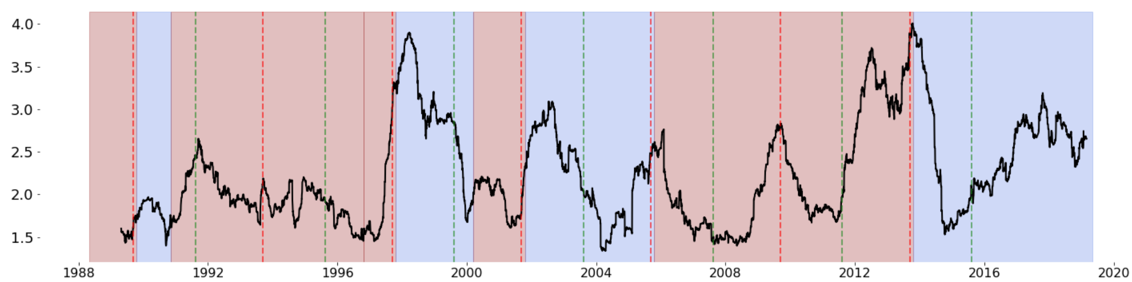
(b) *Mergers & Acquisitions* uncertainty



(c) *International conflicts* uncertainty



(d) *Politics* uncertainty



Note: For details see Figure 4. In Panel (d) the vertical red dashed lines represent parliamentary, and the green dashed lines local, elections in Norway. The areas shaded in red represent periods with a left-leaning government, and the blue shaded areas represent right-leaning governments.

3.2 Comparison with alternative measures

I compare the topic-based uncertainty measures with some alternative measures of uncertainty. As there is limited availability of uncertainty measures for Norway, I generate two alternative measures: First, Norway has no options-based stock market volatility index, and I calculate a realized stock market volatility (RSMV) measure. The RSMV series is calculated as the monthly standard deviation of the Oslo stock exchange benchmark index (OSEBX).

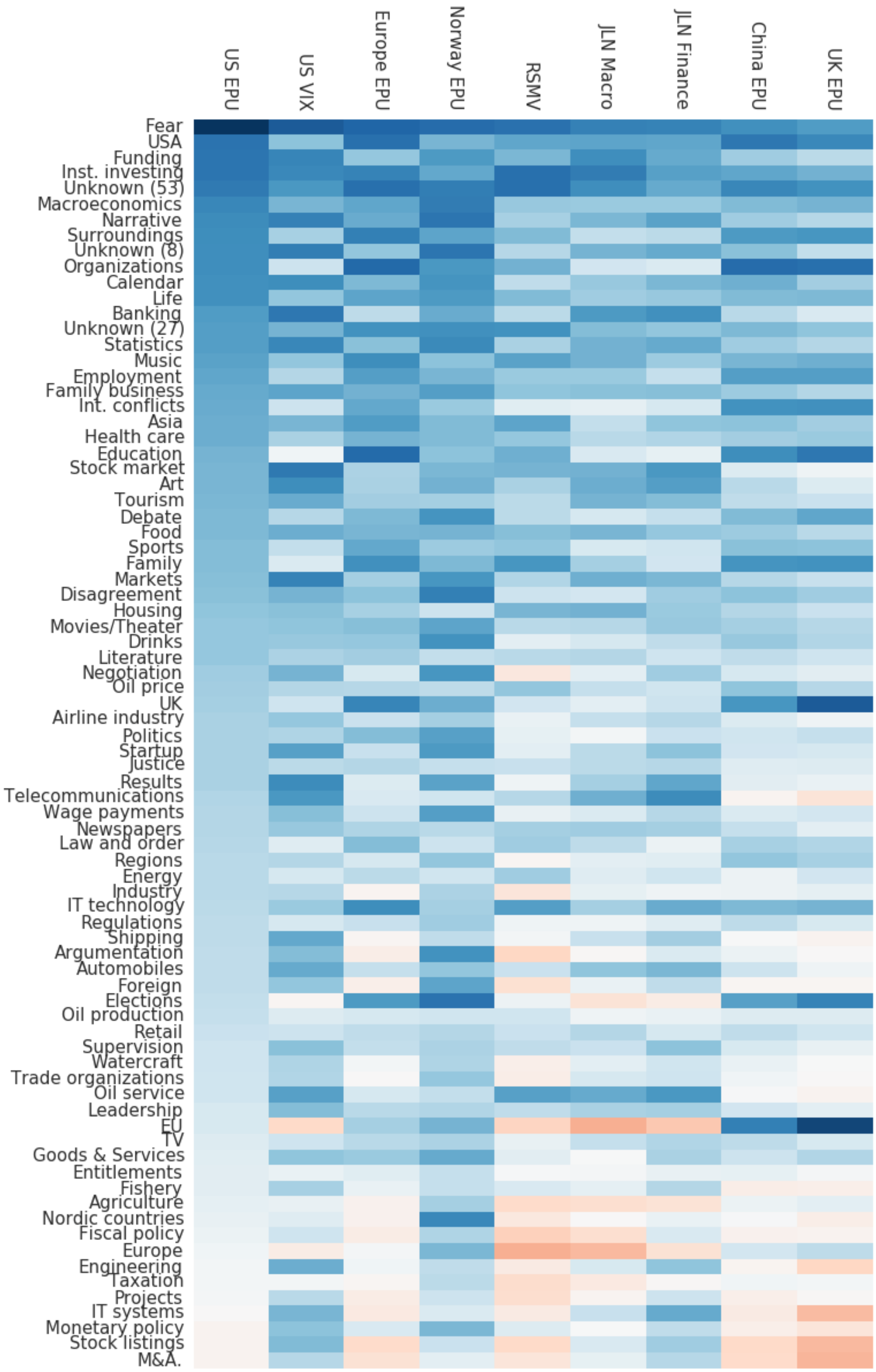
The second measure is a Norwegian version of the EPU as created by Baker et al. (2016). The details on how the Norwegian EPU is computed can be found in Appendix D. In addition, I look at seven foreign measures. Those are: the US VIX, the macroeconomic and financial uncertainty measures from Jurado et al. (2015), and the EPU measures for the US, the UK, Europe and China created by Baker et al. (2016).[17]

Figure 6 displays the correlations between all the 80 topic-based measures and the nine alternative ones. The figure is a heat map where negative correlations are in shades of red and positive correlations are in shades of blue. The highest correlation, 0.71, is between the *Fear* measure and the US EPU. This observation is placed in the top left corner of the heat map, and I sort the rows and columns in descending order away from this point. The lowest correlation, -0.25, is between the *Europe* and the RSMV measure. Some notable results emerge from Figure 6:

First, almost all the topic-based measures have a positive correlation with the alternative ones. This indicates that most of the measures, both the topic-based and the alternative ones, capture similar events. Two notable exceptions are the *EU* and *Europe* measures, which have a negative correlation with several of the alternative measures. In part of the sample, these measures capture a Norway-specific type of uncertainty related to the referendum on membership of the European Union.

[17] Jurado et al. (2015) create an uncertainty measure based on the unforecastable component of a large set of economic variables. The Jurado et al. (2015) paper focuses on macroeconomic uncertainty. In a related paper, using similar data, Ludvigson et al. (2015) disentangle macro and financial uncertainty. I refer to both the macro and the finance measure as JLN measures (Jurado, Ludvigson, and Ng), because I downloaded the measures from the supplementary material from Jurado et al. (2015) (`http://www.columbia.edu/~sn2294/pub.html`).

Figure 6. Correlations with alternative measures



Note: The correlations are computed at a monthly frequency. Blue represents a positive correlation, while red represents a negative one. The topics are sorted by the correlation with the US EPU, where the correlations range from 0.71 to -0.03.

Second, the *Fear* measure captures a type of uncertainty that is common to all the alternative measures. The topic-based measures do not seem to capture much heterogeneity between the alternative ones, but there are some exceptions: Finance-related measures such as *Funding*, *Banking*, and *Stock market* have a relatively high correlation with the US VIX of 0.47, 0.51, and 0.51 respectively. Political measures such as *Politics* and *Elections*, on the other hand, capture more Norway-specific events and have a relatively high correlation with the Norwegian EPU of 0.38 and 0.52 respectively. The *Elections* measure has a high correlation with all the EPU measures. The *USA* uncertainty measure has a high correlation with the US EPU of 0.52, while the *UK* measure has a high correlation with both the UK EPU and the EU EPU of 0.59 and 0.65 respectively.

Third, given that the topic-based measures capture relevant types of uncertainty, the RSMV measure does not look like a good measure for uncertainty: the average correlation between the topic measures and the RSMV is 0.15. Given that no options-based volatility measure exists for Norway, a measure such as *Stock market* uncertainty can be a good alternative as a proxy for a Norwegian VIX.

The topic-based uncertainty measures do capture the type of events we expected, and different measures capture category-specific events. Most measures are positively correlated with the alternative measures of uncertainty, which suggests that there are some common components captured across uncertainty measures. This motivates an analysis of the underlying components of uncertainty.

4 Uncertainty and its effect on the economy

In times of economic distress, the uncertainty count in most types of news tends to increase. We saw in the previous sections that during the Global Financial Crisis, uncertainty increased in most of the topic-based measures. Moreover, the measures are not orthogonal, and they have a between-topic correlation varying from -0.32 to 0.87.[18]

While many of the measures are correlated, Figures 4 and 5 show that the different uncertainty measures capture different historical events confirming the narrative realism of the measures. I investigate whether the different uncertainty measures can be used to capture structural uncertainty shocks that have different implications for the overall

[18] A heat map of all the between-topic correlations is given in Figure 9 in Appendix B. The correlations are calculated at a quarterly frequency. At a daily frequency they vary between 0.03 and 0.56.

economic developments. Now I discuss possible mechanisms through which uncertainty can have different macroeconomic effects and provide five potential channels through which uncertainty affects the economy, three with a negative effect and two with a positive effect.

First, there is a large theoretical literature related to what is called "real options", see e.g., Bernanke (1983), McDonald and Siegel (1986) and Dixit and Pindyck (1994). Higher uncertainty can make firms hold back on investments and hirings that are costly to reverse, and can therefore be postponed for later. The same mechanism can play a role for households when they choose whether to make a durable goods purchase under income uncertainty. Given that they can delay the purchase, it might be their preferred action to wait and see if the uncertainty is resolved, see e.g., Eberly (1994).

Second, if agents are risk-averse, more uncertainty makes agents demand a higher risk premium to invest their money. This raises borrowing costs, and curbs growth in the economy, see e.g., Gilchrist et al. (2014).

A third potential negative channel is precautionary savings, meaning that households become more cautious in uncertain times and increase their savings. In a closed economy, these increased savings lead to increased investments, which is positive for growth. However, in a small open economy some of the savings will be made abroad, and precautionary savings can reduce economic activity, see e.g., Fernandez-Villaverde et al. (2011).

A first potential channel through which uncertainty can have a positive effect on the economy is the theory of "growth options". This refers to a mechanism in which uncertainty can encourage investment because the upside when the uncertainty is resolved can be high while there is a limited downside. One common example is the dot-com bubble in the late 1990s when many IT companies' stock boomed. These are often companies that require very limited physical capital.

A second channel that can lead to a positive effect of increased uncertainty is referred to as the "Oi-Hartman-Abel" effect, originating from Oi (1961), Hartman (1972) and Abel (1983). The mechanism here relies on firms being able to easily expand or contract when they are hit by positive or negative news. In this case, an increase in uncertainty is an increase in both potential good outcomes and bad outcomes: being able to easily contract works as an insurance against bad outcomes and increased risk is looked upon positively. This mechanism makes firms investing in large capacity since it will make them able to take advantage of potential positive news, if the news is bad they will (with low effort) scale back.

Can we map some of these theoretical channels into the topic-based uncertainty measures described in Section 2.3? For the theoretical mechanisms outlined here, any type of uncertainty can potentially give both positive and negative effects on the economy. For uncertainty measures such as *Macroeconomics*, *Stock market* etc., it is reasonable to think that uncertainty makes households and firms more cautious when making spending decisions.

For the positive channels we might look at measures that have a potential upside when the uncertainty is resolved. Examples can be *Stock listings*, *Startups* and *Mergers & acquisitions*, where the uncertainty is related to an action that can have a potential positive outcome.

We have seen some channels for how uncertainty can affect the economy in both positive and negative directions. It is important to note that even if these mechanisms has an exclusively positive or negative effect, this does not need to be the case for the topic measures discussed here. It might very well be that some increases in uncertainty related to e.g., *Mergers & acquisitions* lead to a more cautious behavior by investors and therefore an isolated negative impact on investment. Whether the total effect on the economy is positive or negative is an empirical question, and is where I turn next.

In the next section I estimate the effect on the economy from two different uncertainty shocks, that is *Macroeconomics* and *Mergers and acquisitions* uncertainty. We will see that the different measures give responses that align with the different theories discussed here, showing both positive and negative effects of uncertainty.

5 The macroeconomic effect of uncertainty shocks

This section shows that shocks to different measures of uncertainty can have different effects on the economy. Motivated by the previous section, I look at two types of uncertainty that can be put in the category of good and bad types of uncertainty.

The first type is *Macroeconomics* uncertainty. This can be categorized as a typical bad type of uncertainty and the measure is clearly countercyclical. *Macroeconomics* uncertainty has a correlation coefficient of 0.2 with the growth rate of Norwegian GDP.[19]

[19] See Figure 6 for the correlations with some alternative uncertainty measures.

The second type of uncertainty is labeled *Mergers & acquisitions* and can be related to a potential positive outcome if the merger or the acquisitions end up going well. Figure 6 show that the *Macroeconomics* and *Mergers & acquisitions* measures differ substantially by how they correlate with the alternative uncertainty measures from the literature and they are therefore good candidates for uncertainty measures that can have different effects on the economy.

The standard modeling framework in the literature is to estimate the effect of uncertainty shocks using a structural VAR model, using a recursive identification scheme. The main finding from these studies is that, using various proxies, uncertainty shocks are followed by a decline in real activity, see e.g. Bloom (2009), Jurado et al. (2015), and Baker et al. (2016). However, this approach can be challenging due to the endogeneity issues between the uncertainty measures and the macroeconomic variables, and this is especially true for uncertainty measures that are countercyclical. The strong timing restrictions imposed to identify the shocks are problematic, and even more so with quarterly variables, since assuming that uncertainty does not react to economic activity within the same quarter is unreasonable. As my baseline identification strategy, I rely on some well-known events and estimate a structural VAR using narrative sign restrictions as proposed by Antolín-Díaz and Rubio-Ramírez (2018). These type of restrictions constrain the structural shocks so that they agree with the narrative realism of known historical episodes. This is a convenient framework since the narratives in different episodes can be identified by reading the actual underlying news stories. It is shown by Antolín-Díaz and Rubio-Ramírez (2018) that the narratives of a few or even one single episode can improve identification. For details on the method and estimation, I refer to Antolín-Díaz and Rubio-Ramírez (2018).

I estimate a VAR with six variables: *Macroeconomics* uncertainty, *Mergers & acquisitions* uncertainty, log asset prices (OSEBX), interest rates (the policy rate), log investment and log output.[20] The episodes behind the narrative restrictions will be discussed below, and in addition to these restrictions, I impose two regular sign restrictions: For both types of uncertainty, I assume that an uncertainty shock increases the measures of uncertainty for two subsequent quarters, making the uncertainty shocks somewhat persistent.

[20]Investment and output are real variables for mainland Norway downloaded from Statistics Norway. The OSEBX index is downloaded from Yahoo finance, and the policy rate is from Norges Bank.

5.1 Identification from historical narratives

What type of information should be used for the narrative sign restrictions? Section 2.3 shows that many of the uncertainty measures capture the same events, e.g., for the *LTCM collapse*, we see in Figure 4 that uncertainty increased related to both *Macroeconomics* and *Stock market* (as well as for many other measures). What type of uncertainty should the *LTCM collapse* shock be attributed to? There were worries that a *LTCM collapse* could lead to a global financial crisis that could eventually have real economic implications, and I assume that uncertainty related to macroeconomic conditions increased in the quarter the *LTCM collapsed*. This does not rule out that uncertainty also increased for other categories at the same time, but the event contributed to higher macroeconomic uncertainty. This will be my first narrative sign restriction:

Narrative Sign Restriction 1: *The Macroeconomics uncertainty shock must take a positive value in 1998Q3.*

The second restriction relates to the 9/11 attacks, and as for for many events, we observe an increase in many types of uncertainty. I assume that uncertainty related to the macroeconomic conditions in Norway increased for this shock and the second narrative sign restriction is given by:

Narrative Sign Restriction 2: *The Macroeconomics uncertainty shock must take a positive value in 2001Q3.*

The third restriction relates to the *Mergers & acquisitions* uncertainty. In December 2006 a merger proposal was announced for the two largest oil companies in Norway: Statoil and Norsk Hydro.[21] The companies were officially merged in October 2007 and the new company was called StatoilHydro. Later, in 2009 StatoilHydro changed its name back to Statoil, and in 2018 it changed to Equinor. Only the part of Hydro involved in oil and gas production was merged. The remaining part, which is mainly involved in aluminum production, is still named Norsk Hydro. Merging the two companies was generally looked upon positively, and the value of both stocks increased on the announcement. Given the size of the two companies and their importance for the Norwegian economy, this was a big event, but did the merger proposal also cause uncertainty related to the future of oil

[21] The merger was approved by the Norwegian Parliament in June 2007.

production in Norway? A proposition presented by the Norwegian Ministry of Petroleum and Energy stated that the merger was associated with some uncertainty and that this merger involved both uncertainty and an element of growth opportunities:[22]

...transactions of this kind involve an element of uncertainty. The success of the merger will depend in part on the company's ability to take effective advantage of growth opportunities and to achieve efficiency improvements, synergies, cost savings and other gains. A successful merger will also depend on the company's ability to unify two strong corporate cultures and expertise communities.

I would like to stress that it is difficult to assess how important this event was for the Norwegian economy, and the restriction imposed here assumes nothing about the magnitude of this event other than that uncertainty related to *Mergers & acquisitions* increased when the merger proposal was announced. The third narrative sign restriction is then given by:

Narrative Sign Restriction 3: *The Mergers & acquisitions uncertainty shock must take a positive value in 2006Q4.*

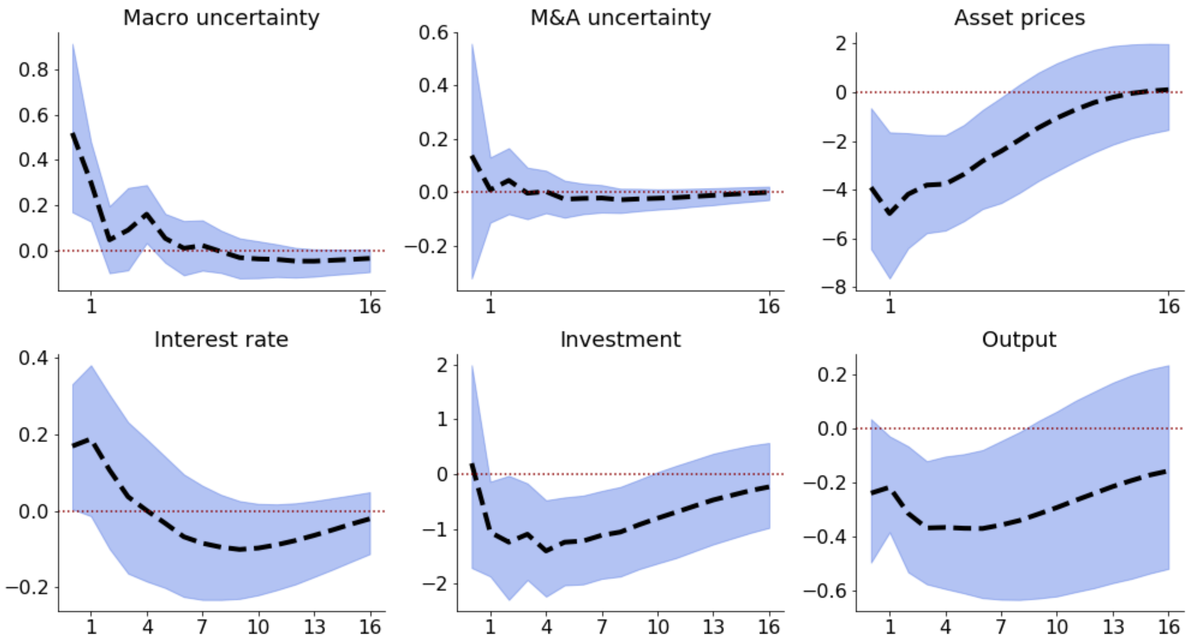
These three narrative restrictions together with the two sign restrictions are the identifying assumption in the baseline specification. Next, I report IRFs for shocks to the two types of uncertainty.

Figure 7 reports the IRFs after a macroeconomic uncertainty shock.[23] The shaded blue areas represent the 68 percent (point-wise) highest posterior density (HPD) credible sets for the IRFs, while the dashed black line represents the point-wise median IRFs. A *Macroeconomics* uncertainty shock gives a significant fall in investment by more than one percent. The response reaches a minimum after four quarters. We also observe a fall in output, with a maximum response after four quarters. *Macroeconomics* uncertainty captures a similar type of uncertainty as the literature has focused on, and the negative impulse responses after this type of uncertainty shock do resemble those in the literature, see e.g. Bloom (2009), Jurado et al. (2015), and Baker et al. (2016).

[22]Storting proposition no 60 (2006-2007) pp. 9 (`https://www.regjeringen.no/globalassets/upload/oed/pdf_filer/stprp_fusjonen_engelsk_uoffisiell_oversettelse.pdf`)

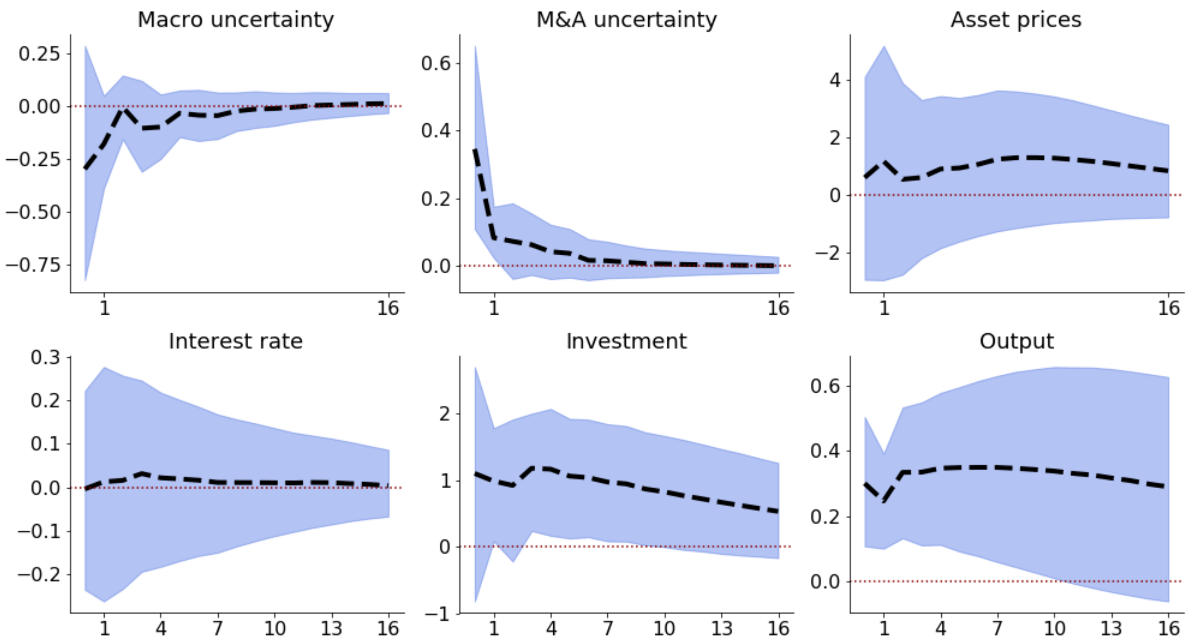
[23]Figure 10 and 11 in Appendix C reports the results with and without the narrative restrictions.

Figure 7. Impulse responses from a shock to *Macroeconomic* uncertainty



Note: The blue shaded area represents the 68 percent (point-wise) HPD credible sets for the IRFs and the dashed lines are the median IRFs.

Figure 8. Impulse responses from a shock to *M&A* uncertainty



Note: The blue shaded area represents the 68 percent (point-wise) HPD credible sets for the IRFs and the dashed lines are the median IRFs.

Figure 8 reports the IRFs after a *Mergers & acquisitions* uncertainty shock. We see a contemporaneous increase in both investment and output after a shock to this type of uncertainty. As such, this indicates that this type of uncertainty is related to a potential upside in the economy, and it increases economic activity.

I have shown here two examples out of the 80 uncertainty measures that have opposite effects on macroeconomic variables. This result indicates that both positive and negative types of uncertainty exists.

5.2 Additional results and robustness

To check the importance of the narrative restrictions for the macroeconomic uncertainty shock, I separately relax restrictions 1 and 2. Figure 12 and 13 in Appendix C shows the IRFs where only the narrative restriction 1 and 3 and thereafter 2 and 3 are imposed. The responses in this specification closely resemble those from the baseline specification, but with slightly wider confidence bands. Imposing additional restrictions can further strengthen the results (not reported).

In the baseline specification an uncertainty shock would increase the uncertainty measure for two consecutive quarters imposing that the shocks are somewhat persistent. Figure 14 and 15 in Appendix C reports the IRFs from a model with only the three narrative restrictions and no sign restrictions. The IRFs are very similar, but the positive effect on investment after a *Mergers & acquisitions* uncertainty shock is no longer significant.

The uncertainty measures used is normalized with the total number of words in the newspaper. If the coverage of a topic increase, this can potentially give an artificial increase in uncertainty.[24] To control for the attention devoted to *Mergers & acquisitions* in the newspaper, I estimate the baseline specification with the amount of *Mergers & acquisitions* news coverage as an additional variable. The IRFs from this model are reported in Figure 16 in Appendix C and are very similar to the baseline responses.

Section 4 gave three suggestions for types of uncertainty that potentially could have a positive effect on the economy. We have seen that an increase in *Mergers & acquisitions* uncertainty can give positive effects on the real economy. In Figure 17 and 18, using a recursive identification scheme, I show that this is also the case for uncertainty related to *Startups* and *Stock listings* respectively.

[24] See the discussion of Equation 4 and the alternative uncertainty measure in Section 2.4.

Lastly, I estimate a baseline model with a recursive ordering in stead of the narrative sign restrictions. A recursive identification of a VAR applies a strong structure to the contemporaneous relationship between the variables. However, it has the advantage that it demands very little supervision compared with the choices one needs to make to use narrative sign restrictions. I estimate a SVARs with a Cholesky identification scheme as a cross check for the baseline model discussed above. The ordering is as follows: asset prices, *Macroeconomic* uncertainty, *Mergers & acquisitions* uncertainty, interest rates, investment, and output, which is a similar specification as Bloom (2009) and Jurado et al. (2015). The IRFs from this model are reported in Figures 19 and 20 in Appendix C. We observe responses that are very similar to those in the baseline specification.

6 Conclusion

This paper introduces a text-based approach to creating category-specific measures of uncertainty, taking advantage of text classification tools from the machine learning literature. I classify more than 30 years of newspaper articles from Norway's largest business newspaper. The articles are classified according to their underlying theme using a topic model. I measure the degree of uncertainty conveyed by the different articles by counting the uncertainty terms within the articles. I obtain uncertainty measures related to a wide range of categories, often related to the economy, such as *Oil price*, *Monetary policy*, *Politics*, and *Stock market*.

The uncertainty measures capture well-known episodes of heightened uncertainty both at an aggregate level and at a more category-specific level. Using a SVAR I estimate the effect of two different types of uncertainty shocks on the Norwegian economy, i.e. uncertainty related to *Macroeconomics* and *Mergers & acquisitions*. Using a combination of regular sign restrictions and narrative sign restrictions, I find that these two types of uncertainty have very different effects on the Norwegian economy. A shock to *Macroeconomics* uncertainty is followed by a contraction in the Norwegian economy; this type of uncertainty resembles the bad type of uncertainty the empirical literature has focused on. In contrast to this bad type of uncertainty, a shock to *Mergers & acquisitions* leads to increased investment and output, indicating that we also can have a good type of uncertainty.

References

Abel, A. B. (1983). Optimal investment under uncertainty. *The American Economic Review 73*(1), 228.

Alexopoulos, M. and J. Cohen (2009). Uncertain times, uncertain measures. *University of Toronto Department of Economics Working Paper 352*.

Antolín-Díaz, J. and J. F. Rubio-Ramírez (2018, October). Narrative sign restrictions for svars. *American Economic Review 108*(10), 2802–29.

Arellano, C., Y. Bai, and P. Kehoe (2010). Financial markets and fluctuations in uncertainty. *Federal Reserve Bank of Minneapolis Working Paper*.

Baker, S. R., N. Bloom, and S. J. Davis (2016). Measuring economic policy uncertainty. *The Quarterly Journal of Economics 131*(4), 1593.

Basu, S. and B. Bundick (2017). Uncertainty shocks in a model of effective demand. *Econometrica 85*(3), 937–958.

Bernanke, B. S. (1983, February). Irreversibility, Uncertainty, and Cyclical Investment. *The Quarterly Journal of Economics 98*(1), 85–106.

Bjørnland, H. C. and L. A. Thorsrud (2016). Boom or gloom? examining the dutch disease in two-speed economies. *The Economic Journal 126*(598), 2219–2256.

Blei, D. M., A. Y. Ng, and M. I. Jordan (2003). Latent dirichlet allocation. *Journal of machine Learning research 3*(Jan), 993–1022.

Bloom, N. (2009, May). The Impact of Uncertainty Shocks. *Econometrica 77*(3), 623–685.

Chang, J., S. Gerrish, C. Wang, J. L. Boyd-graber, and D. M. Blei (2009). Reading tea leaves: How humans interpret topic models. In Y. Bengio, D. Schuurmans, J. Lafferty, C. Williams, and A. Culotta (Eds.), *Advances in Neural Information Processing Systems 22*, pp. 288–296. Curran Associates, Inc.

Dixit, A. K. and R. S. Pindyck (1994). *Investment under uncertainty.* Princeton university press.

Eberly, J. C. (1994). Adjustment of consumers' durables stocks: Evidence from automobile purchases. *Journal of Political Economy 102*(3), 403–436.

Fernandez-Villaverde, J., P. Guerron-Quintana, K. Kuester, and J. Rubio-Ramirez (2015, November). Fiscal volatility shocks and economic activity. *American Economic Review 105*(11), 3352–84.

Fernandez-Villaverde, J., P. Guerron-Quintana, J. F. Rubio-Ramirez, and M. Uribe (2011, October). Risk matters: The real effects of volatility shocks. *American Economic Review 101*(6), 2530–61.

Gentzkow, M., B. Kelly, and M. Taddy (2019, September). Text as data. *Journal of Economic Literature 57*(3), 535–74.

Gilchrist, S., J. W. Sim, and E. Zakrajšek (2014). Uncertainty, financial frictions, and investment dynamics. Technical report, National Bureau of Economic Research.

Griffiths, T. L. and M. Steyvers (2004). Finding scientific topics. *Proceedings of the National academy of Sciences of the United States of America 101*(Suppl 1), 5228–5235.

Gudmundsson, J. and G. J. Natvik (2012). That uncertain feeling - how consumption responds to economic uncertainty in norway. *Staff Memo, Norges Bank*.

Hamilton, J. D. (2013). Historical oil shocks. In R. E. Parker and R. M. Whaples (Eds.), *Routledge Handbook of Major Events in Economic History*, pp. 239–265. New York: Routledge Taylor and Francis Group.

Hansen, S., M. McMahon, and A. Prat (2017). Transparency and deliberation within the fomc: a computational linguistics approach. *The Quarterly Journal of Economics 133*(2), 801–870.

Hartman, R. (1972). The effects of price and cost uncertainty on investment. *Journal of Economic Theory 5*(2), 258–266.

Heinrich, G. (2005). Parameter estimation for text analysis. Technical report, Technical report.

Jurado, K., S. C. Ludvigson, and S. Ng (2015). Measuring uncertainty. *American Economic Review 105*(3), 1177–1216.

Ke, S., J. L. M. Olea, and J. Nesbit (2019). A robust machine learning algorithm for text analysis. Technical report, Working paper.

Kraft, H., E. S. Schwartz, and F. Weiss (2013, February). Growth options and firm valuation. Working Paper 18836, National Bureau of Economic Research.

Larsen, V. H. and L. A. Thorsrud (2019). The value of news for economic developments. *Journal of Econometrics 210*(1), 203–218.

Leduc, S. and Z. Liu (2015). Uncertainty shocks are aggregate demand shocks. Federal Reserve Bank of San Francisco Working Paper 2012-10.

Ludvigson, S. C., S. Ma, and S. Ng (2015). Uncertainty and business cycles: Exogenous impulse or endogenous response? Technical report, National Bureau of Economic Research.

Manela, A. and A. Moreira (2016). News implied volatility and disaster concerns. *Journal of Financial Economics*.

McDonald, R. and D. Siegel (1986). The value of waiting to invest. *The Quarterly Journal of Economics 101*(4), 707–727.

Oi, W. Y. (1961). The desirability of price instability under perfect competition. *Econometrica 29*(1), 58–64.

Segal, G., I. Shaliastovich, and A. Yaron (2015). Good and bad uncertainty: Macroeconomic and financial market implications. *Journal of Financial Economics 117*(2), 369–397.

Appendices

Appendix A Approximation of the posterior inference for LDA

The Latent Dirichlet Allocation (LDA) is estimated using the algorithm described in Griffiths and Steyvers (2004). The goal of the algorithm is to compute the posterior distribution of the hidden variables $(\mathbf{z}, \mathbf{\Theta}, \mathbf{\Phi})$ given the documents consisting of the observed words, $\mathbf{w}$. The corpus consists of M documents, each with N_m words and the index v will run through this vocabulary, which has a total size of V. Following the notation in Section 2.2, we have that $\mathbf{z} = \{\mathbf{z_k}\}_{\mathbf{k=1}}^{\mathbf{K}}$ is a vector of the K latent topics, $\mathbf{\Theta} = \{\boldsymbol{\theta}_k\}_{k=1}^{K}$ are the word mixtures and $\mathbf{\Phi} = \{\boldsymbol{\varphi_m}\}_{m=1}^{M}$ are the topic mixtures. The total probability of the model can be written as

$$p(\mathbf{w}, \mathbf{z}, \mathbf{\Theta}, \mathbf{\Phi}|\alpha, \beta) = \prod_{\mathbf{k=1}}^{\mathbf{K}} \mathbf{p}(\boldsymbol{\theta}_\mathbf{k}|\beta) \prod_{\mathbf{m=1}}^{\mathbf{M}} \left[\mathbf{p}(\boldsymbol{\varphi}_\mathbf{m}|\alpha) \prod_{\mathbf{i=1}}^{\mathbf{N_m}} \mathbf{p}(\mathbf{z_{m,i}}|\boldsymbol{\varphi}_\mathbf{m}) \mathbf{p}(\mathbf{w_{m,i}}|\theta_{\mathbf{w_{m,i}}})\right] \quad (5)$$

The method we use here is called collapsed Gibbs sampling and the goal is to approximate the distribution $p(\mathbf{z}|\mathbf{w}, \alpha, \beta)$. We start by integrating out $\mathbf{\Theta}$ and $\mathbf{\Phi}$:

$$p(\mathbf{z}, \mathbf{w}|\alpha, \beta) \quad = \quad \int_{\mathbf{\Phi}} \int_{\mathbf{\Theta}} p(\mathbf{w}, \mathbf{z}, \mathbf{\Theta}, \mathbf{\Phi}|\alpha, \beta) \mathbf{d\Phi d\Theta} \quad (6)$$

The probability $p(\mathbf{w}|\alpha, \beta)$ does not depend on $\mathbf{z}$ and the conditional distribution $p(\mathbf{z}|\mathbf{w}, \alpha, \beta)$ can be derived from $p(\mathbf{z}, \mathbf{w}|\alpha, \beta)$ directly using Gibbs simulations. Heinrich (2005) shows on page 22 that the conditional distribution $p(\mathbf{z}|\mathbf{w}, \alpha, \beta)$ can be simplified as:

$$p(z_{(m,n)} = k|\mathbf{z}_{-(m,n)}, \mathbf{w}, \alpha, \beta) \propto (n_{m,-(m,n)}^{(k)} + \alpha) \frac{n_{k,-(m,n)}^{(v)} + \beta}{\sum_{i=1}^{V} n_{k,-(m,n)}^{(i)} + \beta} \quad (7)$$

where (m, n) represents the nth word in the mth document and $-(m, n)$ denotes everything except (m, n).

With one simulated sample of the posterior distribution for $p(\mathbf{z}|\mathbf{w}, \alpha, \beta)$, the $\boldsymbol{\theta}$'s and the $\boldsymbol{\varphi}$'s can be estimated from:

$$\hat{\theta}_{k,v} = \frac{n_k^{(v)} + \beta}{\sum_{i=1}^{V} n_k^{(i)} + \beta} \tag{8}$$

$$\hat{\varphi}_{m,k} = \frac{n_m^{(k)} + \alpha}{\sum_{k=1}^{K} n_m^{(k)} + \alpha} \tag{9}$$

where $n_m^{(k)}$ denotes the number of words in the mth document that is assigned to the kth topic, and $n_k^{(v)}$ is the number of times the vth word in the vocabulary has been assigned to the kth topic.

I use the average of the estimated $\hat{\boldsymbol{\theta}}$'s and $\hat{\boldsymbol{\varphi}}$'s from the last 10 samples of the stored Gibbs simulations to approximate the word and topic mixtures used in the analysis. Before estimation, three parameters must be set. That is $(K = 80, \alpha = \frac{50}{K}, \beta = \frac{200}{V})$. It is possible to treat the priors as vectors, but I keep α and β as scalars.

Appendix B Additional results

Table 2. Estimated topics and labeling

Topic	Label	First words
Topic 0	Calendar	january, march, october, september, november, february
Topic 1	Family business	family, foundation, name, dad, son, fortune, brothers
Topic 2	Institutional investing	fund, investments, investor, return, risk, capital
Topic 3	Justice	lawyer, judge, appeal, damages, claim, supreme court
Topic 4	Surroundings	city, water, meter, man, mountain, old, outside, nature
Topic 5	Housing	housing, property, properties, apartment, square meter
Topic 6	Movies/Theater	movie, cinema, series, game, producer, prize, audience
Topic 7	Argumentation	word, besides, interesting, i.e., in fact, sure, otherwise
Topic 8	Unknown	road, top, easy, hard, lift, faith, outside, struggle,fast
Topic 9	Agriculture	industry, support, farmers, export, production, agriculture
Topic 10	Automobiles	car, model, engine, drive, volvo, ford, møller, toyota
Topic 11	USA	new york, dollar, wall street, president, usa, obama, bush
Topic 12	Banking	dnb nor, savings bank, loss, brokerage firm, kreditkassen
Topic 13	Leadership	position, chairman, ceo, president, elected, board member
Topic 14	Negotiation	solution, negotiation, agreement, alternative, part, process
Topic 15	Newspapers	newspaper, media, schibsted, dagbladet, journalist, vg
Topic 16	Health care	hospital, doctor, health, patient, treatment, medication
Topic 17	IT systems	it, system, data, defense, siem, contract, tandberg, deliver
Topic 18	Stock market	stock exchange, fell, increased, quote, stock market
Topic 19	Macroeconomics	economy, budget, low, unemployment, high, increase
Topic 20	Oil production	statoil, oil, field, gas, oil company, hydro, shelf, stavanger
Topic 21	Wage payments	income, circa, cost, earn, yearly, cover, paid, salary
Topic 22	Regions	trondheim, llc, north, stavanger, tromsø, local, municipality
Topic 23	Family	woman, child, people, young, man, parents, home, family
Topic 24	Taxation	tax, charge, revenue, proposal, remove, wealth tax, scheme
Topic 25	EU	eu, eea, commission, european, brussel, membership, no
Topic 26	Industry	hydro, forest, factory, production, elkem, industry, produce
Topic 27	Unknown	man, he, friend, smile, clock, evening, head, never, office
Topic 28	Mergers and acquisitions	orkla, storebrand, merger, bid, shareholder, acquisitions
Topic 29	UK	british, london, great britain, the, of, pound, england
Topic 30	Narrative	took, did, later, never, gave, stand, happened, him, began
Topic 31	Shipping	ship, shipping, dollar, shipowner, wilhelmsen, fleet, proud
Topic 32	Projects	project, nsb, development, fornebu, entrepreneurship
Topic 33	Oil price	dollar, oil price, barrel, oil, demand, level, opec, high
Topic 34	Sports	olympics, club, football, match, play, lillehammer, sponsor
Topic 35	Organizations	leader, create, organization, challenge, contribute, expertise
Topic 36	Drinks	wine, italy, taste, drinks, italian, fresh, fruit, beer, bottle
Topic 37	Nordic countries	swedish, sweden, danish, denmark, nordic, stockholm
Topic 38	Airline industry	sas, fly, airline, norwegian, braathens, airport, travel
Topic 39	Entitlements	municipality, public, private, sector, pension, scheme
Topic 40	Employment	cut, workplace, measures, salary, labor, working, employ
Topic 41	Politics	conservatives, party, ap, labor party, stoltenberg, frp
Topic 42	Funding	loan, competition, creditor, loss, bankruptcy, leverage
Topic 43	Literature	book, books, read, publisher, read, author, novel, wrote

Continued on next page

Table 2 – continued from previous page

Topic	Label	First words
Topic 44	Statistics	count, increase, investigate, share, average, decrease
Topic 45	Watercraft	ship, boat, harbor, strait, shipowner, on board, color
Topic 46	Results	quarter, surplus, deficit, tax, group, operating profit, third
Topic 47	TV	tv, nrk, channel, radio, digital, program, media
Topic 48	International conflicts	war, africa, irak, south, un, army, conflict, troops, attack
Topic 49	Elections	election, party, power, politics, vote, politician, support
Topic 50	Music	the, music, record, of, in, artist, and, play, cd, band, song
Topic 51	Oil service	rig, dollar, contract, option, offshore, drilling, seadrill
Topic 52	Tourism	hotel, room, travel, visit, stordalen, tourist, guest Â
Topic 53	Unknown	no, thing, think, good, always, pretty, actually, never
Topic 54	Engineering	aker, kværner, røkke, contract, shipyard, maritime
Topic 55	Fishery	fish, salmon, seafood, norway, tons, nourishment, marine
Topic 56	Europe	german, russia, germany, russian, west, east, french, france
Topic 57	Law and order	police, finance guards, aiming, illegal, investigation
Topic 58	Weekdays	week, financial, previous, friday, wednesday, tdn, monday
Topic 59	Supervision	report, information, financial supervision, enlightenment
Topic 60	Retail	shop, brand, steen, rema, reitan, as, group, ica, coop
Topic 61	Startups	bet, cooperation, establish, product, party, group
Topic 62	Food	food, restaurant, salt, nok, pepper, eat, table, waiter
Topic 63	Stock listings	shareholder, issue, investor, holding, stock exchange listing
Topic 64	Asia	china, asia, chinese, india, hong kong, south, authorities
Topic 65	Art	picture, art, exhibition, gallery, artist, museum, munch
Topic 66	Disagreement	criticism, express, asserting, fault, react, should, alleging
Topic 67	Debate	degree, debate, context, unequal, actually, analysis
Topic 68	Life	man, history, dead, him, one, live, church, words, strokes
Topic 69	Goods and services	customer, post, product, offers, service, industry, firm
Topic 70	Telecommunication	telenor, mobile, netcom, hermansen, telia, nokia, ericsson
Topic 71	IT technology	internet, net, pc, microsoft, technology, services, apple
Topic 72	Monetary policy	interest rate, central bank, euro, german, inflation, point
Topic 73	Education	school, university, student, research, professor, education
Topic 74	Regulations	rules, authorities, competition, regulations, bans
Topic 75	Trade organizations	lo, nho, members, forbund, strike, organization, payroll
Topic 76	Fear	fear, emergency, hit, severe, financial crisis, scared
Topic 77	Fiscal policy	suggestions, parliamentary, ministry, selection, minister
Topic 78	Energy	energy, emissions, statkraft, industry, environment
Topic 79	Foreign	foreign, abroad, japan, japanese, immigration, games

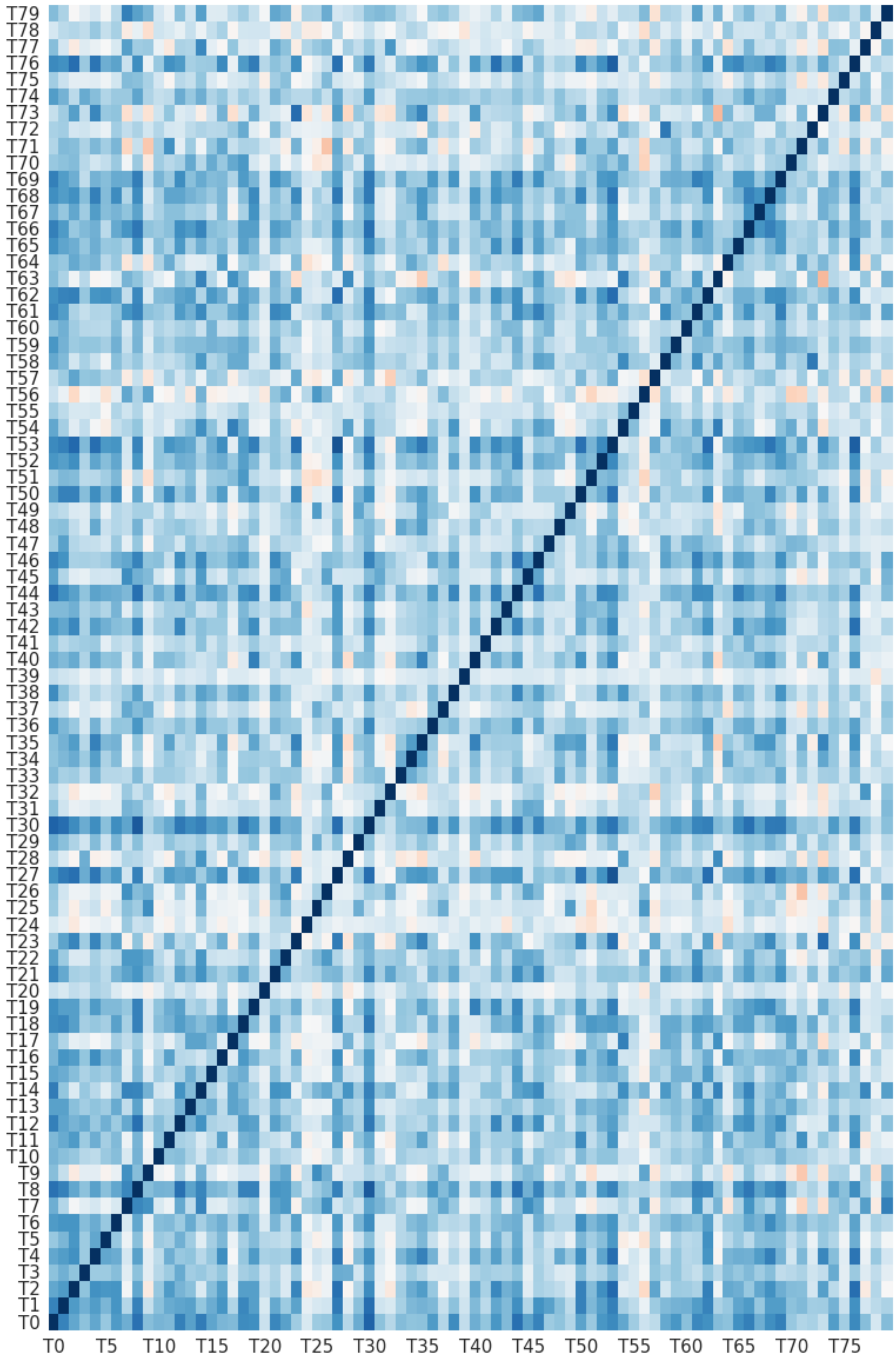
Note: The topics are labeled based on the meaning of the most important words, see the text for details. The "# of articles" column reports the number of articles, in the full sample that, according to the model, belong to that specific topic. The words are translated from Norwegian to English using Google Translate. For some words the translation from Norwegian to English creates phrases or bigrams, e.g., *central bank* in Norwegian is *sentralbank*.

Table 3. Historical events

Event	Date	Description
1990s		
GW1	1990-08-02	Gulf War 1 (2 Aug. 1990 – 28 Feb 1991)
Fixed rate (ECU)	1990-10-01	MPR: Fixed exchange rate pegged to ECU
Bank bailout	1991-12-20	The nationalization of several Norwegian banks
Talk of bank mergers	1992Q1	Talk of bank mergers related to banking crisis
Free float	1992-12-01	MPR: Free float exchange rate (leaving the ECU)
Moland	1994-01-01	New central bank governor: Moland
Stability t.w. EU	1994-05-01	MPR: Stability t.w. EU currencies
EU vote	1994-11-28	Norwegian referendum for EU membership
Storvik	1996-01-01	New central bank governor: Storvik
Failed bank merger	1997-06-26	Failed merger between Storebrand and Kreditkassen
Asian crisis	1997-11-01	The Asian financial crisis
LTCM Default	1998-09-23	Collapse of Long-Term Capital Management
Gjedrem	1999-01-01	New central bank governor: Gjedrem
2000s		
Inflation tgt.	2001-03-01	MPR: Inflation target and floating FX
Nordic Coop merger	2001-04-05	3 Scandinavian retail cooperatives merged into Coop Norden
9/11	2001-09-11	The Al-Qaeda attack on 9/11
War in Afghanistan	2001-10-07	War in Afghanistan (7 Oct 2001 – 28 Dec 2014)
GW2	2003-03-20	Gulf War 2 (20 Mar 2003 – 01 May 2003)
Statoil-Hydro merger	2006-12-18	Norway's two national oil companies are to merge to create the world's biggest offshore operator.
Credit crunch	2007-08-01	Start of the Global Financial Crisis
Lehman	2008-09-15	The collapse of Lehman Brothers
2010s		
Libyan civil war	2011-02-15	15 Feb 2011 – 23 Oct 2011
Olsen	2011-01-01	New central bank governor: Olsen
Stock market crash	2011-08-01	Stock market crash
Greek prop. refer.	2011-10-31	Greek proposed economy referendum
Egyptian coup	2013-07-03	Egyptian coup d'état
OPEC meeting	2014-11-28	OPEC chose not to reduce production
Brexit	2016-06-26	UK voted to leave the EU
Aleppo recaptured	2016-12-01	Syrian Civil War: Russian/Iranian/Turkish-backed ceasefire

Note: Details on the historical events that are indicated in the plots of the uncertainty indexes. MPR is an abbreviation for Monetary Policy Regime. Statoil-Hydro merger: https://www.ft.com/content/836f18c8-8e75-11db-a7b2-0000779e2340.

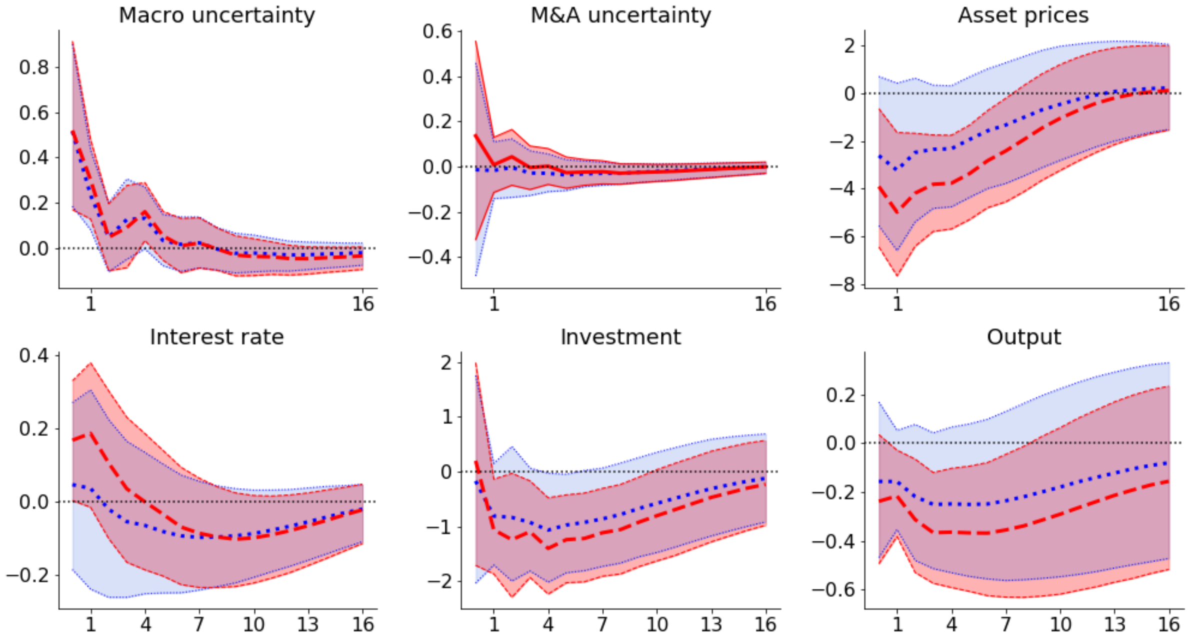
Figure 9. Correlation between the topic-based measures



Note: The correlations are computed at a quarterly frequency. See Table 2 for the corresponding topic labels. Blue represents a positive correlation while red represents a negative one.

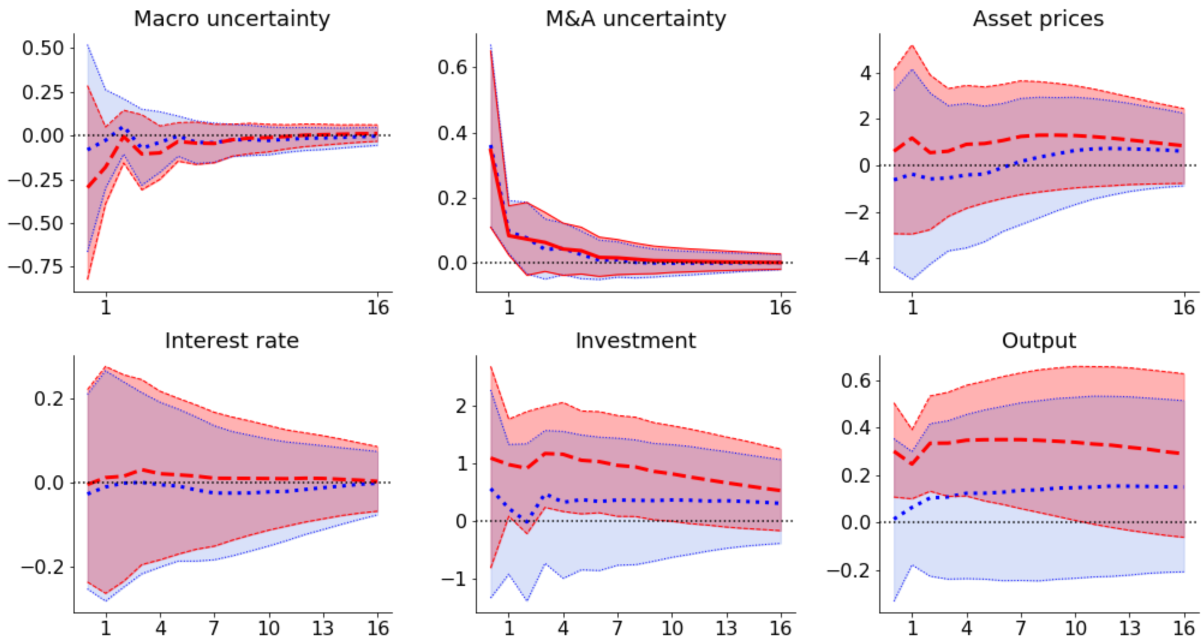
Appendix C Alternative models

Figure 10. Impulse responses from a shock to *Macroeconomic* uncertainty



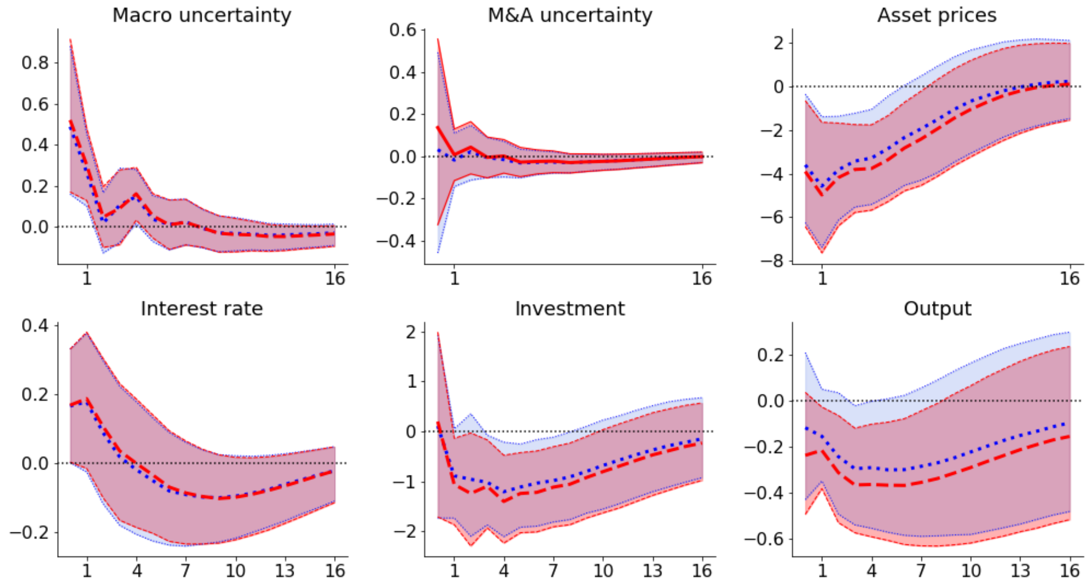
Note: The shaded areas represents the 68 percent (point-wise) HPD credible sets for the IRFs and the dashed lines are the median IRFs. The results in red is from the baseline model and results in blue are from a model where there is no narrative sign restrictions.

Figure 11. Impulse responses from a shock to *M&A* uncertainty



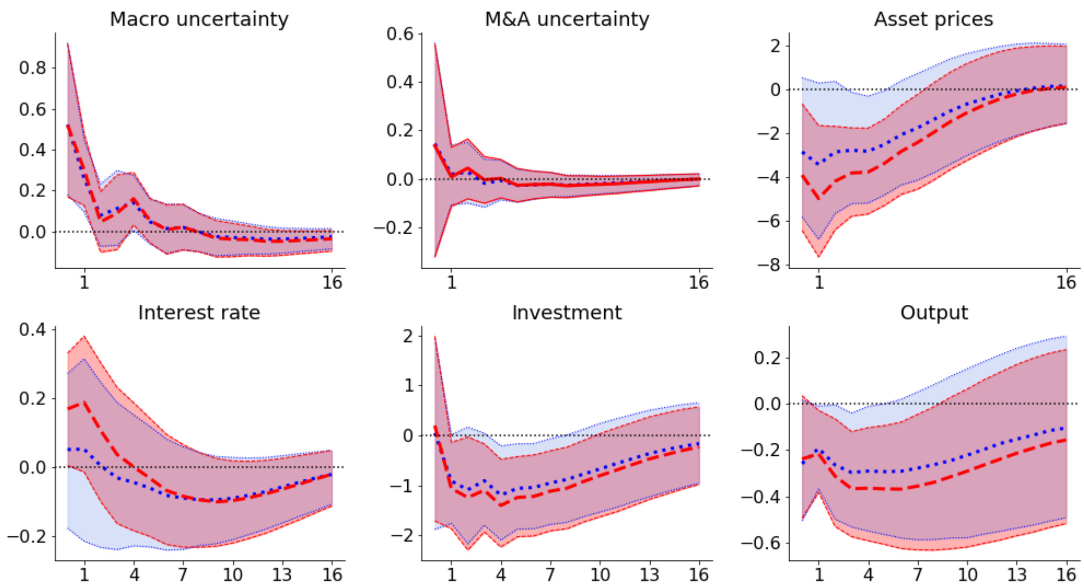
Note: See note in Figure 10.

Figure 12. Impulse responses from a shock to *Macroeconomic* uncertainty



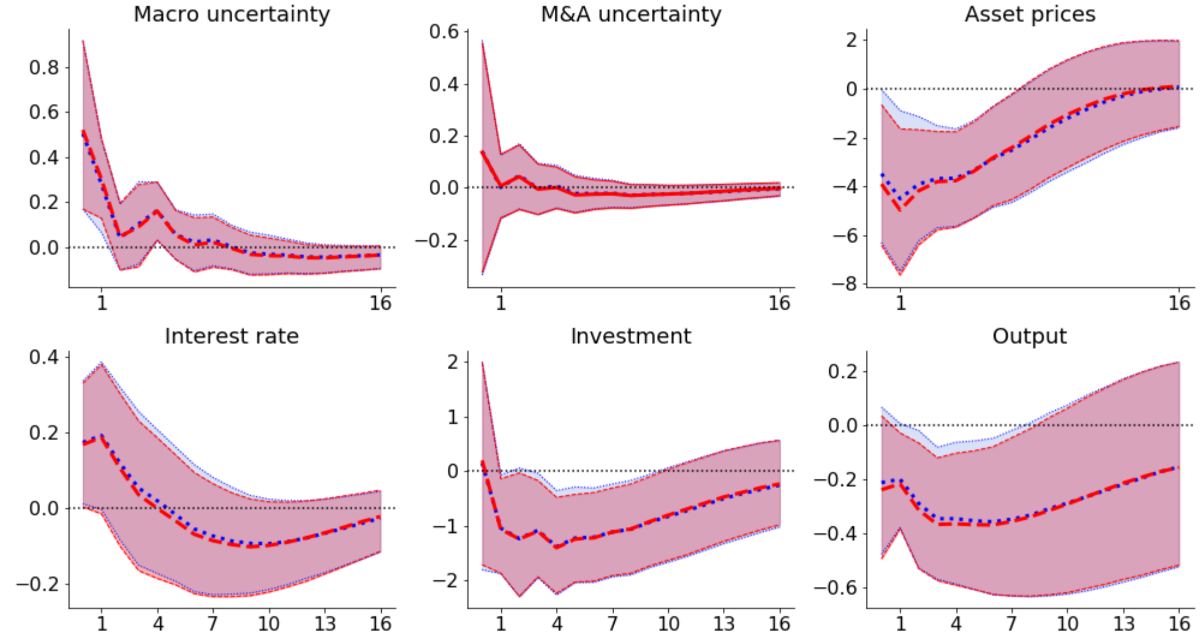
Note: The shaded areas represents the 68 percent (point-wise) HPD credible sets for the IRFs and the dashed lines are the median IRFs. The results in red are from the baseline model and results in blue are from a model where only the narrative restrictions 1 and 3 are imposed.

Figure 13. Impulse responses from a shock to *Macroeconomic* uncertainty



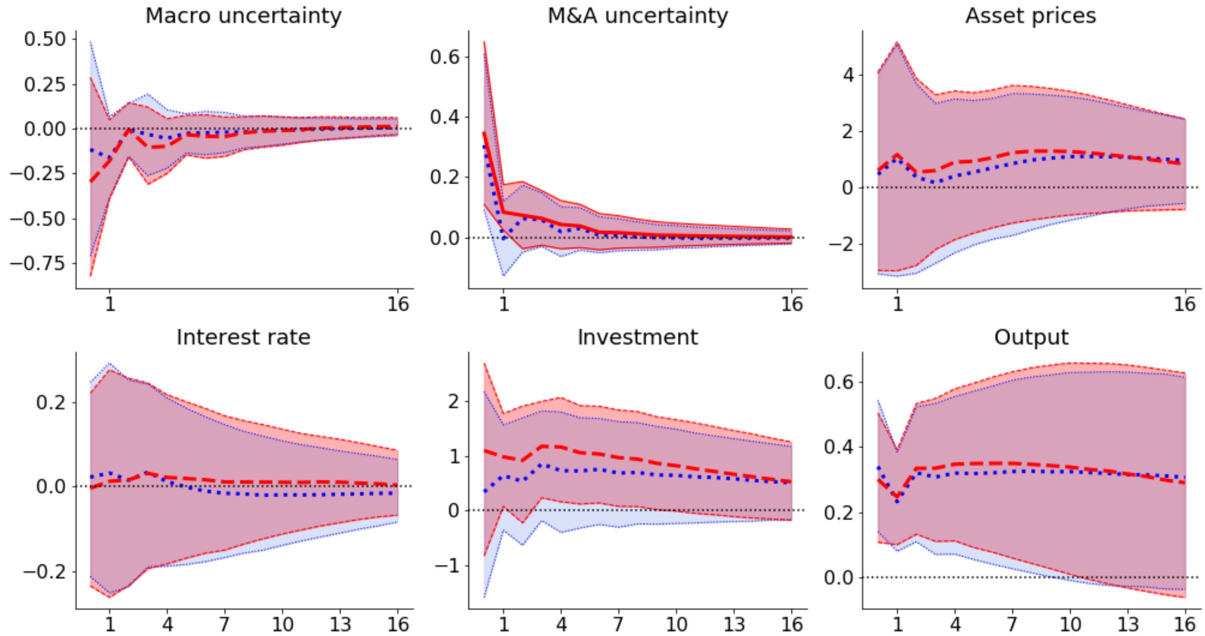
Note: The shaded areas represents the 68 percent (point-wise) HPD credible sets for the IRFs and the dashed lines are the median IRFs. The results in red are from the baseline model and results in blue are from a model where only the narrative restrictions 2 and 3 are imposed.

Figure 14. Impulse responses from a shock to *Macroeconomic* uncertainty



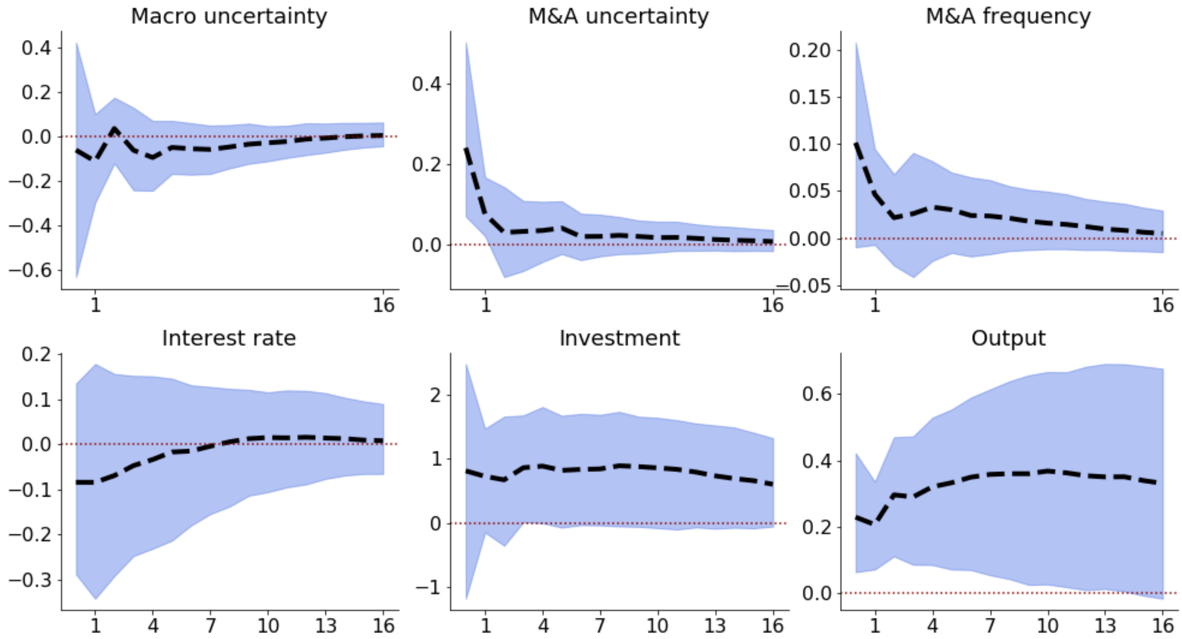
Note: The shaded areas represents the 68 percent (point-wise) HPD credible sets for the IRFs and the dashed lines are the median IRFs. The results in red is from the baseline model and results in blue are from a model where there is no restriction on the persistence of the uncertainty shock.

Figure 15. Impulse responses from a shock to *M&A* uncertainty



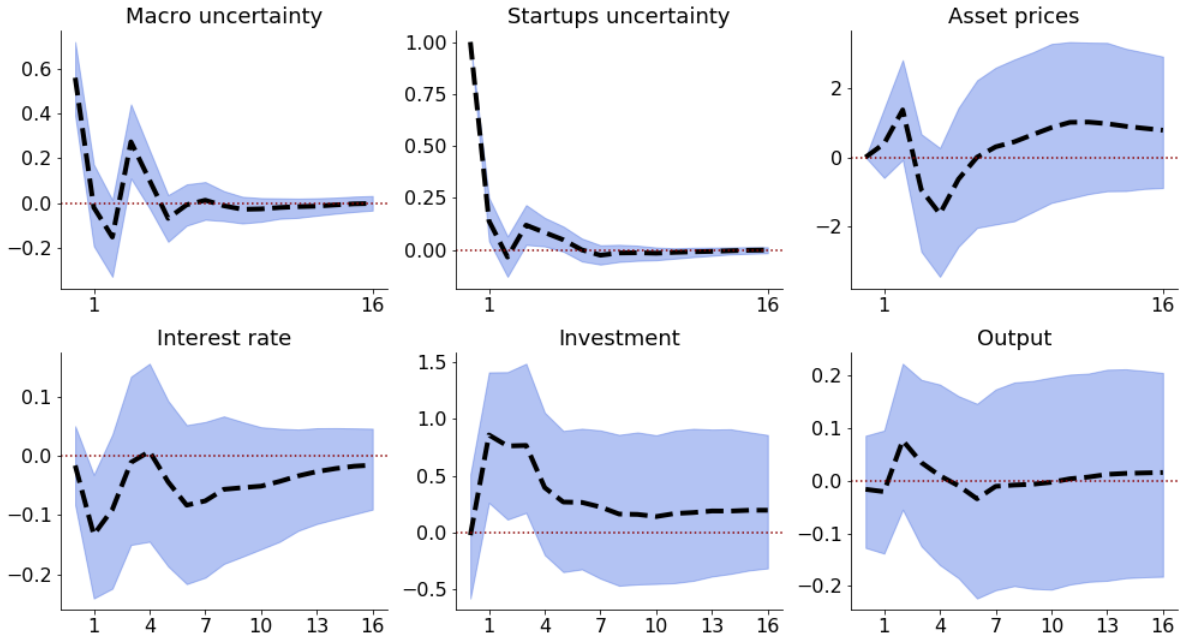
Note: See note in Figure 14.

Figure 16. Impulse responses from a shock to *M&A* uncertainty



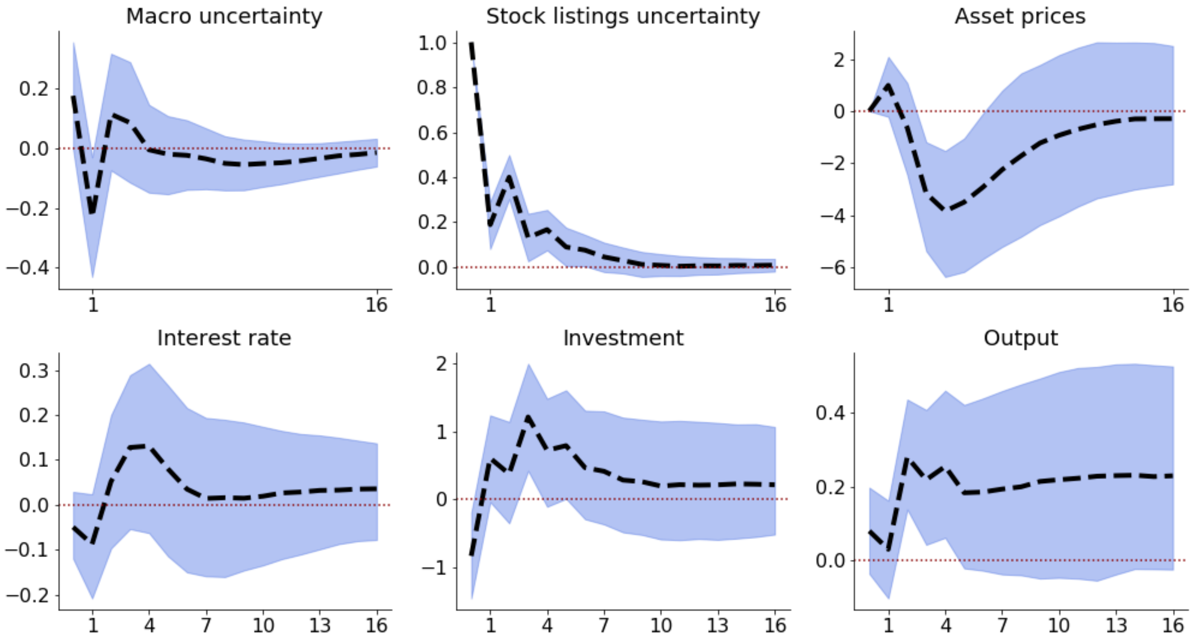
Note: Controlling for the amount of coverage of *M&A* news in the newspaper. The shock is a one standard deviation increase in the uncertainty measure. To control for the media coverage of *M&A* news I have added the frequency of this type of news in the newspaper. The model is identified using the narrative sign restrictions as explained in Section 5. The 68 percent confidence bands are plotted. The asset pricing response is not shown to conserve space.

Figure 17. Impulse responses from a shock to *Startups* uncertainty



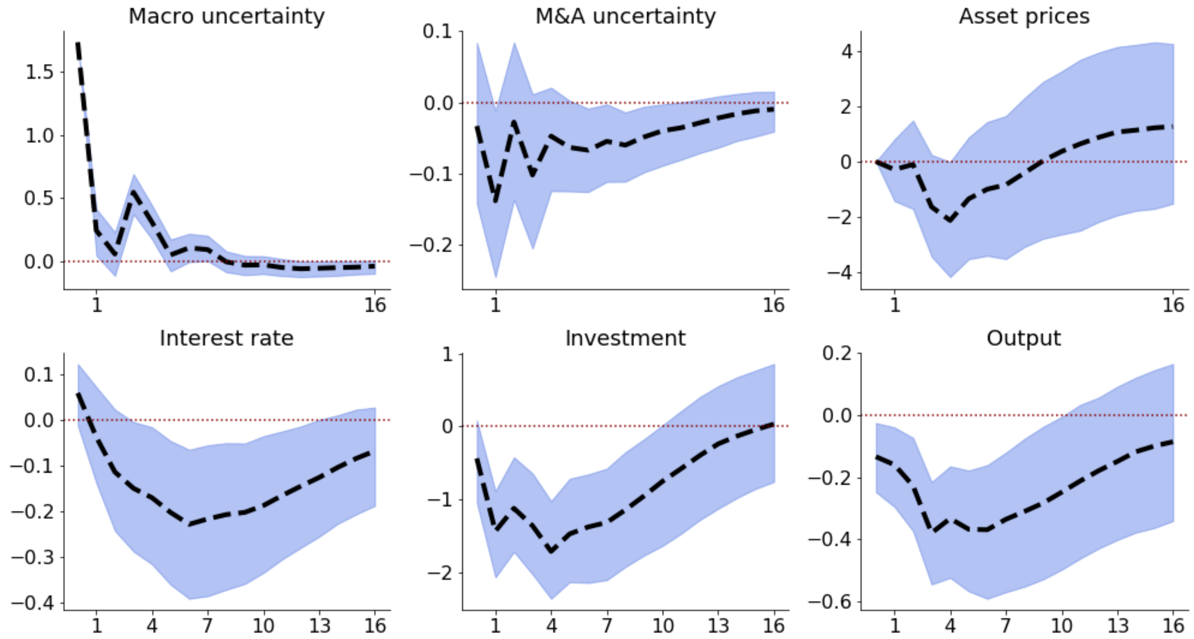
Note: The shock is a one standard deviation increase in the uncertainty measure. The model is identified using a recursive ordering and the ordering is as follows: asset prices, *Macroeconomic* uncertainty, *Startups* uncertainty, interest rates, investment, and output. The 68 percent confidence bands are plotted.

Figure 18. Impulse responses from a shock to *Stock listings* uncertainty



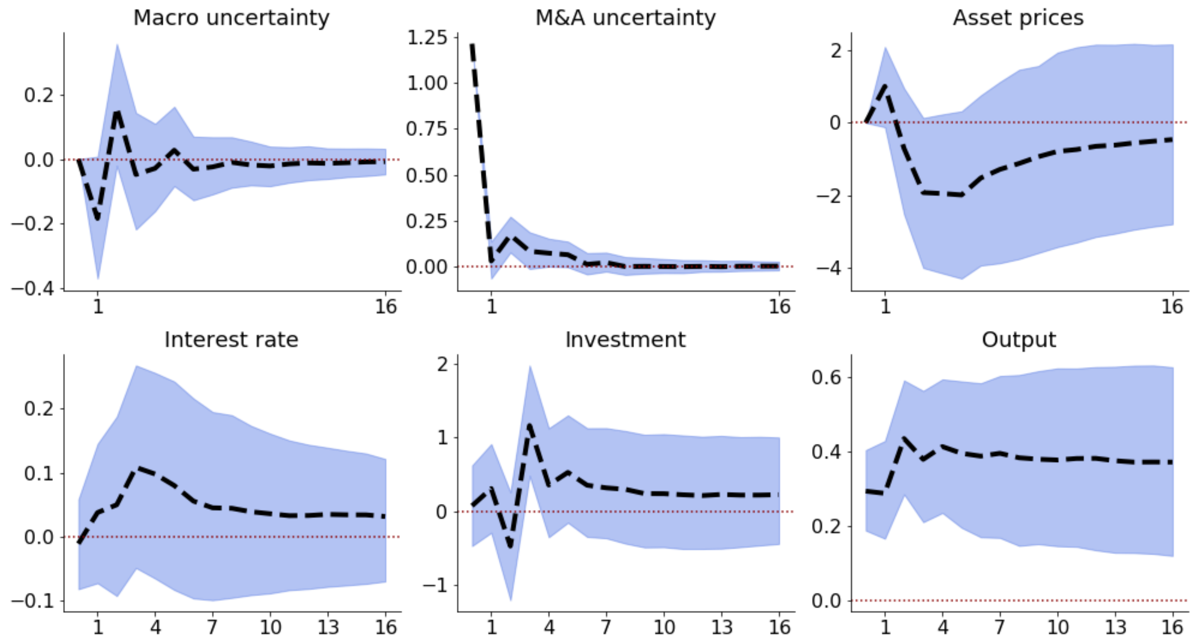
Note: The shock is a one standard deviation increase in the uncertainty measure. The model is identified using a recursive ordering and the ordering is as follows: asset prices, *Macroeconomic* uncertainty, *Stock listings* uncertainty, interest rates, investment, and output. The 68 percent confidence bands are plotted.

Figure 19. Impulse responses from a shock to *Macroeconomic* uncertainty



Note: The shock is a one standard deviation increase in the uncertainty measure. The model is identified using a recursive ordering. The ordering is as follows: asset prices, *Macroeconomic* uncertainty, *Mergers & acquisitions* uncertainty, interest rates, investment, and output. The 68 percent confidence bands are plotted.

Figure 20. Impulse responses from a shock to *M&A* uncertainty



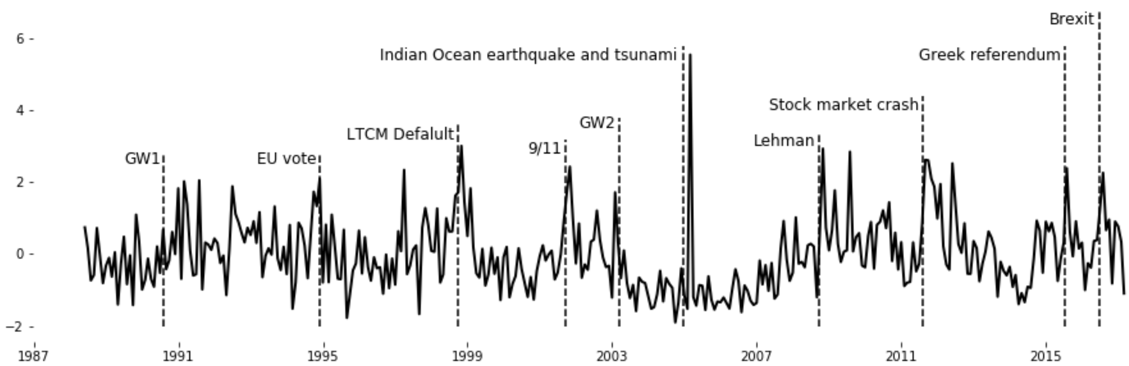
Note: See note in Figure 19.

Appendix D Norwegian Economic Policy Uncertainty

Baker et al. (2016) create an economic policy uncertainty (EPU) index for 11 countries. These indices have proven to be popular and are available through commercial data resources such as Bloomberg, FRED, and Reuters. This index is not available for Norway, and this section will follow Baker et al. (2016) and create this index based on the Norwegian newspaper DN. The series is plotted in Figure 21.

The index is created by counting the articles that contain words from three categories of words: *uncertainty* or *uncertain*; *economic* or *economy*; and Baker et al. (2016) use the following policy terms: *congress*, *deficit*, *Federal reserve*, *legislation*, *regulation* or *white house*. These three categories are named: uncertainty, economy, and policy. The counted articles contain words from all of them. Each day the final count is divided by the total number of articles that day to control for changes in total news coverage over time. The words need to be translated into their Norwegian counterparts to suit a Norwegian setting, the translations used are given in Table 4.

Figure 21. Economic Policy Uncertainty Index for Norway



Note: The index is calculated by counting articles that contain words from the *uncertainty* terms, the *economy* terms, and the *policy* terms.

Table 4. Term Sets for the Norwegian EPU index

Category	English	Norwegian
Uncertainty	*uncertainty* or *uncertain*	*usikker* or *usikkerhet*
Economy	*economic* or *economy*	*økonomisk* or *økonomi*
Policy	*government*	*regjering*
	parliament	*storting*
	authorities	*myndigheter*
	tax	*skatt*
	regulation	*regulering*
	budget	*budsjett*
	deficit	*underskudd*
	ministry of finance	*Finansdepartementet*
	central bank	*sentralbank*

Note: I also include variations of the words given in this table such as *taxation* and *regulations*.