WORKING PAPER

MICRO DATA SCIENCE

DANMARKS NATIONALBANK

24 NOVEMBER 2020 — NO. 165

Uncertainty and the real economy: Evidence from Denmark

Mikkel Bess
mbes@nationalbanken.dk
DANMARKS NATIONALBANK

Alessandro Tang-Andersen Martinello
alem@nationalbanken.dk
DANMARKS NATIONALBANK

Erik Grenestam
egr@nationalbanken.dk
DANMARKS NATIONALBANK

Jesper Pedersen
jpe@nationalbanken.dk
DANMARKS NATIONALBANK

WORKING PAPER — DANMARKS NATIONALBANK
24 NOVEMBER 2020 — NO. 165

Uncertainty and the real economy: Evidence from Denmark

Abstract

We construct an index for economic policy uncertainty in Denmark using articles in Denmark's largest financial newspaper. We adapt a Latent Dirichlet Allocation (LDA) model to sort articles into topics. We combine article-specific topic weights with the occurrence of words describing uncertainty to construct our index. We then incorporate the index in a structural VAR model of the Danish economy and find that uncertainty is a robust predictor of investments and employment. According to our model, increased uncertainty contributed significantly to the drop in investments during the Sovereign Debt Crisis. Conversely, uncertainty has so far had a smaller, but still significant, impact on investments during the COVID-19 pandemic.

Resume

Vi konstruerer et indeks for økonomisk politisk usikkerhed i Danmark ved hjælp af artikler fra Danmarks største finansielle avis. Vi tilpasser en Latent Dirichlet Allocation (LDA) model for at kategorisere artiklerne i emner. Vi kombinerer artikelspecifikke emnevægte med forekomsten af ord, som beskriver usikkerhed, for at konstruere vores indeks. Derefter indarbejder vi indekset i en strukturel VAR model for dansk økonomi og finder, at usikkerhed er en robust indikator for investeringer og beskæftigelse. Ifølge vores model bidrog øget usikkerhed betydeligt til faldet i investeringer under statsgældskrisen. Omvendt har usikkerhed indtil videre haft en mindre men stadig signifikant effekt på investeringer under COVID-19-pandemien.

Key words

Economic activity and employment; Current economic and monetary trends; Statistical method

JEL classification

D80; E66.

Acknowledgements

The authors wish to thank Børsen for kindly providing the text data used in this paper. The authors also thank colleagues from Danmarks Nationalbank for useful feedback and comments.

The authors alone are responsible for any remaining errors.

UNCERTAINTY AND THE REAL ECONOMY: EVIDENCE FROM DENMARK

WORKING PAPER

Mikkel Bess
Macroeconomic Analysis
Economics and Monetary Policy
Danmarks Nationalbank
mbes@nationalbanken.dk

Erik Grenestam
Data Analytics and Science
Financial Statistics
Danmarks Nationalbank
egr@nationalbanken.dk

Alessandro Tang-Andersen Martinello
Data Analytics and Science
Financial Statistics
Danmarks Nationalbank
alem@nationalbanken.dk

Jesper Pedersen
Macroeconomic Analysis
Economics and Monetary Policy
Danmarks Nationalbank
jpe@nationalbanken.dk

November 24, 2020

ABSTRACT

We construct an index for economic policy uncertainty in Denmark using articles in Denmark's largest financial newspaper. We adapt a Latent Dirichlet Allocation (LDA) model to sort articles into topics. We combine article-specific topic weights with the occurrence of words describing uncertainty to construct our index. We then incorporate the index in a structural VAR model of the Danish economy and find that uncertainty is a robust predictor of investments and employment. According to our model, increased uncertainty contributed significantly to the drop in investments during the Sovereign Debt Crisis. Conversely, uncertainty has so far had a smaller, but still significant, impact on investments during the COVID-19 pandemic.

Keywords Uncertainty · Natural language processing · SVAR · Time-series · Denmark

1 Introduction

Economic policy uncertainty is a growing source of concern (Ghirelli et al., 2019a; Bachmann et al., 2013a). The theoretical literature suggests that shocks to uncertainty can explain changes in real economic activity (Fernández-Villaverde et al., 2011, 2015). The empirical evidence shows that heightened uncertainty contributed to the steep economic decline and slow recovery in the wake of the global financial crisis (Fund, 2013).

This paper constructs a first measure of economic policy uncertainty specific to the Danish economy using articles published in the leading Danish business-focused newspaper (Børsen) and estimates the effect of uncertainty on economic activity. We build on the work of Baker et al. (2016), who use newspaper articles as a source of information on the level of policy-related uncertainty perceived by the general public. We augment this index by applying the methodological advances proposed by Larsen (2017), Huang et al. (2019), and Thorsrud (2018). Specifically, we expand the baseline dictionary of Baker et al. (2016) by including semantic neighbors identified in the Danish Wikipedia, and weight all articles in our corpus by the relevance of economic-related topics identified in the corpus through a Latent Dirichlet Allocation (LDA) model.

The paper consists of two parts. First, we show that while our index correlates with that of Baker et al. (2016) for the Danish economy, it is more responsive to historical events from a narrative perspective. Furthermore, by identifying topics in the article corpus, our approach allows to refine the index by focusing on specific sources of economic uncertainty (e.g. uncertainty in financial markets). We show that the main contributors to uncertainty in the Danish economy are articles covering international monetary policy, financial crises, economic growth and politics.

Second, we estimate the impact of uncertainty on the Danish economy using a structural vector auto-regressive (SVAR) model, and find that investments and employment react significantly to changes in uncertainty. If the economy is hit by an exogenous uncertainty shock at a magnitude similar to the increase in uncertainty during the Sovereign Debt Crisis (SDC), investments and employment would decrease by up to 4.5 and 0.4 per cent, respectively. We do not find that the initial drop in investment is followed by a period of increased activity, suggesting there is no pent-up demand. We examine the historical decomposition of shocks during the Sovereign Debt Crisis and the COVID-19 pandemic. We find evidence that structural shocks to uncertainty affected investments negatively during the crises, but so far to a smaller extent during the 2020 pandemic than during the SDC, even though the nominal rise in uncertainty was similar during the two crises.

Our approach measures a specific form of uncertainty. Knight et al. (1921) distinguish uncertainty from risk by defining risk as a known distribution over events and uncertainty as an unknown distribution over events. This paper focuses on the general public's perception of uncertainty. A higher level of uncertainty can be interpreted as a widening of the perceived distribution over future events. A shock to uncertainty can thus be defined as a mean preserving shock to the standard deviation of the structural shocks that drive the business cycle.[1]

We expect that changes in this type of uncertainty affect the behavior of risk-averse firms and households through specific channels. On one hand, higher uncertainty about the future state of the economy might increase household savings, and thereby decrease private consumption, through precautionary behavior. On the supply side, uncertainty about the future can cause firms to become cautious, postponing hiring and investments — they "wait and see" (Bernanke, 1983). Increased uncertainty may also dampen firm activity through financial markets by increasing bond premia and the cost of capital. This effect will induce firms to decrease investments to prevent default (Gilchrist et al., 2014; Arellano et al., 2016).

On the other hand, rather than rising uncertainty being the cause of depressed economic activity, high uncertainty might also be the consequence of depressed economic activity. In Bachmann et al. (2013a), this is referred to as the "by product" hypothesis. Periods of recessions are natural times of severed business practices, and the reestablishment may in itself generate uncertainty. While our SVAR model is silent on causality, we find evidence that uncertainty is a predictor of lower investment while controlling for other measures of economic activity, offering some support to the idea that uncertainty can be the cause of shocks to economic activity.

The rest of the paper is organized as follows: Section 2 presents our measure of economic policy uncertainty in Denmark and examines the core components of uncertainty over the last 20 years; In section 3, we set up an empirical model and assess how uncertainty affects real economic activity in Denmark; Section 4 decomposes the trends in investments during the Sovereign Debt Crisis and the COVID-19 pandemic to determine the role played by uncertainty and section 5 concludes the paper.

[1] Following the literature, we do not take a stand on whether the shock is asymmetric or not, while to truly be a risk shock and to distinguish a risk shock from a standard shock to the economy, the mean should be preserved.

2 Measuring Uncertainty

The seminal work by Baker et al. (2016) (BBD) has spurred the development of a number of approaches to measure economic policy uncertainty using news articles, with the aim of improving efficiency in terms of information extraction. Baker et al. (2016) construct their index as a normalized article headcount based on the co-occurrence of predetermined sets of words within an article.[2] Their approach is very selective, and it discards all information not directly conveyed by this arbitrary set of words. While well performing on average, this method therefore requires a large collection of articles, and does not reveal much about the sources of economic policy uncertainty.

To improve information efficiency, researchers have turned to computational language-processing methods when analyzing newspaper articles. Huang et al. (2019) expand the sets of words identifying concepts such as "uncertainty" and "fear" by selecting semantic neighbors to the original word in a high-dimensional word-vector representation. Larsen (2017) and Thorsrud (2018) use a Latent Dirichlet Allocation (LDA) model (Blei et al., 2003), a popular approach for unsupervised labeling and grouping of large collections of documents into granular topics. Bybee et al. (2020) further group these topics by similarity of their characterizing words in an embedded vector space using hierarchical clustering. They are thereby able to construct a topical, data-driven map of an article corpus from micro-level topics (e.g. mergers and acquisitions) to aggregated ones (e.g. corporate finance), which provides an intuitive representation of a corpus.

This paper combines all these approaches to construct a state-of-the-art uncertainty index for the economy. We begin by selecting words related to uncertainty. We expand this dictionary by adding semantic neighbors, and combine LDA and hierarchical clustering to construct a data-driven topic map of our article corpus. This map allows us to construct a normalized count of uncertainty-related words in articles belonging to macro topics that are most relevant for the economy (e.g., we can automatically exclude articles about sports or career advice). Alternatively, this approach allows us to focus on uncertainty related to specific hand-picked topics in which a researcher is arbitrarily more interested, such as uncertainty in the financial market. We construct indices using both strategies.

2.1 Data

Our data consists of articles published in the Danish newspaper Børsen, Denmarks largest business-focused daily publication. Our sample covers the period between January 2000 and June 2020. Before constructing our measures, we preprocess articles according to the following steps:

1. Remove articles filed under section names which suggest limited economic news content, such as sports, fashion, cooking, and travel.
2. Remove bylines.
3. Remove any HTML formatting and URL's.
4. Remove all non-alphabetical characters, including digits.
5. Remove any word longer than 25 and shorter than two characters.
6. Remove stopwords. Stopwords are a curated set of short functional words that do not convey any subject matter, such as *the*, *which* and *at*.
7. Lemmatize words: For example, *walking* is transformed into *walk*.
8. Remove all articles shorter than 50 words.
9. Add bigrams, i.e. two words separated by whitespace that commonly co-occur (e.g. *White House*), to the vocabulary as a unique entity (e.g. *White_House*).

Our final sample consists of 1,233,560 articles, with 53,160 unique terms recorded in the dictionary.[3]

2.2 Constructing the Index

We use a dictionary-based approach to measure the uncertainty expressed by an article (Baker et al., 2016). Specifically, we calculate the share of words in each article that signal uncertainty. The starting point for our collection of uncertainty-related words is our Danish translation of the words used by BBD.[4]

[2]We refer to this method as dictionary-based. BBD labels their three word sets "uncertainty", "economy", and "policy". Replications have attempted to expand the sets of words that convey economic policy uncertainty to the use of smaller corpora (Armelius et al., 2017; Arbatli et al., 2017).

[3]We do not include in the dictionary rare words appearing less than a 100 times in the full corpus.

[4]The words in Danish are "usikker", "usikkerhed", and "usikkert".

As in Huang et al. (2019), we expand this set by using a word-embedding model, i.e. a vector space representation of words, trained on the Danish Wikipedia. Similar to a thesaurus, the embedding identifies words that are used in a similar context to that of our original set. We further add some negations that appear in our articles, e.g. "not at all certain" and "far from certain". Our final set of uncertainty-related terms consists of 13 terms.[5] Similar to Larsen (2017), our baseline measure of uncertainty (U_m) for a single article consists simply in the count of uncertainty-related words normalized by the total word count of the article.

A simple aggregation of uncertainty in all our article corpus will, however, include uncertainty that is not relevant for the economy. Børsen also covers sports, and publishes career advice and book reviews. These articles should not affect a measure of economic policy uncertainty. To ensure that the index is based on relevant articles, we score each article according to how much it covers an economically relevant topic.

To score articles, we use topics identified by LDA. The LDA model finds a small set of word distributions (topics) that best represents our corpus. The starting point for fitting the LDA model is our collection of articles, which we can represent numerically as an article-term matrix $\mathcal{M}$. In this $M \times \mathcal{V}$ matrix, the rows correspond to our M articles, and each element $w_{m,v}$ is an integer representing how many times word $v \in \mathcal{V}$ occurs in article m. In this representation, a document is reduced to a discrete distribution over the words in the vocabulary $\mathcal{V}$.

The structural assumption underpinning the LDA model is that a document can be represented as a probabilistic mixture of topics, where a topic is defined as a distribution over $\mathcal{V}$. For each article m, the topic z_m follows a multinomial distribution:

$$z_m \sim \text{Multinomial}(\theta) \tag{1}$$

Given a topic z_m, the conditional probability that word w appears in an article can be written as

$$p(w|z_m, \beta) \tag{2}$$

where β is a $K \times \mathcal{V}$ where each row contains the probability distribution over words in a topic $k \in K$. The joint distribution of words and topics in an article consisting of N words can be written as

$$p(\mathbf{w}, \mathbf{z}|\beta, \theta) = \prod_{n=1}^{N} p(z_n|\theta)p(w_n|z_n, \beta) \tag{3}$$

By summing over topics and integrating over the topic distribution θ , the marginal distribution of words in an article is reduced to a function of β, and the likelihood for our corpus $\mathcal{M}$ can be written as the product of the marginal article probabilities:[6]

$$p(\mathcal{M}|\beta) = \prod_{m=1}^{M} f_m(\beta). \tag{4}$$

By estimating the model, we find the word-topic distribution matrix β that maximizes the likelihood of generating our corpus. To this end, we use the variational Bayes algorithm by Hoffman et al. (2010).

While LDA is an unsupervised model, the researcher needs to determine the appropriate number of topics (K). This choice is a model selection exercise that requires evaluation: While simply maximizing the likelihood function across different choices of K will give us the optimal $\hat{\beta}$ in a statistical sense, there is no guarantee that the resulting topics will be interpretable. As coherently labeling the topics is essential for our study, we evaluate different values for K by using a coherence measure. Our coherence measure quantifies topic quality by determining if a set of high-probability words in a given topic also co-occur in the text. Specifically, we use the C_V measure by Röder et al. (2015). Figure A2 shows coherence scores for a range of different values for the number of topics. The coherence score is maximized for $K = 90$, and consequently this is our preferred model.

[5] The final terms are: "forment", "uklar", "uklart", "usik", "usikker", "uvist", "usikkert", "ikke sikker", "ikke er sikker", "ikke helt sikker", "langt fra sikker" and "sikker havn".

[6] The prior distribution for θ is assumed to be Dirichlet(α) where $\alpha = (\alpha_1, ..., \alpha_K)$. We set $\alpha_i = \frac{1}{K} \forall \alpha_i \in \alpha$. By setting $\alpha_i < 1$, we are priming the model to find sparse topic probabilities for each document, i.e. a document is likely to cover just one or a small number of topics.

Based on the top words from each topic, we manually label each topic. The top words in the 90 topic model are presented along with our labels in table A1[7]. To visualize the relationships between topics, we use the topic-word distributions in $\hat{\beta}$ to hierarchically cluster topics into broader meta-topics similar to Bybee et al. (2020). Specifically, we use the pairwise cosine distance between word distributions as our distance metric and the clustering algorithm by Ward Jr. (1963). The result of this exercise is presented by figure A3. The clustering suggests that a fundamental grouping of our articles distinguishes texts that deal with macroeconomics and financial markets from texts about other topics. We note that topic labels that appear related to a human reader also tend to have similar word distributions. This suggests a high level of coherence between the model-learned topics and our human perception of a textual topic.

To calculate an index for economic policy-related uncertainty, we manually identify 34 topics out of the total of 90 that lie within the domain of economic policy in a broad sense.[8]. The index value for each article is calculated as the sum of topic probability across the 34 topics multiplied by the share of uncertainty words. To minimize the risk that an article unrelated to our 34 topics is included in the index, we impose a minimum topic probability sum of 0.5. This can interpreted as a requirement that an article should be more likely to cover one or more of our 34 topics than anything else. The index contribution of article m is calculated as

$$T_m = \sum_{j \in k} Pr_{LDA}(z_{jm}) \tag{5}$$

$$I_m = \mathbb{1}\left[T_m \geq 0.5\right] T_m U_m \tag{6}$$

$I_m \geq 0$ is the contribution of article m and U_m is the share of uncertainty words in the article. Finally, we calculate the aggregated uncertainty index by simply taking the mean article contribution over all articles published in a given quarter.

2.3 The Børsen Uncertainty Index

Figure 1 presents our index plotted against the VIX and our Danish version of the dictionary-based index developed by BBD. Our topic-based approach and the BBD index are both responsive around major international incidents and crises, albeit with some differences in magnitude: While our approach elicits a stronger response around the 2015 migrant crisis and Brexit, the dictionary-based index has a greater spike during the COVID-19 pandemic. Both indices are by construction more responsive to European events with respect to the VIX, which, as a function of the (implied) volatility of the S&P 500, is more responsive to events related to the U.S. and cannot capture uncertainty pertaining mainly to Denmark.

While selecting relevant topics limits the influence of economically irrelevant articles, the topic composition is somewhat arbitrary. A more data-driven approach would be to use the hierarchy of topics produced by our clustering algorithm to prune irrelevant groups of topics (see figure A3). Using the branches in a dendrogram, we create an alternative index using all topics except the topics in the branches labeled "Communication", "Miscellaneous", "Sports", "Career" and "Arts and leisure". The resulting broad index contains 61 topics as opposed to the 34 in our curated index. As shown by figure A6, the two indices are highly correlated, as the topics most responsible for driving uncertainty are included in both.

An attractive feature of our index is that we can calculate the contributions of each topic at any given point in time. We can use this property to understand which subject matter is driving uncertainty and to verify that the index is able to correctly attribute the source of historical events driving uncertainty. We can also calculate indices related to specific topics by simply limiting the calculation to a subset of the 34 topics that make up our main index. The contribution for each article-topic pair is calculated as

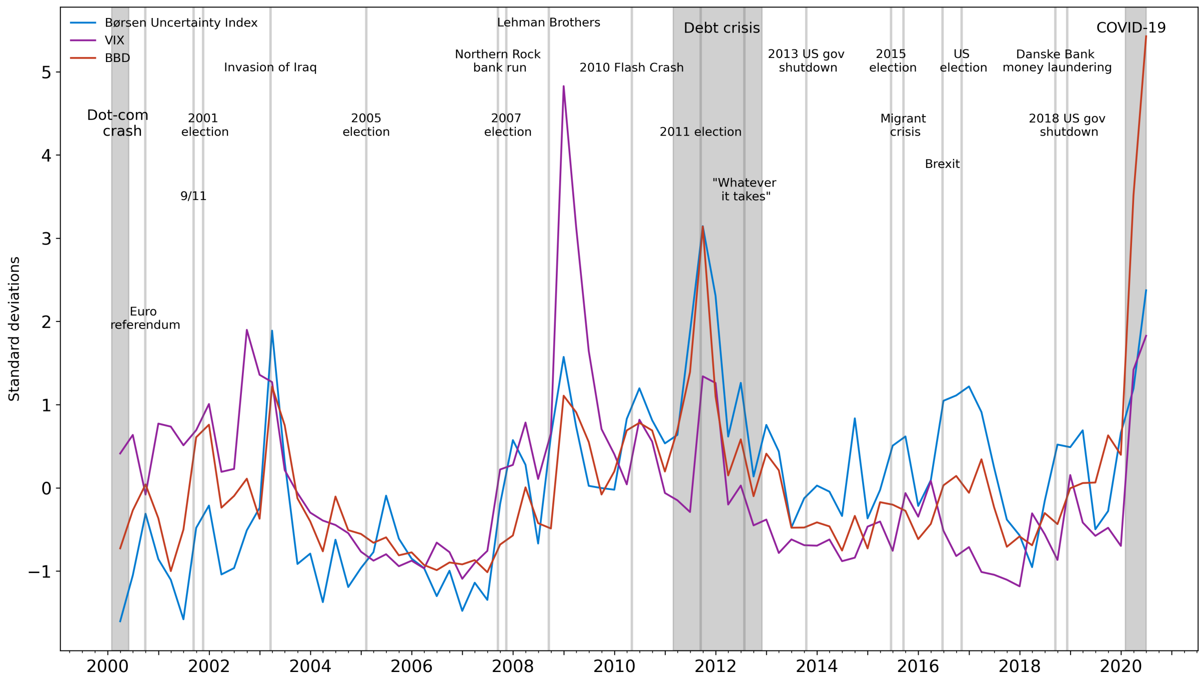
$$I_{m,k} = \mathbb{1}\left[T_m \geq 0.5\right] Pr_{LDA}(z_{km}) U_m \tag{7}$$

Next, we aggregate the index to a quarterly basis by taking the mean of $I_{m,k}$ across all articles m published during the quarter for each topic k and standardize topic contributions to sum to one for each quarter. Because some topics are more likely to contain an uncertainty word than others (e.g. the topic labeled "financial crisis" is more likely to talk

[7]Note that some of the top words are names, reflecting e.g. individual policy makers that are consistently quoted on a certain topic.

[8]Table A1 lists the 34 topics we identify manually, together with their most characteristic associated words.

Figure 1
Comparison of text-based indices with the VIX



Note: All indices are standardized to mean zero and unit standard deviation. The BBD index is calculated applying the method proposed by Baker et al. (2016) to our article corpus. The VIX index is calculated on the S&P 500 stock market. The correlation between our index, BBD, and VIX is 0.73 and 0.36, respectively.

about uncertainty than the topic labeled "infrastructure"), we have permanent level differences in the contributions of each topic to the index. Figure 2 illustrates the top three topics contributing to uncertainty in each period.[9]

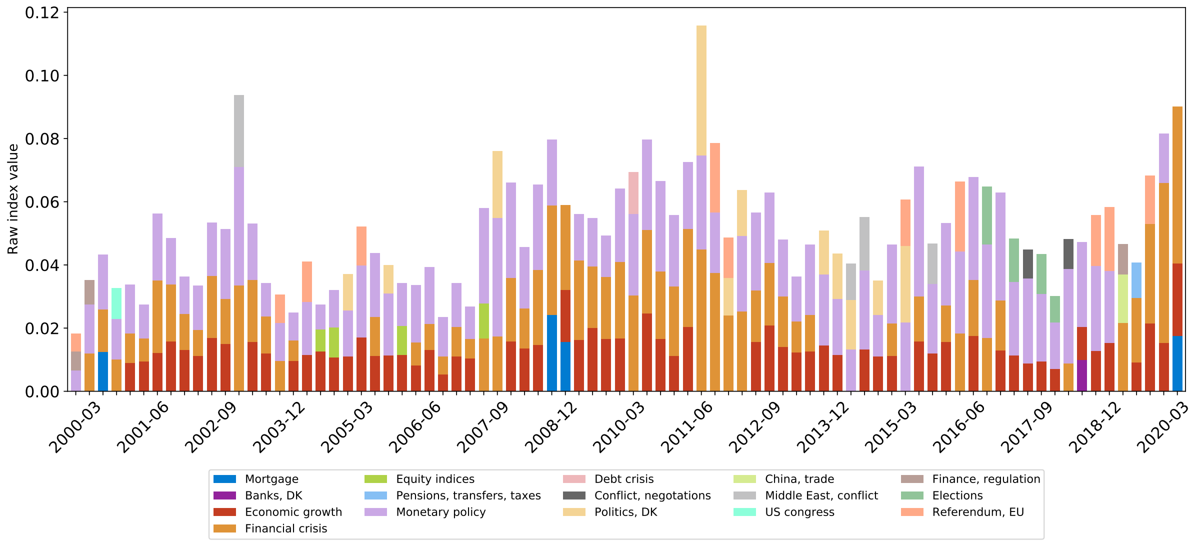
A more revealing illustration of the drivers behind uncertainty in a given quarter is to show topics that contribute abnormally to the index. To this end, we normalize each topic's contribution to mean zero and unit standard deviation. This gives us a measure of the abnormal contribution of each topic in a given quarter, i.e. how much a given topic contributed relative to its mean contribution over the entire period. Figure 3 presents the standardized topic contributions. Overall, the highs for each topic correspond well to major events and crises at the time. To give a few examples, we note that the topics on Danish banks peak around the great financial crisis. Brexit is an extreme outlier in the contributions stemming from the "UK" topic and "China, trade" flares up during the recent trade war with the U.S.

Certain topics correlate well with important economic indicators. Figure A4 shows a sub-index constructed using only articles where the "mortgage" topic has a weight of 0.5 or above. We note that the index spikes during periods of high interest rate volatility (dot-com bubble and the great financial crisis going into the Sovereign Debt Crisis). The index also spikes during the COVID-19 pandemic. Initially, there was a fear of a house price crash as unemployment rose and investors fled mortgage bonds for safer government bonds. This scenario has not materialized. This suggests that the index is able to capture uncertainty related to possible future events.

Our index tends to spike during economic crises, suggesting that our index may simply capture negative sentiment and expectations rather than uncertainty. While our index would ideally only capture changes to second order moments, confidence and uncertainty are historically closely related. BBD finds that his uncertainty index complements measures of confidence, although there is some overlap. The correlation between our index and a Danish indicator for consumer confidence is -0.34, suggesting that it also captures separate information.

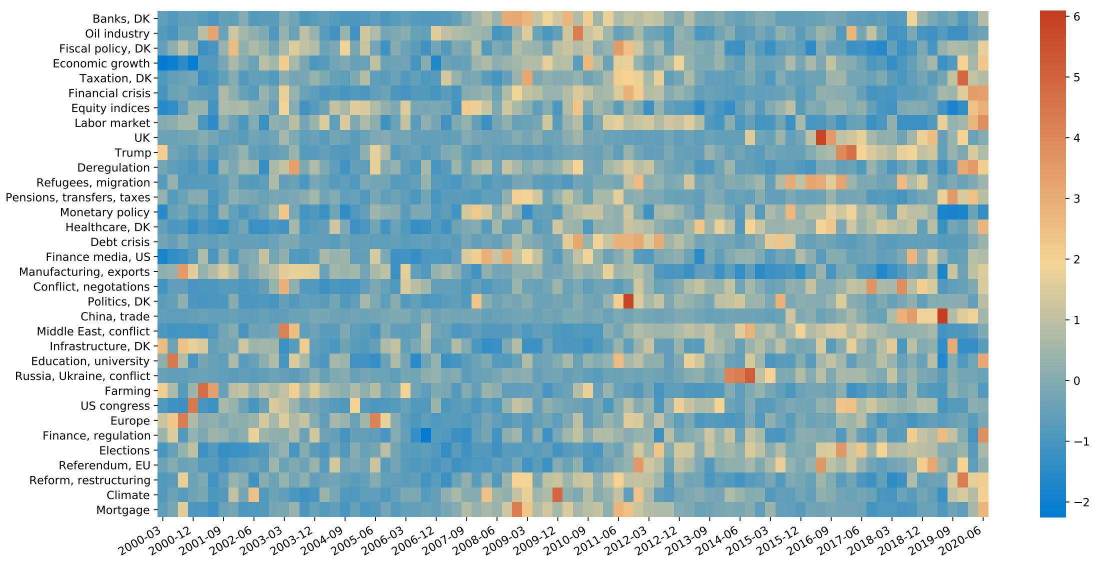
[9] The main contributing topics are "Financial crisis", "Monetary policy", and "Economic growth". See figure A5 for the raw index contributions.

Figure 2
Top three topics contributing to uncertainty in each period.



Note: The graph decomposes the total uncertainty index shown in figure 1 into contributions from each topic. Monetary policy, politics, and financial markets tend to be the biggest contributors to uncertainty in any given period. This large contribution is partially due to the fact that these three topics are some of the largest in our corpus.

Figure 3
Abnormal contributions from each of the topics in our index.



Note: The figure shows the row standardized contributions from each topic identified through the LDA model to our index. Large contributions indicate that the topic is a distinct driver of during the given period, e.g the "UK" topic during Brexit.

News reflecting increased uncertainty could be skewed towards perceived increases in downside risk. However, a bias towards negative coverage may simply reflect the demand from readers. Several studies have found that concepts like loss aversion, i.e. that investors and consumers are more concerned with downside risk, well explain phenomena in, e.g., financial markets (Haigh and List, 2005; Barberis and Huang, 2001). As our measure of uncertainty is constructed from the same news that is consumed by consumers and investors, an asymmetric responsiveness to downside risk may simply reflect the type of uncertainty that matters for the decisions of economic agents, which is what we ultimately care about.

3 Impact of Uncertainty on the Danish Economy

Having derived a measure of uncertainty we proceed to analyze how uncertainty affects the Danish economy. The analysis is focused on business investments and employment, as we expect these variables to be particularly sensitive to uncertainty. To address to what extent uncertainty affects real activity, we follow, e.g., Baker et al. (2016), Bachmann et al. (2013b) and, Ghirelli et al. (2019b), and build a structural vector auto-regressive (SVAR) model for the Danish economy.

Drawing causal inferences from VAR models is challenging because changes in uncertainty can respond to current economic conditions and anticipated future economic developments. However, SVAR models are useful for characterizing dynamic relationships with a minimal set of assumptions. Despite not being able to draw causal conclusions, an identified SVAR allows us to identify the structural shocks in the economy, and thus allows us to investigate whether changes in uncertainty implies weaker real economic activity conditional on other standard macroeconomic variables.

Our baseline model is a standard VAR model with six endogenous variables: Foreign trade-weighed GDP (F_t); the uncertainty index presented previously (U_t); business investments (I_t); the yearly log change in employment (Q_t); the yearly log inflation rate extracted from the EU-harmonized consumer price index (IN_t); and the money market interest rate (R_t). Variables are in constant prices and seasonally adjusted whenever relevant. All variables, except the uncertainty index and the interest rate, are log-transformed. The structure of the VAR model is

$$Y_t = \alpha + \Gamma T_t + \sum_{i=1}^{p} \beta_i Y_{t-i} + u_t, \tag{8}$$

where $Y_t = [F_t, U_t, I_t, Q_t, IN_t, R_t]'$ is the vector of endogenous variables.

The baseline specification includes a constant, α, and a linear trend, T_t, to take care of trending stationary variables. $u_t = [f_t, u_t, i_t, q_t, in_t, r_t]'$ is the vector of reduced-form residuals with variance-covariance matrix, $E_t[u_t u_t']$. We assume that trade-weighted foreign GDP, F_t, is exogenous with respect to the domestic variables, including uncertainty. This assumption captures the fact that Denmark is a small open economy where shocks to the global economy affect Danish variables, but shocks to Danish variables do not affect the global economy. The assumption is implemented by restricting the parameters relating domestic variables to developments in FEU_t to zero in the β matrix. This implies that FEU_t is estimated as a univariate AR(p) process with only its own lags.

Using the marginal likelihood ratio and Jeffrey's guideline (see e.g. Dieppe et al. (2018)) with no prior beliefs, we choose to estimate the baseline model with $p = 1$ lag. The DIC information criteria also indicates a preferred lag length of $p = 1$, see appendix table A7.

As it is standard in the literature, we use a Cholesky decomposition to identify structural shocks. The causal ordering follows the above presentation of the variables, starting with the foreign block, moving to the interest rate at the bottom. Let $\varepsilon_t = [e_t^f, e_t^u, e_t^i, e_t^q, e_t^{in}, e_t^r]'$ be the vector of mutually orthogonal structural innovations to the endogenous variables, the identification scheme assumes the following relation:

$$\begin{pmatrix} f_t \\ u_t \\ i_t \\ q_t \\ in_t \\ r_t \end{pmatrix} = \begin{pmatrix} a, 0, 0, 0, 0, 0 \\ b, c, 0, 0, 0, 0 \\ d, e, f, 0, 0, 0 \\ g, h, i, j, 0, 0 \\ k, l, m, n, o, 0 \\ p, q, r, s, t, u \end{pmatrix} \begin{pmatrix} e_t^f \\ e_t^u \\ e_t^i \\ e_t^q \\ e_t^{in} \\ e_t^r \end{pmatrix} \tag{9}$$

The choice of letting uncertainty enter the causal ordering first in the domestic block follows the dominant view in the literature (see e.g. Baker et al. (2016), Bachmann et al. (2013b) or Ghirelli et al. (2019b)), and reflects the fact that uncertainty affects real variables instantaneously: Uncertainty is determined before real variables in the causal ordering.

Nonetheless, Rice (2020) puts uncertainty in the end of the causal ordering, reflecting that changes in real variables affect uncertainty instantaneously instead. We show that our results are robust towards changes in the causal ordering.

The model is estimated on quarterly data using Bayesian techniques with independent Normal-Wishart priors with a univariate AR coefficient. The Minnesota prior is often used in these types of exercises. However, the Minnesota prior suffers from the drawback of assuming that the residual covariance matrix is known (Dieppe et al., 2018). Instead, we use the independent Normal-Wishart prior, thereby relaxing this assumption. Giannone et al. (2012) argue that the most natural way of choosing the hyperparameters of a model is based on their posterior distribution. This posterior distribution is proportional to the product of the hyperprior and the marginal likelihood (ML), which is the likelihood of the observed data as a function of the hyperparameters, and can be obtained by integrating out the model's coefficients. We therefore choose hyperparameters in order to maximize the marginal likelihood of the posterior model. In appendix tables A3 and A4 we report the preferred prior specification together with numerous other marginal likelihoods for different hyperparameters.

The outbreak of the COVID-19 pandemic resulted in changes in the macroeconomic variables of a magnitude never seen before in 2020:Q2. At the time of writing, it is too early to tell whether these extremely large shocks will permanently affect the economic relations. However, for inference purposes the extreme 2020:Q2 observation could affect the parameter estimates and bias the analysis. Lenza and Primiceri (2020) propose to include a parameter capturing changes in the volatility in the data. We choose a more simple approach and simply disregard the 2020:Q2 observation in the estimation sample. In the later shock decomposition of the COVID-19 pandemic, we fix the estimated parameter values to the ones estimated over the sample 2000:Q1 to 2020:Q1, and thus run the shock decomposition for the subsequent quarters using these parameters.

3.1 Results

Figure 3 shows the response of investments and employment to an increase in uncertainty of one standard deviation. To ease the interpretation of the magnitude of a one standard deviation shock, we note that uncertainty increased by 2.5 standard deviations during the SDC.

As expected, a positive shock to uncertainty reduces investment levels. Looking at the left panel, an exogenous shock to uncertainty reduces investments significantly for 14 periods, with a decrease of 1.9 per cent at its maximum after five quarters. Scaled to match the rise in uncertainty during the SDC, investments would decrease by 4.7 per cent. We find that the magnitude of the fall in investments is larger than that found in similar investigations. Ghirelli et al. (2019b) find in a SVAR model on Spanish data that investments decrease by 0.9 per cent as a response to an uncertainty shock of the same magnitude. Looking at increased uncertainty for small open economies, and in particular Chile, Cerda et al. (2016) find that a positive uncertainty shock might decrease investments by 2-3 per cent.

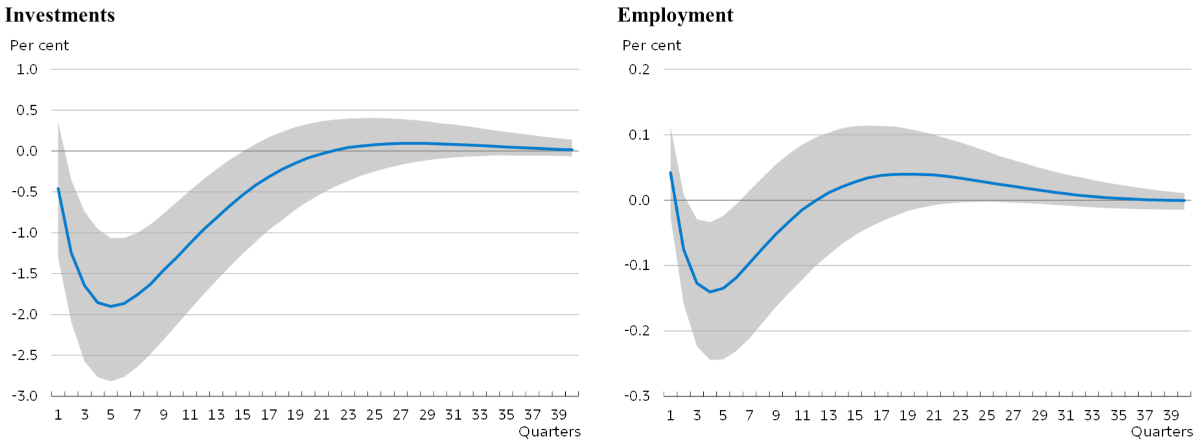
The "wait and see" hypothesis states that the initial decrease in investments is due to precautionary behavior, where firms hold back investments until the outlook is more certain. The initial decrease is then followed by a boom due to postponed investments. We are not able to conclude whether it is precautionary behavior which reduces investments in the first place, but the results do not support the idea of pent-up investment demand. Instead the investment level returns to its original level and is never significantly positive.

Looking at employment, we would expect that a rise in uncertainty would have a direct negative effect on employment level because firms might postpone hiring in uncertain times. Also, a indirect negative effect on aggregate demand due to precautionary behavior might depress the employment level as well. The right panel supports the hypothesis that the employment level responds negatively to a rise in uncertainty. The employment level decreases significantly by 0.15 per cent at its maximum.

The magnitude is similar to that found in the literature. Cerda et al. (2016) find that employment decreases by 0.2-0.7 per cent in response to uncertainty. Using data for Northern Ireland Rice (2020) finds that employment might decrease as much as 0.6-1 per cent due to heightened uncertainty. However, Nyman et al. (2018) find modest effects on employment. Nyman et al. (2018) find that a shock similar to the one for the Sovereign Debt Crisis reduces employment by 0.2 per cent. Baker et al. (2016) find that a similar chock would reduce US employment by 0.35 per cent. Scaling our results to match the Sovereign Debt Crisis, we find that employment would decrease by 0.375 per cent.

After the initial decrease, the median employment response increases above zero, but not significantly. Hence, we cannot reject the null hypothesis that firms do not postpone hiring until the uncertainty shock subsides. The median response crosses zero after 14 quarters which is remarkably less than investments which reach zero after 22 quarters. This indicates that investments react more to uncertainty in terms of maximum effect and that the effect on investments is more persistent compared to the effect on employment.

Figure 4
Investments and employment response to uncertainty



Note: Business investments and employment responses to a one standard deviation uncertainty shock. Grey areas mark 68 percent confidence intervals.

3.2 Robustness

Figure 5 shows that our results are robust towards a number of alternative specifications and causal orderings.

A potential concern about the baseline model is whether and to what extent the estimated impulse responses reflect bad news generally instead of uncertainty shocks. As a first robustness check, we therefore estimate the model with the Danish stock market index OMXC25 included.[10] Baker et al. (2016) argue that adding a leading stock market index mitigates this concern, given that stock markets are forward looking and stock prices incorporate many sources of information. However, adding the OMXC25 to the VAR model do not change the results from the baseline model as shown in figure 5.

The index presented in this paper is meant to capture a large range of topics related to economic policy uncertainty. The VIX index measures the volatility of the US stock market and is often used as a measure of uncertainty. However, as highlighted previously, the VIX index falls short in capturing political events such as Brexit. As a second robustness check, we therefore estimate the VAR model, including the VIX, index in order to investigate whether our uncertainty index has predictive power conditional on financial instability. We include the VIX index before the uncertainty index in the casual ordering for the same reasons as for the OMXC25. Figure 5 shows that the results from the baseline model still hold, despite the VIX index.

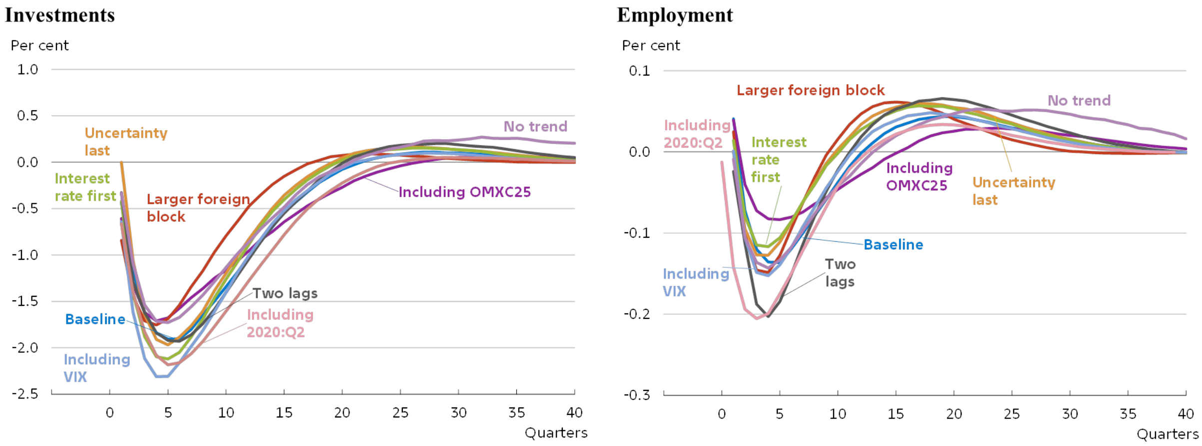
In order to keep the baseline model as simple as possible, the foreign block was created using only the developments in foreign trade-weighted GDP. Therefore, as a third robustness check we make sure that our results are robust to a more well specified foreign block. We extend the exogenous block to include a measure of relative competitiveness defined as the relative wage developments internationally and domestically (PEU_t) and oil prices OIL_t. Despite the larger foreign block, our results are robust.

The identification scheme imposes a causal ordering on the model variables. In the baseline model, we implicitly assume that rising uncertainty would affect investments and employment instantaneously. To ensure that our estimated effects are robust to the "by product" hypothesis, stating that rising uncertainty is a consequence of depressed real activity, rather than the cause, we estimate the model that allows the uncertainty index to enter the causal ordering last. Our estimated effects are robust to this change. The causal ordering is of minor importance as long as we are interested in the quantitative effects.

We perform four further robustness checks. First, we let the policy interest rate enter the causal ordering first. As Denmark adapts its monetary policy from the euro area, this change is a simple way of checking whether our results are sensitive to alternative specifications of the monetary policy transmission. Second, we change the number of lags in the

[10] Due to the forward looking behavior we include this before the uncertainty index in the causal ordering.

Figure 5
Business investments and employment responses to a one standard deviation uncertainty shock



Note: *Baseline* is the model presented in 3.1. *Interest rate first*: the money market rate is included first in the endogenous block; *Larger foreign block*: Growth on export markets, measure of competitiveness and oil prices; *Uncertainty last*: The uncertainty index is last in the Cholesky ordering; *Including OMXC25*: Adds the largest Danish stock market index before uncertainty; *Including VIX*: Adds the VIX index before uncertainty; *Two lags*: Two lags of endogenous variables; *No trend*: Drops the deterministic trend; *Including 2020:Q2*: Estimated model on sample period 2000:Q1-2020:Q2

model. Third, we disregard the trend term. Fourth, we estimate the model on a sample covering 2020:Q2. The results are robust to all these changes.

4 Does Uncertainty Drive Investments? Comparing the Sovereign Debt Crisis to the COVID-19 Pandemic

In the previous section, we found evidence that shocks to uncertainty matter for investments. However, that does not necessarily imply that uncertainty shocks have been a driving force in the historical developments in the variables of interest. This section exploits our model to examine to what extent the change in business investments during the Sovereign Debt Crisis in 2011-2012 and the COVID-19 pandemic in the first half of 2020 can be explained by uncertainty. To this end, we decompose historical movements in the endogenous variables into the contributions by deterministic variables and structural shocks.

The model in (8) has the following infinite moving average representation

$$Y_t = \alpha + \beta(L)^{-1}T_t + \Psi_0\varepsilon_t + \Psi_1\varepsilon_{t-1} + ... \tag{10}$$

where Ψ_i is the impulse response functions of the structural VAR. Each endogenous variable can then be written as

$$y_{j,t} = \underbrace{d^{(t)}}_{\substack{\text{historical contribution of}\\ \text{deterministic variables}}} + \underbrace{\sum_{i=0}^{t-1}\phi_{i,j1}\varepsilon_{1,t-i}}_{\substack{\text{historical contribution of}\\ \text{structural shock 1}}} + \cdots + \underbrace{\sum_{i=0}^{t-1}\phi_{i,jn}\varepsilon_{n,t-i}}_{\substack{\text{historical contribution of}\\ \text{structural shock n}}} \tag{11}$$

where $\phi_{i,jk}$ is element (j,k) of IRF Ψ_i. Using our estimated parameters over the sample from 2000:Q1-2020:Q1, (11) allows us to model the past contribution of structural shocks present in one variable to the development of another variable, giving us a measure of the orthogonal impact that uncertainty had on, e.g., investments.

When we estimate the parameters of the model over the sample from 2000:Q1 to 2020:Q1 and keep these estimates constant to evaluate the historical shock decomposition in 2020:Q2, we implicitly assume that the estimated economic relations in the model have not changed because of the COVID-19 pandemic. This assumption allows us to decompose the shock without worrying whether the parameter estimates are biased due to the extreme event.

The Sovereign Debt Crisis and the COVID-19 pandemic both constituted major adverse shocks to the economy, but were fundamentally different in terms of causes and policy response. On one hand, while Denmark was not at the core of the Sovereign Debt Crisis, uncertainty rose between 2011-2012, likely due to the unstable conditions in the Euro Area. On the other hand, the (still ongoing at the time of writing) pandemic has had direct consequences on the real Danish economy, partly driven by restrictions imposed by policymakers and changing demand patterns. In 2020:Q2 the level of uncertainty was almost the same as at the end of 2011.

Figure 6 displays the contribution of uncertainty to the quarterly growth rate in Danish business investments during the Sovereign Debt Crisis from 2011 to 2012 and the beginning of the COVID-19 crisis from 2020:Q1 to 2020:Q2, respectively.[11] Structural shocks to uncertainty contributed negatively to business investments during the Sovereign Debt Crisis. Uncertainty reduced the growth in investments by -1.6 and -1.7 percentage points in 2011:Q3 and Q4, respectively. The actual drop in investments in the third quarter was 1.8 percent, suggesting that uncertainty played a significant role in the fall in investments. In contrast to 2011, decreasing uncertainty actually contributed positively to the growth rate of investments in the second half of 2012. This effect could be a result of the actions undertaken by the ECB, including the *Whatever it takes*-speech by Mario Draghi held in July 2012.

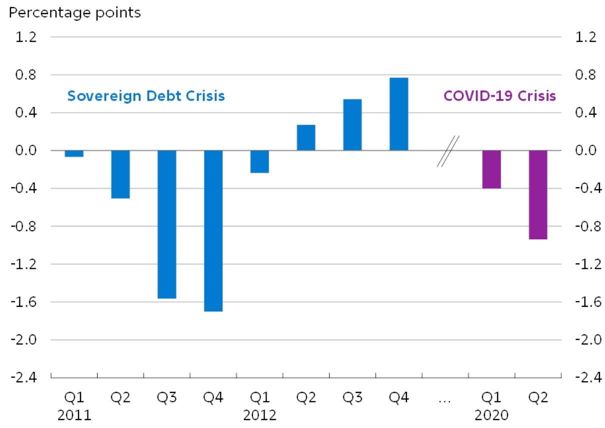
In 2020, we find that uncertainty is at almost the same level as during the peak of the SDC. However, the impact on investments is so far smaller. In the first two quarters of 2020, structural shocks to uncertainty, reduced the growth rate of investments by 0.4 and 0.9 percentage points, respectively. In reality, investments decreased by more than 10 per cent in the 2020:Q2. The marginal contribution of uncertainty suggests that uncertainty accounts for a smaller fraction of the total drop in investments when compared to the Sovereign Debt Crisis.

There might be several possible explanations for why uncertainty matters less in 2020:Q2 than in 2011. One explanation is that the full effect of the structural uncertainty shock has not yet materialized. In section 3, we found that a structural shock to uncertainty has its maximum impact on the investment level after 5 quarters. Since we are only able to address the contribution of uncertainty to investment growth within the period of the large event, it might be the case that uncertainty will continue to contribute even at a larger scale, to decreasing investments in the quarters to come.

Another explanation is that the uncertainty impact during 2011 is the effect of a longer period of positive uncertainty shocks. Figure A7 in the appendix shows the magnitude of all observed shocks to uncertainty over the sample period,

[11] Our results are robust to using the broad uncertainty index.

Figure 6
Contribution from uncertainty shocks to quarterly growth rate in investments



Note: Historical decomposition of quarterly growth rate in business investments. Sovereign debt crisis: 2011Q1 to 2012Q4. COVID-19 crisis: 2020Q1 to 2020Q2.

and shows that the economy had experienced a number of positive shocks to uncertainty, also prior to 2011. These shocks might have accumulated over time, increasing the total effect on investment growth.

Lastly, it might also be the case that the structural uncertainty shock was smaller in 2020 than in 2011 even though the uncertainty index moved by approximately the same magnitude, i.e. that shocks to other variables contributed to the changes in uncertainty. Figure A8 outlines the historical decomposition of the uncertainty index and shows that structural shocks to international and domestic variables accounted for a larger part of the increase in the uncertainty index during 2020 than 2011, i.e. other shocks contributed to the changes in uncertainty. It could therefore be argued that the sudden drop in e.g. output contributed to rising uncertainty during COVID-19.

Ludvigson et al. (forthcoming) find that different kinds of uncertainty matter differently for real activity. The Sovereign Debt Crisis sparked a fear of financial instability and the concern that a number of European countries declare bankruptcy. In contrast, during the beginning of the 2020 pandemic, firms did not face the same financial uncertainties, and uncertainty was instead related to the fiscal and labor market impact, as suggested by figure 3, while the decrease in investments was also directly affected by the shutdown. Therefore, different dynamics in the two crises might also contribute to the difference in the effects.[12]

5 Conclusion

Economic policy uncertainty has come into focus in the last decade as new methods to measure this abstract concept have become widely available. Standard economic theory suggests that a more uncertain environment might affect agents' choices due to e.g. precautionary behavior. This paper analyzes how changes in policy-related economic uncertainty affects a small open economy.

Our measure of uncertainty is constructed using an LDA model that classifies texts into a small number of topics with minimal input from the researcher. The generated topics are coherent and easy to interpret and label. When comparing our index to other measures, we find a high degree of correlation. An advantage of our method is that we can decompose the index into single-topic sub-indices. We find that these do a good job of capturing past drivers of uncertainty, which further validates our measure.

[12]While we are not able to conclude whether different kinds of uncertainty play a role in explaining the differing contributions to investment growth in the current model set-up, Ludvigson et al. (forthcoming) investigate whether rising uncertainty is an endogenous response to drops in economic activity or whether uncertainty causes drops in economic activity. They find that markedly higher macroeconomic uncertainty in recessions is often an endogenous response to output shocks, while uncertainty about financial markets is a likely source of output fluctuations. Their findings might suggest that the effect of uncertainty is bigger in crises similar to the SDC than in crises similar to the COVID-19 pandemic.

Using a structural VAR model of the Danish economy, we are able to assess whether uncertainty affects real economic activity. The methodology enables us to assess the dynamics between uncertainty and real activity, conditional on other standard financial and macroeconomic variables. Focusing on business investment and employment, we find that an exogenous shock to uncertainty of the same magnitude as during the Sovereign Debt Crisis will reduce investments and employment by up to 4.5 and 0.4 per cent, respectively. Our results are significant, and robust to alternative specifications. The magnitude of the results is in line with the literature.

Lastly, we examine crises of the past to examine how much of the variation in investment growth can be explained by uncertainty. In particular, we focus on the 2011 Sovereign Debt Crisis and the 2020 COVID-19 pandemic. We find that during the Sovereign Debt Crisis, uncertainty explains investment decisions to a larger extent than during the COVID-19 pandemic. Rising uncertainty reduced the quarterly growth rate of investments by 1.6 percentage points in 2011:Q3, which suggests that uncertainty was a key driver of reduced investment growth during the SDC. In 2020:Q2, uncertainty reduced the quarterly growth rate of investments by a mere 0.9 percentage points. Several possible explanations are outlined, one being that during the pandemic, increased uncertainty was not the main cause of depressed economic activity, but a symptom of a sudden drop in demand.

References

ARBATLI, E. C., S. J. DAVIS, A. ITO, AND N. MIAKE (2017): "Policy Uncertainty In Japan," Working Paper 23411, National Bureau of Economic Research.

ARELLANO, C., Y. BAI, AND P. J. KEHOE (2016): "Financial Frictions and Fluctuations in Volatility," Working Paper 22990, National Bureau of Economic Research.

ARMELIUS, H., I. HULL, AND H. S. KÖHLER (2017): "The timing of uncertainty shocks in a small open economy," *Economics Letters*, 155, 31–34.

BACHMANN, R., S. ELSTNER, AND E. R. SIMS (2013a): "Uncertainty and Economic Activity: Evidence from Business Survey Data," *American Economic Journal: Macroeconomics*, 5, 217–49.

——— (2013b): "Uncertainty and Economic Activity: Evidence from Business Survey Data," *American Economic Journal: Macroeconomics*, 5, 217–49.

BAKER, S. R., N. BLOOM, AND S. J. DAVIS (2016): "Measuring Economic Policy Uncertainty*," *The Quarterly Journal of Economics*, 131, 1593–1636.

BARBERIS, N. AND M. HUANG (2001): "Mental Accounting, Loss Aversion, and Individual Stock Returns," *The Journal of Finance*, 56, 1247–1292.

BERNANKE, B. S. (1983): "Irreversibility, Uncertainty, and Cyclical Investment," *The Quarterly Journal of Economics*, 98, 85–106.

BLEI, D. M., A. Y. NG, AND M. I. JORDAN (2003): "Latent Dirichlet Allocation," *J. Mach. Learn. Res.*, 3, 993–1022.

BYBEE, L., B. T. KELLY, A. MANELA, AND D. XIU (2020): "The structure of economic news," Tech. rep., National Bureau of Economic Research.

CERDA, R., A. SILVA, AND J. T. VALENTE (2016): "Economic Uncertainty Impact In a Small Open Economy: The Case of Chile," Working paper, Centro Latinoamericano de Políticas Económicas y Sociales, Pontificia Universidad Católica de Chile.

DIEPPE, A., R. LEGRAND, AND B. V. ROYE (2018): "The Bayesian Estimation, Analysis and Regression (BEAR) Toolbox," Technical guide, ECB.

FERNÁNDEZ-VILLAVERDE, J., P. GUERRÓN-QUINTANA, K. KUESTER, AND J. RUBIO-RAMÍREZ (2015): "Fiscal volatility shocks and economic activity," *American Economic Review*, 105, 3352–84.

FERNÁNDEZ-VILLAVERDE, J., P. GUERRÓN-QUINTANA, J. F. RUBIO-RAMÍREZ, AND M. URIBE (2011): "Risk matters: The real effects of volatility shocks," *American Economic Review*, 101, 2530–61.

FUND, I. M. (2013): "World Economic Outlook: Hopes, Realities, Risks," Economic outlook, IMF Press.

GHIRELLI, C., M. GIL, J. J. PÉREZ, AND A. URTASUN (2019a): "Measuring economic and economic policy uncertainty and their macroeconomic effects: the case of Spain," *Empirical Economics*.

GHIRELLI, C., J. J. PÉREZ, AND A. URTASUN (2019b): "A New Economic Policy Uncertainty Index for Spain," Working Paper 1906, Banco de Espana.

GIANNONE, D., M. LENZA, AND G. E. PRIMICERI (2012): "Prior Selection for Vector Autogregressions," Working Paper 1494, ECB.

GILCHRIST, S., J. W. SIM, AND E. ZAKRAJŠEK (2014): "Uncertainty, Financial Frictions, and Investment Dynamics," Working Paper 20038, National Bureau of Economic Research.

HAIGH, M. S. AND J. A. LIST (2005): "Do Professional Traders Exhibit Myopic Loss Aversion? An Experimental Analysis," *The Journal of Finance*, 60, 523–534.

HOFFMAN, M., F. R. BACH, AND D. M. BLEI (2010): "Online Learning for Latent Dirichlet Allocation," in *Advances in Neural Information Processing Systems 23*, ed. by J. D. Lafferty, C. K. I. Williams, J. Shawe-Taylor, R. S. Zemel, and A. Culotta, Curran Associates, Inc., 856–864.

HUANG, C., S. SIMPSON, D. ULYBINA, A. ROITMAN, ET AL. (2019): "News-based Sentiment Indicators," Tech. rep., International Monetary Fund.

KNIGHT, F. H., U. RISK, AND H. PROFIT (1921): "Schaffner & Marx," *New York*.

LARSEN, V. (2017): "Components of uncertainty," Tech. rep., Norges Bank Working Paper 5/2017.

LENZA, M. AND G. E. PRIMICERI (2020): "How to Estimate a VAR after March 2020," Working paper, ECB and NBER.

LUDVIGSON, S., S. MA, AND S. NG (forthcoming): "Uncertainty and Business Cycles: Exogenous Impulse or Endogenous Response?" *American Economic Journal: Macroeconomics*.

NYMAN, R., S. KAPADIA, D. TUCKETT, D. GREGORY, P. ORMEROD, AND R. SMITH (2018): "News and narratives in financial systems: exploiting big data for systemic risk assessment," Working Paper 704, Bank of England.

RICE, J. (2020): "Economic Policy Uncertainty in Small Open Economies, a Case Study of Ireland," Working Paper 1, Central Bank of Ireland.

RÖDER, M., A. BOTH, AND A. HINNEBURG (2015): "Exploring the Space of Topic Coherence Measures," in *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining*, New York, NY, USA: ACM, WSDM '15, 399–408.

THORSRUD, L. A. (2018): "Words are the New Numbers: A Newsy Coincident Index of the Business Cycle," *Journal of Business & Economic Statistics*, 0, 1–17.

WARD JR., J. H. (1963): "Hierarchical Grouping to Optimize an Objective Function," *Journal of the American Statistical Association*, 58, 236–244.

Appendix

This appendix contains all supplementary figures and tables referenced in the main body of the text.

Figure A1
Uncertainty terms per 1,000 words, 90-day moving average.

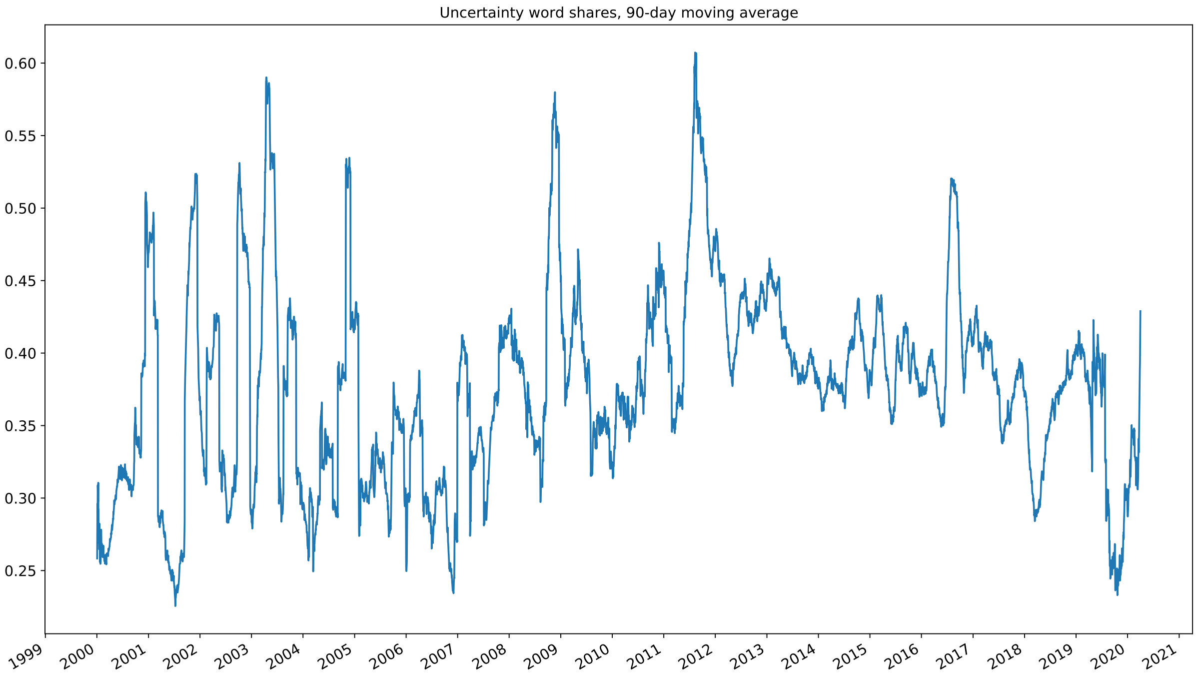


Table A1
LDA model topics

Topic #	Label	Top words
1	English words	the, of, lego, and, in, twitte, on, com, is, amerikansk, as, world, new, global, us, drone
2	**Banks, DK**	bank, banke, dansk, nordea, kunde, jysk, pengeinstitut, skriver, udlåne, finanstilsyn, finans
3	**Oil industry**	møller, mærske, møller_mærske, selskab, olie, nordsøen, maersk_oil, olieselskab, felt
4	**Fiscal policy, DK**	offentlig, regering, stat, privat, sektor, økonomisk, finansministerie, økonomi, finansminister
5	Financial markets	kur, pct, indeks, falde, aktie, stigning, ligge, stige, marked, dansk, kfor_eksempel, selskab
6	IT, software	microsoft, software, it, system, selskab, sap, damgaard, kunde, oracle, business, virksomhed
7	Wind power, Vestas	vesta, ordre, selskab, vindmølle, mw, finans, projekt, marked, ørsted, flsmidth, møller, megawatt
8	Commodity markets	per, prise, dollar, tønde, falde, stige, olieprise, olie, produktion, efterspørgsel, mio
9	Soccer, Spain	spansk, spanien, minut, barcelona, real_madrid, tivoli, mål, madrid, spanier, portugal
10	Sports, DK	vinde, nummer, første, dansker, turnering, dansk, sæt, ol, runde, spille, finale, gange, god
11	Insurance, DK	tryg, dsv, pandora, novozyme, forsikringsselskab, selskab, topdanmark, forsikring, alme_brande
12	**Economic growth**	pct, falde, vente, stige, tal, måned, økonomi, stigning, vækste, kvartal, økonom, dansk, ifølge
13	**Taxation, DK**	regering, dansk, pct, skat, virksomhed, forslag, øge, sænke, erhvervsliv, die, danmark
14	Organization	virksomhed, ny, kunde, god, bruge, udvikling, skab, strategi, samarbejde, løsning, fokus, del
15	**Financial crisis**	økonomisk, økonomi, krise, vækste, land, problem, global, politisk, situation, finansiel, risiko
16	Retail, food	butik, forbruger, kæde, kunde, produkt, vare, købe, sælge, supermarked, dansk, ny, salg, coop
17	Corporate governance	aktie, selskab, krone, mio, meddelelse, aktionær, pct, bestyrelse, samle, egne, eje, finans
18	Danish names	nielsen, direktør, jensen, bestyrelse, medlem, formand, hansen, erik, jørgen, jens, larsen
19	Cars	bil, motor, ny, model, kilometer, bmw, køre, hestekraft, tesla, vw, ford, audi, krone, volkswagen
20	Telecom operators	tdc, kunde, net, selskab, teleselskab, dansk, dankort, marked, ny, tele, mobil, prise, telia
21	**Equity indices**	pct, falde, indeks, stige, aktie, selskab, amerikansk, europæisk, stigning, dow_jone, marked
22	Technical analysis	niveau, teste, teknisk, bryde, uge, lars, positiv, falde, stige, fortsat, momentum, trend, brud
23	Newspapers, DK	avis, politiken, berlingske, jylland_post, skriver, lund, dagblad, medie, mediere, ifølge
24	Family	barn, kvinde, familie, ung, mand, årig, pige, forælder, menneske, dreng, liv, dø, gå, datter
25	Maritime, shipping	norsk, rederi, skib, norge, havn, norden, skibe, torm, sejle, oslo, flåde, havne, dansk, værft
26	Sweden, Scandinavia	danmark, dansk, svenske, sverige, land, udenlandsk, nordisk, international, norge, udland, finland
27	Soccer, DK	klub, spiller, jesper, brøndby, morten, spille, kontrakt, dansk, fodbold, olsen, sommer, trænere
28	**Labor market**	medarbejder, arbejde, ansætte, jobbe, virksomhed, arbejdsmarked, leder, folk, arbejdsplads
29	**UK**	britisk, storbritannien, london, pund, england, brite, engelsk, falck, bbc, the_guardian
30	Breweries	carlsberg, bryggeri, øl, marked, manchester_city, royal_unibrew, fifa, heinek, pct, coca_cola
31	Design, fashion	butik, design, ny, brande, poste, salg, sælge, tøj, marked, produkt, mærke, bang_olufsen, brand
32	Telecom	apple, google, telefon, mobiltelefon, ny, iphone, amazon, mobil, nokia, selskab, smartphone
33	**Trump**	trumpe, filme, film, donald_trumpe, trump, prise, dansk, god, serie, skuespiller, nominere
34	IPO	aktie, investor, selskab, børs, dansk, marked, aktiemarked, børsnotering, købe, fondsbørs
35	**Deregulation**	dk, jakob, paul, sundhed, bech, rune, portal, kåre, liberalisering, brandt, ulla, apotek
36	Royal family	rasmussen, københavn, dansk, udstilling, besøge, arrangement, båd, borgmester, både, deltage
37	**Refugees, migration**	italien, italiensk, flygtning, pol, brasilien, polsk, land, asylansøger, grænse, brasiliansk
38	Sports, cycling	køre, løb, holde, team, hold, etape, tid, rytter, sidste, kilometer, træning, gange, formel
39	**Pensions, transfers, taxes**	skatte, betale, krone, penge, konkurs, selskab, mio, udbetale, betaler, ifølge, beløbe
40	Books, magazines	bog, skrive, historie, boge, læser, journalist, læse, bøge, artikel, side, forlag, magasin
41	**Monetary policy**	rente, amerikansk, centralbank, pct, dollar, marked, torsdag, onsdag, euro, ecb, fredag, dansk
42	**Healthcare, DK**	kommune, region, borger, bruge, patient, mene, formand, kommunal, borgmester, god, land, læge
43	**Debt crisis**	grækenland, græsk, gæld, mia, obligation, euro, pct, lån, statsobligation, imf, rente, auktion
44	Analysis reports	pct, aktie, krone, falde, stige, dansk, kursmål, anbefaling, indeks, regnskab, torsdag, onsdag
45	Crime	politi, årig, mand, oplyse, københavn, person, anholde, blive, ifølge, sag, sigte, fund, finde
46	South East Asia	iss, thailand, pandora, dubai, smykke, ur, malaysia, bendten, nicklasse_bendten, hartmann
47	Marketing, adverstising	tv, dr, facebook, bruge, bruger, digital, internet, hjemmeside, nette, kanal, ny, app, radio
48	Soccer, FCK	kamp, point, mål, sæson, minut, første, score, hold, spille, holde, vinde, halvleg, sejre, spill
49	Investing	investering, afkast, investere, pct, investor, kunde, marked, risiko, aktie, selskab, penge
50	Legal, court	sag, ret, dom, advokat, afgørelse, anklager, tiltale, dømme, bøde, mene, domstol, dommer
51	M&A	selskab, købe, sælge, eje, salg, fond, opkøbe, kapitalfond, krone, mia, virksomhed, ejer, pct
52	Names	peter, steen, clausen, søndergaard, bang, arne, frede, bella_center, tk_developmene, jeppes
53	**Finance media, US**	dollar, mia, amerikansk, ifølge, skriver, selskab, bloomberg_new, aktie, per, bloomberg, finans
54	Pharma, biotech	lundbeck, behandling, patient, selskab, studie, produkt, genmabe, lægemiddel, middel, medicin
55	Entrepreneurship	virksomhed, penge, gå, gange, bruge, ny, firma, iværksætter, starte, første, folk, rigtig, tid
56	**Manufacturing, exports**	dansk, virksomhed, marked, produktion, industri, danmark, eksport, fabrik, land, produkt, ny
57	**Conflict, negotations**	land, usa, iran, møde, nordkorea, aftale, fn, international, præsident, verden, ifølge
58	Executive	direktør, bestyrelse, ledelse, topchef, formand, chef, bestyrelsesformand, tid, beslutning
59	**Politics, DK**	regering, parti, venstre, dansk, folkeparti, sf, statsminister, socialdemokrat, konservativ
60	East Asian markets	indeks, pct, japansk, japan, yen, ligge, falde, tokyo, asiatisk, marked, stige, morgen, nikkeu

Table A2
LDA model topics, continued

Topic #	Label	Top words
61	**China, trade**	kina, kinesisk, usa, trumpe, amerikansk, vare, land, verden, beijing, kinese, global, tolde, yuan
62	Performing arts	musik, teater, festival, scene, forestilling, publikum, bureau, spille, synge, københavn
63	Art, wine	vin, museum, krone, flaske, sælge, auktion, kunst, maleri, smage, samling, værke, købe, diamant
64	**Middle East, conflict**	angribe, dræbe, by, menneske, ifølge, syrien, blive, soldat, oplyse, reut, irak, person, land
65	**Infrastructure, DK**	dsb, tage, bro, køre, tog, dfds, trafik, færge, forbindelse, passager, københavn, bus, motorvej
66	Church, religion	ole, christiansen, kjær, dan, kirke, fls, patent, præst, niels, koch, ld, lauritzen, thygesen
67	IT, hardware	it, system, flemming, ibm, data, leverandør, columbus, kmd, atea, selskab, hp, csc, kontrakt
68	C25 firms	novo_nordisk, novo, marked, pct, finans, selskab, maersk_line, usa, ny, konkurrent, william_demant
69	**Education, university**	universitet, uddannelse, skole, professor, studerende, forsker, ung, forskning, lære, københavn
70	**Russia, Ukraine, conflict**	rusland, russisk, ukraine, putin, nato, sanktion, russer, land, moskva, ukrainsk, ifølge
71	**Farming**	landbrug, landmand, dansk, arla, danmark, fødevare, danish_crown, fisker, økologisk, produktion
72	Housing, construction	københavn, by, bolig, byggeri, ny, projekt, ejendom, bygge, område, hus, bygning, ligge, lejlighed
73	**US congress**	amerikansk, usa, præsident, new_york, obama, amerikaner, washington, ifølge, skriver, kongres
74	Airlines	sas, fly, lufthavn, flyselskab, selskab, passager, københavn, flyve, norwegian, rejse, ryanair
75	Earnings report	krone, mio, omsætning, pct, koncern, selskab, direktør, resultat, overskud, året, underskud
76	**Europe**	tysk, euro, tyskland, fransk, frankrig, pari, holland, europæisk, europa, berlin, tysker
77	**Finance, regulation**	regel, rapport, lov, undersøgelse, krav, dansk, myndighed, finanstilsyn, sag, mene, oplysning
78	Accidents	klokke, søndag, aften, fredag, morgen, mandag, lørdag, nat, oplyse, time, stede, vagtchef
79	**Elections**	præsident, valg, parti, regering, stemme, politisk, tyrkiet, støtte, land, premierminister
80	Equity analysis	mio, krone, kvartal, vente, resultat, regnskab, forventning, pct, omsætning, første, analytiker
81	**Referendum, EU**	eu, kommission, land, aftale, europæisk, parlament, europa, forhandling, møde, bruxelles
82	Career	ansætte, stilling, arbejde, direktør, inden, virksomhed, ny, ble, danmark, ansvar, udnævne
83	Soccer, World Cup	kamp, vm, danmark, spiller, spille, dansk, emme, holde, gruppe, landstræner, første, landshold
84	Statistics	procent, visere, tal, antal, stige, falde, dansker, sidste, undersøgelse, pct, danmark, sen
85	**Reform, restructuring**	krone, mia, milliard, million, penge, mio, samle, cirka, beløbe, bruge, koste, knappe, året
86	**Climate**	energi, grøn, dong, co, dansk, danmark, el, mål, bruge, klima, miljø, dong_energy, ny, pct
87	Travel	hotel, restaurant, stede, ligge, gammel, tid, hele, by, stå, holde, rundt, gå, god, går, del
88	Transport, trucking	bil, køre, lastbil, chauffør, indien, indisk, bilist, elbil, køretøj, ulykke, vej, dansk, ny
89	**Mortgage**	rente, lån, pct, nykredit, nationalbank, boligejer, stige, boligmarked, dansk, bolig, realkredit
90	Dates	jan, maj, usa, offentliggøre, danmark, mar, ok, klokke, kvt, januar, april, banke, regnskab

Note: Top words for each topic found by our LDA model. Topics in bold are included in our main index.

Figure A2
Number of topics and coherence scores in LDA model

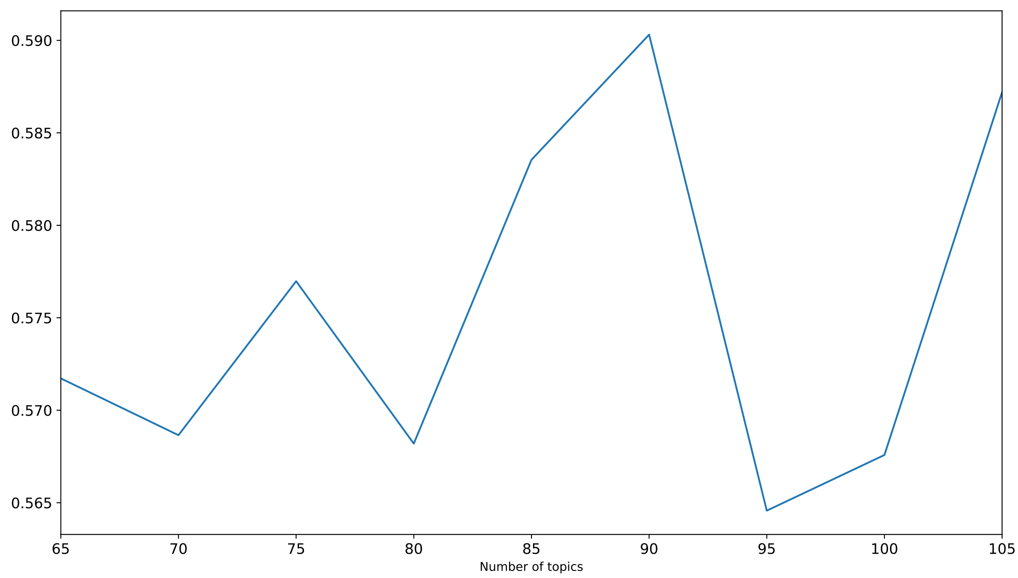
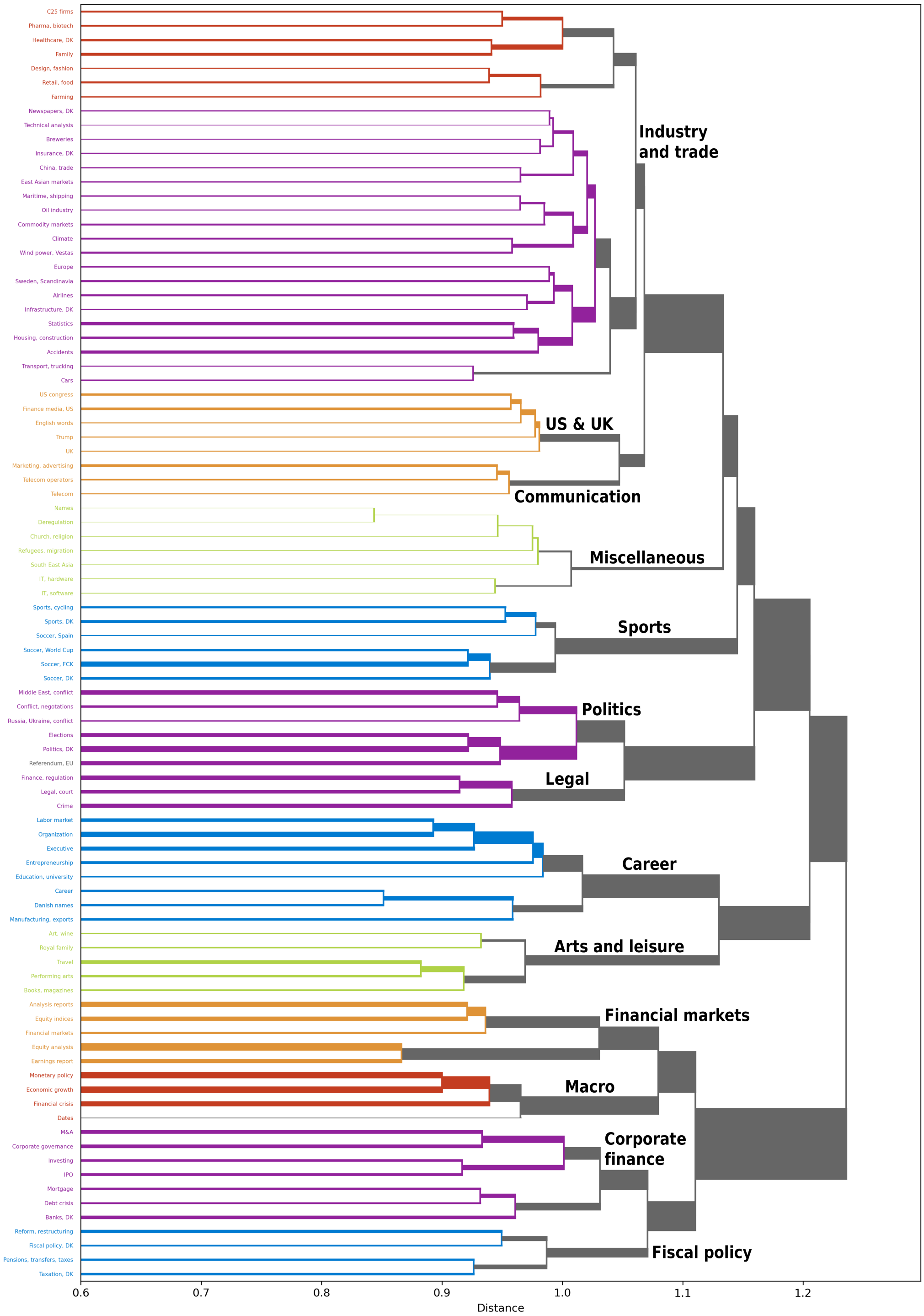
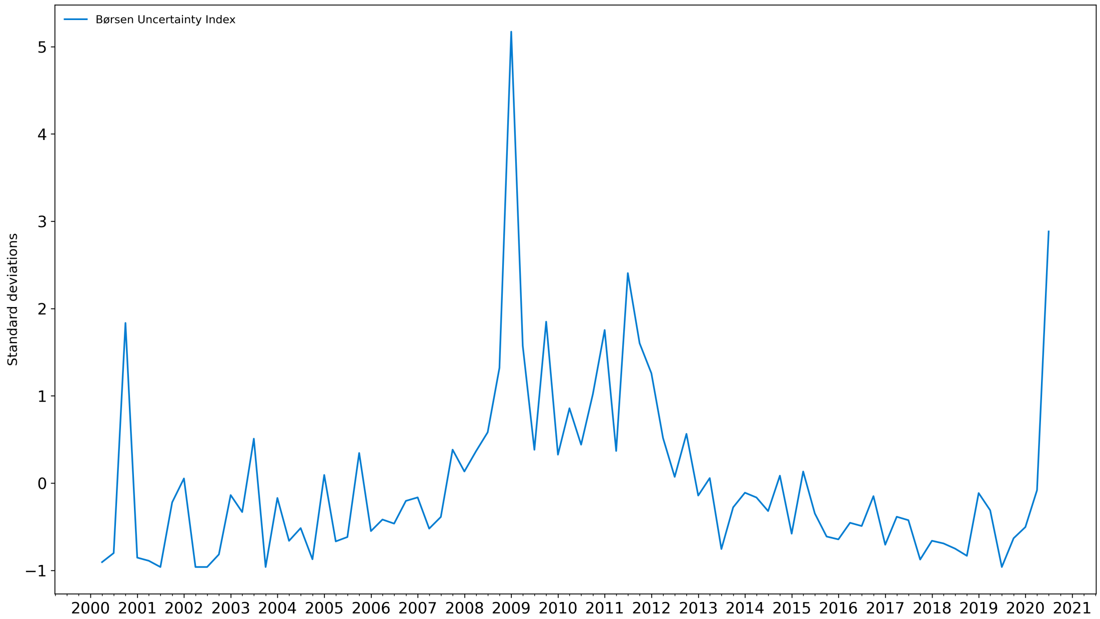


Figure A3
Topic dendrogram



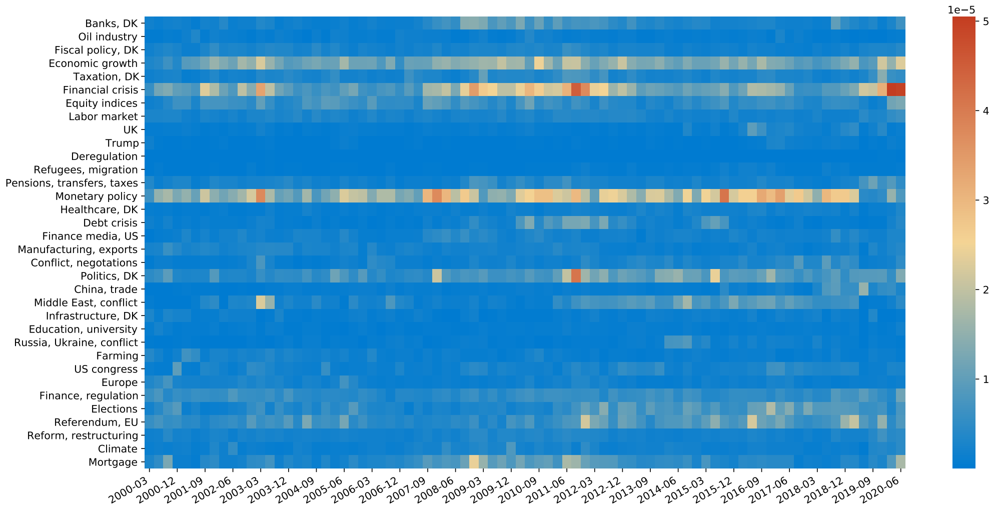
Note: Hierarchical agglomerate clustering of our news articles. Pairs of clusters are formed iteratively based on the similarity of the word distribution that makes up each topic until all topics form one large cluster (the right side of the figure). Line thickness is proportional to the prevalence of the topic in the corpus, i.e. thicker lines are more frequently covered.

Figure A4
Mortgage-related uncertainty.



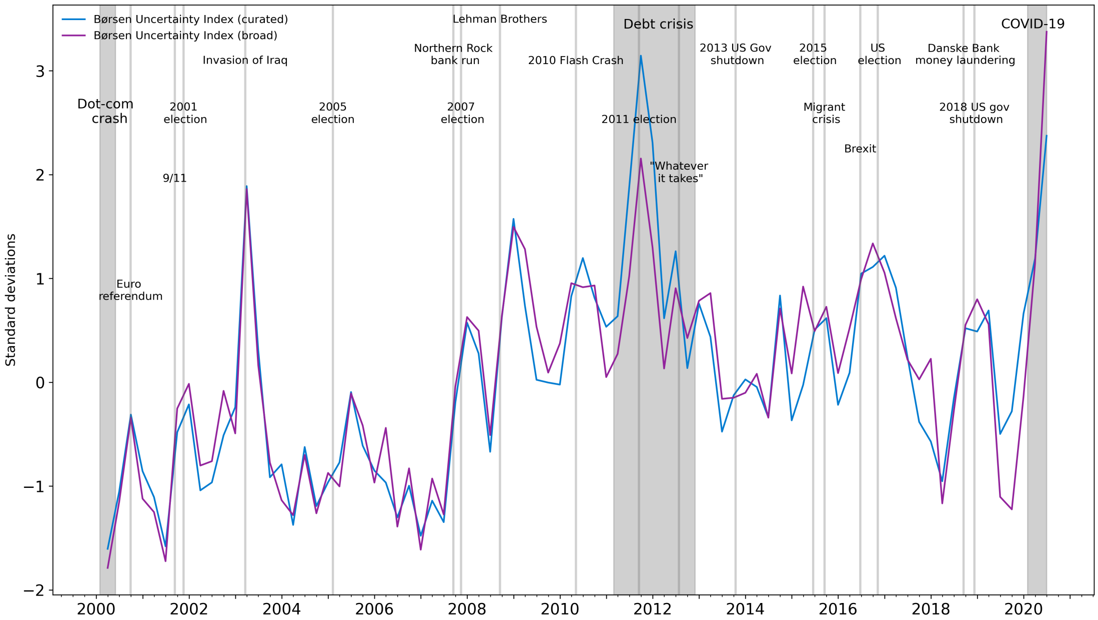
Note: Uncertainty index using only articles where the "mortgage"-topic has a weight of 0.5 or above.

Figure A5
Topic heatmap



Note: A heatmap illustrating raw uncertainty contributions from each topic included in our index.

Figure A6
Broad versus curated index.



Note: A broad 61-topic index and our preferred 34-topic curated index.

Table A3
Preferred settings in the Bayesian estimation procedure

Hyperparameters	Value	Settings for the Gibbs sampler	Value
AR	0.8	Burn-in iterations	5000
λ_1	0.1	Iterations	10000
λ_2	1		
λ_3	1		
λ_4	100		
λ_5	0.001		

Table A4
Hyperparameter selection

		λ_2		
		0.01	0.5	1
λ_1	0.01	452.2	453.6	456.7
	0.05	481.7	482.4	482.0
	0.1	480.2	484.9	**488.0**

		λ_4		
		100	150	200
λ_3	1	**488.0**	483.4	484.4
	2	485.8	480.7	481.1

	λ_5		
	0.001	0.005	0.01
	488.0	481.4	487.4

Note: Marginal likelihood for different hyperparameters. Large numbers are preferred.

Table A5
Ratio of posterior probabilities, marginal likelihood

	$p = 1$	$p = 2$	$p = 3$
$p = 1$	0.0000	0.0156	0.0268
$p = 2$	-0.0156	0.0000	0.0112
$p = 3$	-0.0268	-0.0112	0.0000

Note: The ratio is calculated as the logarithm of the marginal likelihood of row i divided by column j. Any prior beliefs are set to zero. In order to address the preferred model, we use Jeffrey's guideline.

Table A6
Jeffrey's guideline

$x > 2$	Recisive support for row i
$2 > x > 3/2$	Very strong evidence for row i
$3/2 > x > 1$	Strong evidence for row i
$1 > x > 1/2$	Substantial evidence for row i
$1/2 > x > 0$	Weak evidence for row i

Table A7
DIC information criteria

	$p = 1$	$p = 2$	$p = 3$
DIC	**-2439**	-2353	-2257

Note: Small values are preferred.

Figure A7
Structural shocks to uncertainty

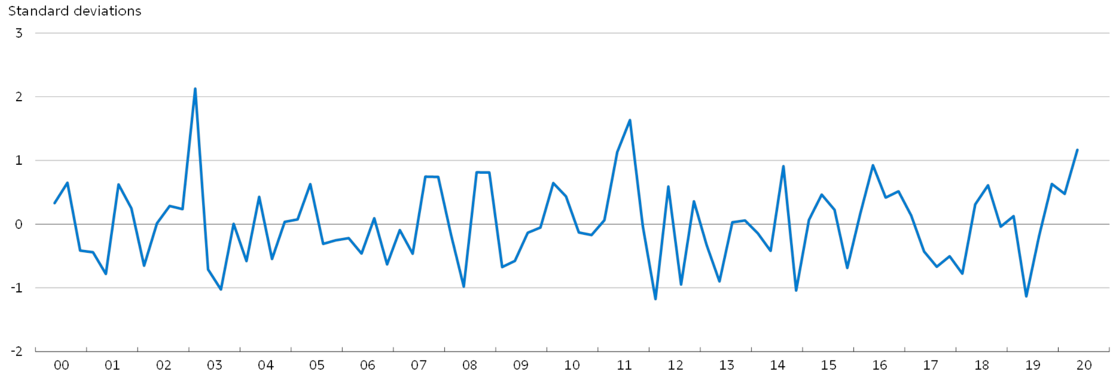
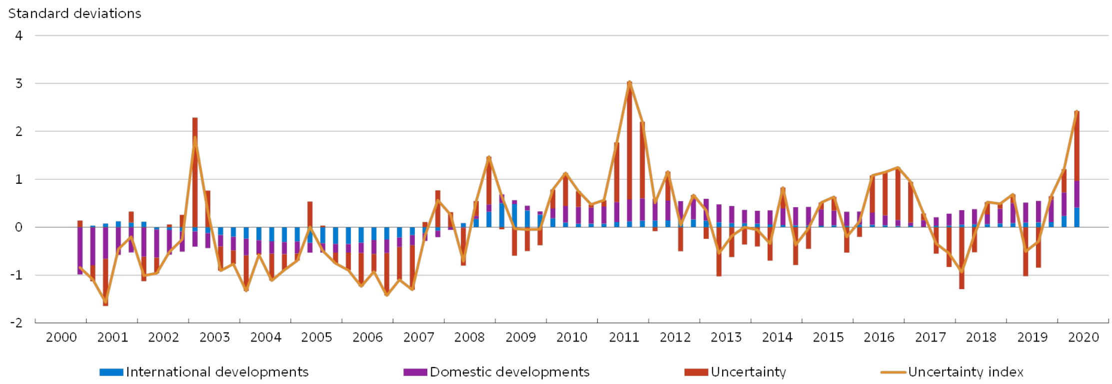
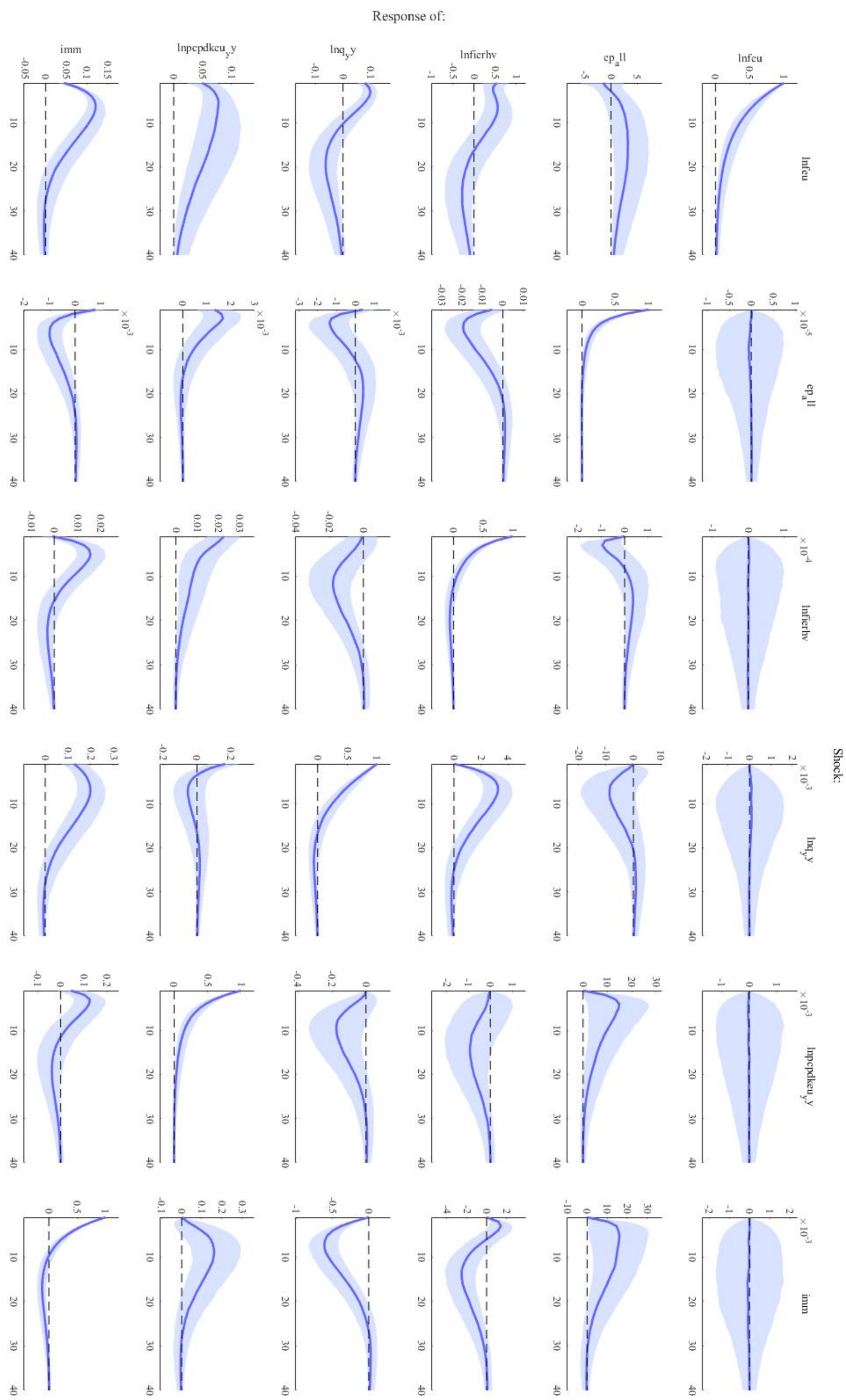


Figure A8
Historical decomposition of the uncertainty index



Note: The historical decomposition of the uncertainty index is calculated from the baseline model. International developments show the contribution to the developments in the uncertainty index from structural shocks to foreign trade-weighted GDP, F_t. Domestic developments show the contribution to the developments in the uncertainty index from structural shocks to business investments, I_t, employment, Q_t, inflation, IN_t, the interest rate, R_t, and exogenous contributions from deterministic variables. Uncertainty shows the contribution to the developments in the uncertainty index from structural uncertainty shocks.

Figure A9
Impulse response functions for our preferred model



DANMARKS NATIONALBANK
ANALYSES

MICRO DATA SCIENCE

Data in new ways

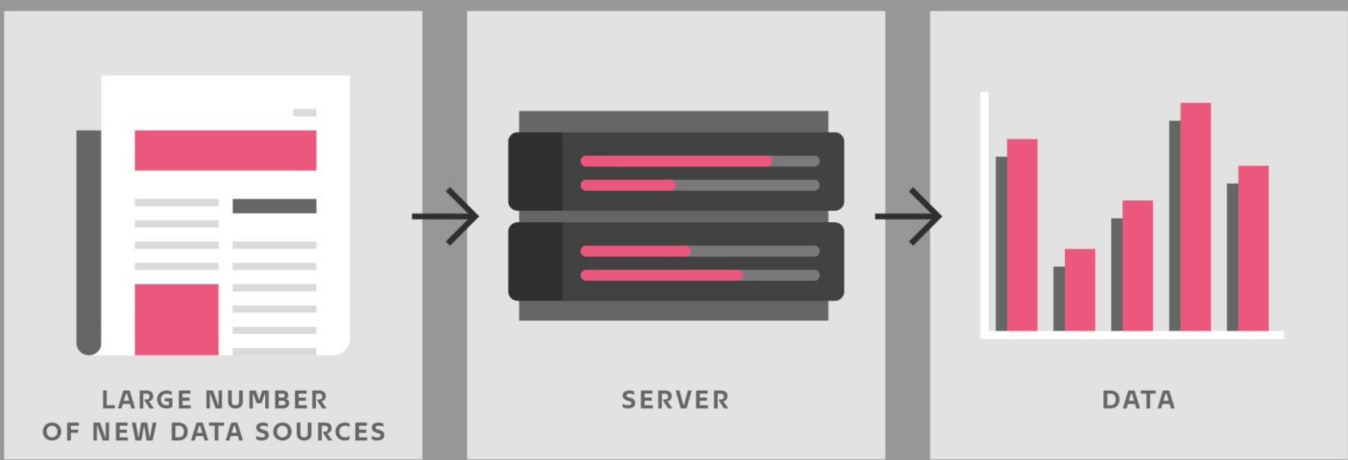
Data volumes have grown exponentially. By 2025, an estimated 450 exabytes of data will be created each day.

This is equivalent to hundreds of millions of personal computers being filled with data on a daily basis. The vast volumes of data are highly diverse, but new and sophisticated methods enable analysis of this data in new and more efficient ways.

New data types and new data collection methods may be used in various contexts in Danmarks Nationalbank's ongoing work.

In order to acquire more knowledge and a better basis for assessing the Danish economy, Danmarks National-bank focuses on new data types and methods in a series of publications of which this Working Paper is one.

New data creates new knowledge



DANMARKS NATIONALBANK

PUBLICATIONS

NEWS

News offers a quick and accessible insight into an Analysis, an Economic Memo, a Working Paper or a Report from Danmarks Nationalbank. News is published continuously.

ANALYSIS

Analysis from Danmarks Nationalbank focuses on economic and financial matter. Some of the analyses are published with a regular frequency e.g. *Outlook for the Danish economy and Financial stability*. Other analyses are published continuously.

REPORT

Report comprises recurring reports and reviews of the functioning of Danmarks Nationalbank. For instance Report includes the *Annual report* and the annual publication *Danish government borrowing and debt*.

ECONOMIC MEMO

Economic Memo is a cross between Analysis and Working Paper and it often shows the ongoing study of the authors. The publication series is primarily targeted at professionals. Economic Memo is published continuously.

WORKING PAPER

Working Paper presents research projects by economists in Danmarks Nationalbank and their associates. The series is primarily targeted at professionals and people with an interest for academia. Working Paper is published continuously.

DANMARKS NATIONALBANK
LANGELINIE ALLÉ 47
DK-2100 COPENHAGEN Ø
WWW.NATIONALBANKEN.DK

As a rule, Working Papers are not translated, but are available in the original language used by the contributor.

Danmarks Nationalbank's Working Paper are published in PDF format at www.nationalbanken.dk. A free electronic subscription is also available at the website. The subscriber receives an e-mail notification whenever a new Working Paper is published.

Please direct any enquiries directly to the contributors or to Danmarks Nationalbank, Communications, *Kommunikation@nationalbanken.dk.*

DANMARKS NATIONALBANK