EUROPEAN CENTRAL BANK
EUROSYSTEM

Working Paper Series

Andrés Azqueta-Gavaldón,
Dominik Hirschbühl, Luca Onorante,
Lorena Saiz

Economic policy uncertainty in the euro area: an unsupervised machine learning approach

No 2359 / January 2020

Abstract

We model economic policy uncertainty (EPU) in the four largest euro area countries by applying machine learning techniques to news articles. The unsupervised machine learning algorithm used makes it possible to retrieve the individual components of overall EPU endogenously for a wide range of languages. The uncertainty indices computed from January 2000 to May 2019 capture episodes of regulatory change, trade tensions and financial stress. In an evaluation exercise, we use a structural vector autoregression model to study the relationship between different sources of uncertainty and investment in machinery and equipment as a proxy for business investment. We document strong heterogeneity and asymmetries in the relationship between investment and uncertainty across and within countries. For example, while investment in France, Italy and Spain reacts strongly to political uncertainty shocks, in Germany investment is more sensitive to trade uncertainty shocks.

keywords— economic policy uncertainty, Europe, machine learning, textual-data

JEL classifications: C80, D80, E22, E66, G18, G31

Non-technical summary

Recently Europe has been affected by an unprecedented number of episodes of uncertainty, including the Great Recession (2008-2014); the euro area sovereign debt crisis (2010-2012); the sanctions imposed on Russia by the European Union (EU) following the Ukraine crisis (March 2014); the Brexit vote (June 2016); and the recent disputes over global trade. Episodes of uncertainty can have detrimental effects on the real economy. In response to uncertainty shocks, firms may reduce their levels of investment, hiring or orders from foreign intermediates, leading to a slowdown in trade and aggregate investment. In turn, consumers may react to increased uncertainty by postponing consumption and increasing precautionary savings.

This paper quantifies different policy-related uncertainty measures across the largest euro area countries, namely Germany, France, Italy and Spain, by looking at the print media. News articles contain information about the current economic outlook. This information shapes the beliefs of economic agents about the state of the economy. However, only a small fraction of all news articles describes relevant economic information. Combining concepts and techniques developed in the context of computational linguistics and data mining, we extract policy uncertainty indicators for these four countries and their respective languages. In particular, we use a continuous bag of words model to retrieve those articles relevant for the uncertainty surrounding the economic outlook, while topic models are used to split the sample into different themes or topics. Combining these two approaches, we extract eight different uncertainty indicators for each country: fiscal, monetary, political, geopolitical, trade/manufacturing, European regulation, domestic regulation and energy.

The diffusion of these indices through time, from January 2000 to May 2019, captures events such as labour market reforms, fiscal adjustments, the Brexit vote and post-referendum period, and the energy crisis. We observe increases in the domestic regulation uncertainty index during events such as the Hartz reforms in Germany, the labour market reforms in Italy and Spain in 2011 and 2012, and the Macron laws in France in 2015. The geopolitical uncertainty index also rose during the Iraq war (particularly in Spain), the Syrian civil war (particularly in France), and more recently during tensions between Russia and the EU. Similarly, the trade uncertainty index has also increased steadily since the beginning of 2018.

To validate our indices, we compare them with several exogenous indices. First, we compare our aggregate EPU index the aggregation of eight individual categories with the EPU indicator developed by Baker et al. (2016), the BBD-EPU index. The BBD-EPU indices for the four largest euro area countries rely on a list of keywords that are an extrapolation of the ones used for the United States. Despite the differences in the methodologies, we observe strong correlations at country level between the two indices. In addition, we set side by side a financial uncertainty index created by adding the sub-indices of finance-related topics and the Eurostoxx implied volatility index (VSTOXX). Once again, we observe a strong correlation between the two. Lastly, we compare the world trade uncertainty indicator created by Ahir et al. (2019) with the one produced at European level by adding all the trade/manufacturing components of our four countries. Both have remained at relatively high levels since the beginning of 2018.

Finally, we assess the economic impact of the policy-related uncertainty indicators in the real economy. More specifically, we assess the effect of these uncertainty indicators on investment in machinery and equipment. We observe that while investment in France, Italy and Spain reacts heavily to political uncertainty shocks, Germany's investment is more sensitive to trade uncertainty shocks. This is plausible since France, Italy and Spain have suffered prolonged periods of political instability, e.g. the yellow vest protests in France, difficulties forming a government in Italy, and the Catalan independence referendum in Spain. With regard to trade uncertainty, which has reached unprecedented high levels recently, it is not surprising that Germany, the biggest exporter country in the euro area, is also the most vulnerable to it.

1 Introduction

Recently the euro area has been affected by an unprecedented number of episodes of uncertainty, including the Great Recession (2008-2014); the euro area sovereign debt crisis (2010-2012); the sanctions imposed on Russia by the European Union (EU) following the Ukraine crisis (March 2014); the Brexit vote (June 2016); and the recent global trade disputes. These episodes have contributed to high levels of policy-related uncertainty in the euro area. Understanding the sources and dynamics of uncertainty affecting the economy is valuable for policymakers, including central banks. Firms are particularly sensitive to uncertainty when making their investment decisions. In response to uncertainty shocks they may reduce their investment, hiring or orders from foreign intermediates, leading to a slowdown in trade and aggregate investment. In turn, consumers may react to increased uncertainty by postponing consumption and increasing precautionary savings, as reflected in the rise in the household saving rate in 2018.

The purpose of this paper is to measure the effect of the different episodes of policy-related uncertainty on investment in the euro area. Economic policy uncertainty (EPU) documents the ambiguity regarding who will make economic policy decisions, and what and when economic policy actions will be undertaken (Baker, Bloom, and Davis (2016)). The EPU is built by aggregating different components such as fiscal policy, monetary policy and geopolitical issues, to name a few. Several studies have reported a strong relationship between investment and overall policy uncertainty (Baker, Bloom, and Davis (ibid.); Gulen and Ion (2015); and Meinen and Röhe (2017)). However, there has not been a study that focuses on specific categories of policy uncertainty in the euro area. This is mainly due to the limitations involved in creating conventional EPU indicators.

The first contribution that this paper makes is to use a method that can consistently categorise the wide sources of economic uncertainty from the media in a wide range of languages and contexts. We do so in two steps: first, we characterise news articles describing economic uncertainty using a continuous bag of words model that represents words as vectors based on their context. This allows us to distinguish the words most closely related to "economy" and "uncertainty" across four languages, namely German, French, Italian and

Spanish, and therefore to retrieve all those articles relevant to economic uncertainty for each country. Failing to do so would induce an increase in the number of false negatives, that is, we would not pick up all the news articles relevant to economic uncertainty.

Second, we use the methodology proposed by Azqueta-Gavaldón (2017) to identify relevant components of economic uncertainty. This approach uses an unsupervised machine learning algorithm that categorises news articles into specific categories of economic uncertainty. The unsupervised nature of the algorithm classifies news articles into topics without the need for previous knowledge on the themes covered in the articles. The algorithm used is called "Latent Dirichlet Allocation" (LDA) and was developed by Blei, Ng, and Jordan (2003). It is a generative probabilistic method that recovers two distributions, namely words-per-topic and topic-per-article distributions. The advantage of this algorithm is that the researcher does not need to come up with individual lists of keywords for each topic, but can apply this method to uncover the structural patterns of any text endogenously.

One of the caveats of this method is that the topics recovered in the form of most probable words need to be interpreted by the researcher. However, in practice the interpretation of topics, even across different languages, is straightforward. Take for instance monetary policy uncertainty. In our application the lowercase words after stemming (i.e. keeping only the root of words) that characterise this topic are: "ezb, "notenbank", "geldpolitik", "zentralbank" or "draghi" for Germany; "taux", "monetair", "europ", "bce", "central" or "inflat" for France; "bce", "tass", "deb", "central", "monetar", "inflazion" or "drag" for Italy; and "tipos", "bce", "monetaria", "inflacion", or "draghi" for Spain. In all languages the words are very similar.

The spikes in this index coincide with episodes of inflation risks (e.g. during the Iraq war due to concerns over oil price increase); the euro area sovereign debt crisis (2010-2012); and the Brexit vote (June 2016). In addition, we examine in detail the evolution of the eight policy-related uncertainty indicators that form the overall EPU index: fiscal, monetary, political, geopolitical, trade/manufacturing, European regulation, domestic regulation, and energy for each country. We observe increases in the domestic regulation uncertainty index during events such as the Hartz reforms in Germany, the labour market reforms in

Italy and Spain in 2011 and 2012, and the Macron laws in France in 2015. The geopolitical uncertainty index rose during the Iraq war (in particular in Spain), the Syrian civil war (in particular in France), and the most recent tensions between Russia and the EU. Furthermore, the trade uncertainty index has increased steadily since the beginning of 2018.

As a validation exercise that goes beyond cross-checks of time-events, we use several exogenous indices (outside our measures) that have a one-to-one mapping (or close to) with our indices. First, we compare our aggregate EPU index (the aggregation of eight individual categories) with the EPU indicator developed by Baker et al. (2016), the BBD- EPU index, for each European country under consideration. The BBD-EPU indices for the four largest euro area countries rely on a list of keywords that are an extrapolation of the ones used for the United States. Despite the differences in the methodologies, we observe strong correlations at country level between the two indices (0.69 for Germany, 0.78 for France, 0.67 for Italy and 0.86 for Spain). Second, we compare a financial uncertainty index created by adding the sub-indices of finance-related topics with the Eurostoxx implied volatility index (VSTOXX). Once again, we observe a strong correlation between the two (0.61 correlation). Both of these indices rose during the 9/11 terrorist attacks, the Iraq war, the financial crisis and the European sovereign debt crisis. We then compare our European trade/manufacturing index (created by adding each country's trade/manufacturing index) with the world trade uncertainty indicator created by Ahir, Bloom, and Furceri (2018).[1] Although this involves less of a one-to-one mapping (the WTU is global, while ours is European), these two items display some similarities (0.55 correlation) and have both remained at relatively high levels since the beginning of 2018.

Following the standard procedure in the literature, we use a structural vector autoregressive (SVAR) model to document the relationship between business investment proxied by investment in machinery and equipment and our EPU index and the eight sub-indices. We first compare the responses of investment to our aggregate EPU index and the one computed through keywords (BBD-EPU). The impact and significance of our index is higher than the BBD-EPU for all countries except for Germany. In the case of the BBD-EPU indices, only the ones for Germany and Italy are statistically significant. This highlights

[1] See https://www.policyuncertainty.com/wui_quarterly.html

the value added of our method when constructing uncertainty indicators.

In addition, the results display heterogeneity in the relationship between investment and the different sub-indices across and within countries. For example, while investment in France, Italy and Spain reacts heavily to political uncertainty shocks, investment in Germany is more sensitive to trade uncertainty shocks. This is plausible, as France, Italy and Spain have suffered prolonged periods of political instability (e.g. the yellow vest protests in France, difficulties forming a government in Italy, and the referendum on Catalan independence in Spain). With regard to trade uncertainty, which has reached unprecedented high levels recently, it is not surprising that Germany, as the biggest exporter country in the euro area, is also most vulnerable to it.

This paper draws on at least two strands of literature. The first concerns research on the impact of uncertainty on investment. Theoretical work on this topic dates back to Bernanke (1983), who finds that high levels of uncertainty give firms an incentive to delay investment when investment projects are costly to reverse.[2] Recently developed macroeconomic models also show that uncertainty has a strong impact on the business cycle. For example, in models with heterogeneous agents, households face periods of high uncertainty in the lower part of the cycle given that uncertainty is endogenously procyclical.[3] From an empirical perspective, there has been an extensive amount of work documenting the detrimental effects of uncertainty on investment (see for example Gulen and Ion (2015), Meinen and Röhe (2017), Jens (2017) or Azzimonti (2018)).

Second, there is a rapidly growing body of literature on textual methods to produce quantitative measures of complex concepts such as uncertainty and risk. In their seminal contribution, Baker, Bloom, and Davis (2016) used newspaper coverage frequency and simple dictionary techniques to measure EPU.[4] Tobback, Nardelli, and Martens (2017) built an indicator of the degree of "hawkishness" or "dovishness" of the media perception of the

[2]A. K. Dixit, R. K. Dixit, and Pindyck (1994) offer a detailed review of the early theoretical literature.

[3]For example, in Bayer et al. (2019), there is a reduction in physical investment as a response to the decline in consumption demand caused by higher uncertainty.

[4]EPU indices have been replicated using more advanced methods (see Azqueta-Gavaldón (2017) and Saltzman and Yung (2018).

ECB's tone using semantic orientation and support vector machine text classification. In addition, they used LDA to detect the dominant topics in the news articles. LDA was also used by Hansen, McMahon, and Prat (2017) to study communication patterns in the Federal Open Market Committee talks. Using simple text-mining techniques, Hassan et al. (2019) built a political risk measure as the share of firm quarterly conference calls that are devoted to political risk for the United States.[5] Finally, Azqueta-Gavaldón (2020) uses LDA and sentiment analysis to study how narratives propagated by the media influence cryptocurrency prices.

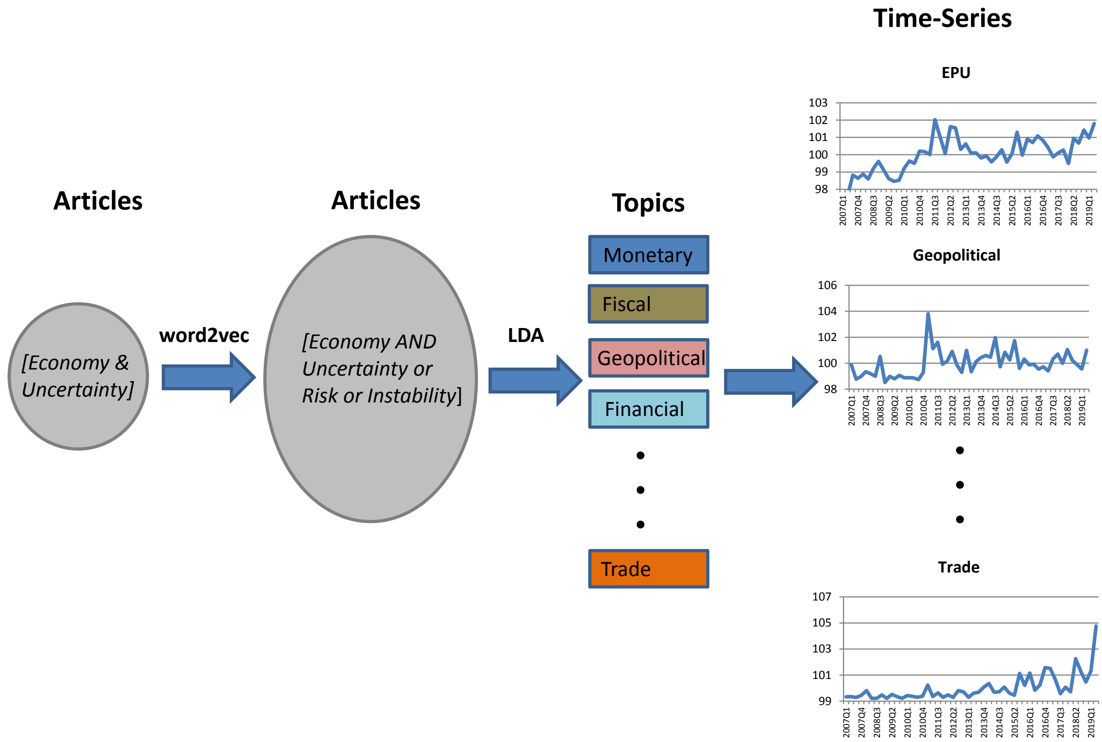
The rest of the paper is structured as follows: Section 2 describes the algorithms and news media data used to produce the EPU indices for Germany, France, Italy and Spain, and compares the resulting aggregate indices with the existing ones; Section 3 describes in detail the individual components that form the aggregate EPU index; Section 4 displays the empirical findings of the effect of EPU sub-indices on the real economy; Section 5 presents the indices validations checks; and Section 6 concludes.

2 Data and Methods

Figure 1 describes the process beginning with gathering news articles to modelling individual components of uncertainty as a time series. This is done in a few simple steps: i) collecting all news articles that contain the words "economy" and "uncertainty"; ii) extending the sample of news articles describing economic uncertainty by including those words that are closest semantically to the above two words in each language ("word2vec" algorithm); iii) running topic modelling algorithms (LDA) to unveil distinctive topics of economic uncertainty; and iv) forming the time series with these topics.

[5] To come up with political topics, they first filter political topics by correlating them to sources using a priori political vocabulary, e.g. political sciences textbooks. They then count the number of instances in which these politics-related words appear together with synonyms of "risk" or "uncertainty".

Figure 1: From News to Time-Series



Notes: The grey circles represent the corpus, i.e. the set of all news articles; "word2vec" stands for the continuous bag of words model developed by Mikolov et al. (2013); and LDA stands for the Latent Dirichlet Allocation algorithm developed by Blei, Ng, and Jordan (2003).

2.1 News articles containing references to economic uncertainty

The first step in creating our indices is to gather all news articles containing any form of the word "economy" and "uncertainty" (language specific). It should be noted that the EPU index developed by Baker, Bloom, and Davis (2016) (BBD) is created using a set of three terms: "uncertainty" or "uncertain"; "economic" or "economy"; and one of the following policy terms: "Congress", "deficit", "Federal Reserve", "legislation", "regulation", or "White House". We select those articles containing the first two of these three terms from the most read newspapers in each country:

- **German newspapers:** Handelsblatt, Frankfurter Allgemeine Zeitung, Die Welt, Süddeutsche Zeitung
- **French newspapers:** Le Figaro, Le Monde
- **Italian newspapers:** Corriere della Sera, La Repubblica, La Stampa
- **Spanish newspapers:** El País, El Mundo, La Vanguardia

From January 2000 to May 2019, the total number of news articles containing any form of the word "economy" and "uncertainty" was 14,695 for Germany, 11,308 for France, 30,346 for Italy and 32,289 for Spain. However, while the words "economy" and "uncertainty" might be well-suited for the English language, this might not hold for other languages. Take for instance the case of German, which has various synonyms for the word "economy" ("Wirtschaft", "Konjunktur", "Volkswirtschaft", "Ökonomie") and the word "uncertainty" ("Unsicherheit") might not map one-to-one onto the English word "uncertainty".[6] Similar complications are also likely to arise in the other languages considered here. For this reason, we need a flexible tool that can perform well in language-specific contexts in order to select all news articles that describe overall economic uncertainty.

To identify the words most similar to "economy" and "uncertainty" for each country (language) we use the continuous bag-of-words model developed by Mikolov et al. (2013), also known as the "word2vec" algorithm. Continuous bag-of-words models are based on the idea that words are similar if they themselves appear near similar words. For example, to the extent that "ECB" or "Fed" tend to appear next to words like "inflation" or "target" one would infer that the two words "ECB" and "Fed" have similar meanings to one another. Continuous bag of words models represent words as a vector, with the elements in each vector measuring the frequency with which other words are mentioned nearby. Given this vector representation, two words are similar if the inner product of their vectors is large.

The most well-known purpose of "word2vec" is to group the vectors of similar words together in the vector space. For example, Atalay et al. (2017) use "word2vec" to create a

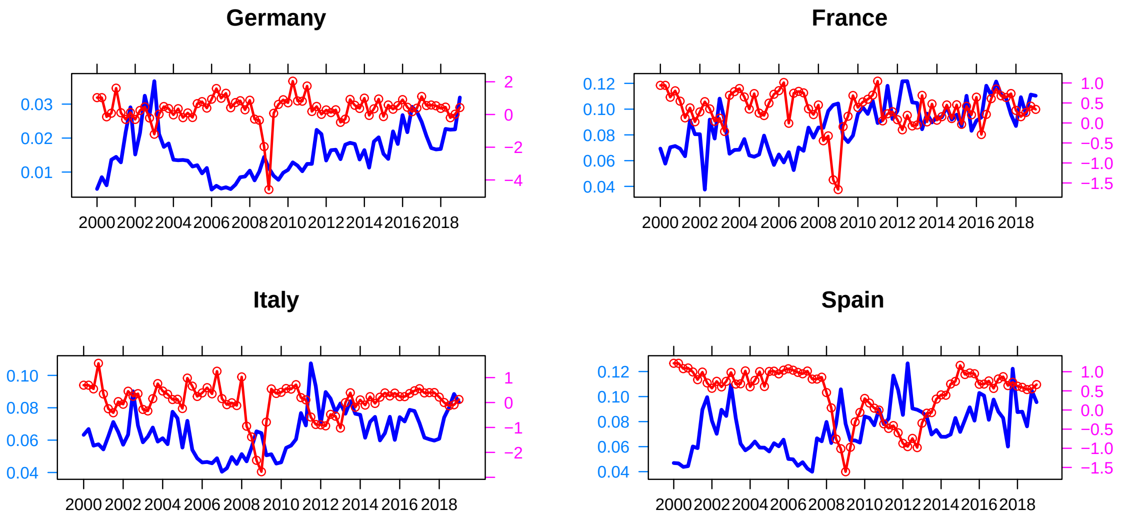
[6]For example, in German the word "Ungewissheit" is often used to express the idea that something is unknown.

list of words related to routine tasks in newspaper job advertisements. Using this method, they show that words related to non-routine tasks have been increasing in frequency, while words related to routine tasks (especially routine manual tasks) declined in frequency between 1960 and 2000. In our case, we want to retrieve the words most similar to "economy" and "uncertainty" across the four different languages. The results reveal that the closest words for "Wirtschaft" are "Konjunktur" (0.61), "Volkswirtschaft" (0.59) and "Ökonomie" (0.56) while for "Unsicherheit" they are "Verunsicherung" (0.73) and "Ungewissheit" (0.63). The number in parenthesis indicates the vector proximity which ranges from 0 (completely opposite or orthogonal) to 1 (exact synonyms).[7] These results seem reasonable, given that, as previously mentioned, "Konjunktur", "Volkswirtschaft", and "Ökonomie" are straight synonyms of the word "economy", while "Ungewissheit" (unknown) is often used to refer to a situation when something is not clear and "Verunsicherung" tends to express a worrisome or a daunting outlook. To see the words retrieved for the rest of the countries, see Appendix A.

As a result, the set of news articles containing the extended list of words related to "economy" and "uncertainty" increases substantially in each country: from 14,695 to 28,941 in Germany's press; from 11,308 to 31,434 for France; from 30,346 to 74,144 for Italy; and from 32,289 to 54,550 for Spain. Figure 2 shows the monthly propagation of this set of news articles (scaled by the total number of them containing the word "today") per country and GDP growth rate. As can be seen, the proportion of news articles describing overall economic uncertainty tends to increase during periods of negative growth rates and events related to geopolitical tensions such as the Iraq war (March 2003) and the recent Brexit referendum (June 2016). This highlights the fact that they are mainly capturing negative events and therefore we do not expect a high level of false positives, e.g. being labelled as characterising rises in economic uncertainty while actually describing falls in economic uncertainty.

[7]The results are based on the standard specification in this literature: size=150; window=10; minimum count=2; and workers=10. For the documentation, see https://radimrehurek.com/gensim/models/word2vec.html

Figure 2: Proportion of news articles describing economic uncertainty in the press (continuous line) and GDP growth rates (dotted line) by country.



Notes: Ratio of the total number of news articles containing words related to "economy" and "uncertainty" over the total number of news articles containing the word "today". Quarterly data from Q1:2000-Q1:2019.

2.2 Topic modelling

Before feeding all the data (raw words per document) into the LDA algorithm to obtain unique topics, we need to pre-process them. *Stopwords*, punctuation, and numbers are removed. *Stopwords* are words that do not contain informative details about an article, e.g., "that" or "me".[8] All words are converted to lower case, and each word is converted to its root in a process known as "stemming".[9]

As mentioned, in order to unveil the distinctive sources of uncertainty, we use the methodology described in Azqueta-Gavaldón (2017). This approach applies an unsupervised machine learning algorithm to all news articles describing economic uncertainty to

[8]Note that the list of stopwords is language-specific. We use the *NLTK* library, see www.nltk.org/

[9]Stemming is language-specific and to carry it out, we use the *SnowballStemmer*: https://www.nltk.org/modules/nltk/stem/snowball.html

unveil their topics. The unsupervised machine learning algorithm, called Latent Dirichlet Allocation (LDA) and was developed by Blei, Ng, and Jordan (2003). It reveals the topics of articles without the need for prior knowledge about their content (unsupervised). Intuitively, the algorithm studies the co-occurrences of words across articles to frame each topic as a composition of the most likely words. In parallel, each article is composed via a distribution of topics. This is done in an unsupervised way, meaning that the algorithm forms these two hidden (or latent) distributions without any labelling of the articles or training of the model before the articles are classified.

The only input observed by the algorithm is the number of words per document. The data generation process (DGP) for each word in each set of documents involves a few simple steps:

1. Select the overall theme of an article by randomly giving it a distribution over topics;
2. For each word in the document:
   (a) randomly pick one topic from the topic distribution chosen in step 1;
   (b) given that topic, randomly choose a word from this topic.

Iterating the second step generates a document while iterating both the first and the second step generates a collection of documents. This does not mean that the algorithm assumes knowledge of topics and words frequencies in them but rather that it uses this simple DGP together with the words from each document to infer the underlying topic structure: topics as a distribution of words, and articles as a distribution of topics. The model recovers these two distributions by obtaining the parameters that maximise the probability of each word appearing in each article given the total number of topics K. In this respect, the probability of a word w_i occurring in an article is:

$$P(w_i) = \sum_{j=1}^{K} P(w_i|z_i = j)P(z_i = j) \tag{1}$$

where z_i is a latent variable indicating the topic from which the *ith* word was drawn (step 1 from the DGP) and $P(w_i|z_i = j)$ is the probability of word w_i being drawn from topic j (step b from the DGP). $P(z_i = j)$ is the probability of drawing a word from topic

j in the current article, which will vary across different articles (step a from the DGP). Intuitively, $P(w|z)$ indicates which words are important to a topic, whereas $P(z)$ is the prevalence of those topics within an article. The goal is therefore to maximise $P(w_i|z_i = j)$ and $P(z_i = j)$ from equation (1).

However, direct maximisation turns out to be susceptible to problems of slow convergence or the algorithm getting stuck in local maxima (Griffiths and Steyvers (2004)). The two most common methods used to approximate the posterior distribution given by these two probabilities are sampling methods (SM) and variational methods (VM). Although SM are asymptotically exact, they are very time consuming as they rely on techniques such as the Gibbs sampler. Alternatively, VMs approximate the posterior distribution of $P(w_i|z_i = j)$ and $P(z_i = j)$ using an alternative and simpler distribution: $P(z|w)$, and associated parameters. Although asymptotically might not be exact, VMs are much faster and therefore suited for larger datasets. We use an advanced type of VM called *online variational Bayes* as proposed by Hoffman, Bach, and Blei (2010) and the practical implementation of Rehurek and Sojka (2010).

Finally, to find the most likely number of topics K, we use a *likelihood* maximisation method. This method involves estimating empirically the likelihood of the probability of words for a different number of topics $P(w|K)$. This probability cannot be directly estimated since it requires summing over all possible assignments of words to topics but can be approximated using the harmonic mean of a set of values of $P(w|z, K)$, when z is sampled from the posterior distribution (Griffiths and Steyvers (2004)). Based on this method we set K to 30 for Germany, France, and Italy, and 40 for Spain.[10]

3 Economic policy uncertainty in the euro area

Baker, Bloom, and Davis (2016) used eight categories to produce their original EPU index for the United States: monetary policy; healthcare; national security; regulation; sovereign debt; entitlement programmes; and trade policy. Although some of these categories will

[10] The likelihood function was run from 10 to 80 topics in intervals of 10.

be common to our four euro area countries, not all will have an exact match. On the one hand, there are categories that are not as relevant in Europe as in the United States. This is the case of national security and healthcare. While there has been some debate over the financing of healthcare systems in some EU countries, in particular during the sovereign debt crisis, this debate did not reach the uncertainty levels of Obama Care in the United States. In the case of the United States, healthcare was a major topic during the 2008 Democratic presidential primaries, as it was meant to affect 30 million uninsured people and went to the Supreme Court in 2012. In addition, while there have been some military interventions by EU states, these did not reach the engagement levels of the United States.

On the other hand, there are certain policy-related events that are unique to EU- countries and are not present in the United States. This is the case of political referendums, such as the Brexit vote or the Catalan referendum for independence, which have greatly contributed to policy uncertainty but do not match any of the eight categories described in the original Baker, Bloom, and Davis (2016) index. Further complications arise from the fact that in the case of the EU, there are policies at the European Union level (e.g. monetary policy), at the individual country level (e.g. military interventions) and at both the EU and country levels (e.g. fiscal policies in the context of the EU Stability and Growth Pact).

The aim is therefore to select those topics that best describe sources of policy uncertainty in the European context. We then select eight categories that best suit the European context and are also easy to identify across our wide range of countries. These categories are: fiscal; monetary; political; geopolitical; trade/manufacturing; European regulation; domestic regulation; and energy. As can be seen in Table 1, with the words that the LDA algorithm gives we can easily label each category/topic. For example, the political topic is framed by words such as “ministry”, “president” or names of heads of states, while the monetary policy topic contains words such as “ECB”, “inflation” and “central bank”.

In addition, we observe some interesting differences across countries regarding the stance taken on specific topics. For example, the words describing the geopolitical category are heavily tuned towards the Russian conflict in the case of Germany, France and Italy, but not in the case of Spain; words relating to Russian-EU tensions such as "Russia", "sanctions" and "Ukraine" appear in all geopolitical indices except in the Spanish one. This is not entirely surprising since the three largest euro area economies (Germany, France and Italy) experienced the highest export losses with Russia in absolute terms as a consequence of the sanctions imposed by the EU. On the other hand, the words in the fiscal category relate to pension and labour reform in the case of Germany (e.g. "Tarifvertrag" meaning collective agreement or "Rente" meaning pension) while for the rest of countries they also include budgetary terms (e.g. "deficit").

To form the aggregate EPU time series at the country level, we follow two simple steps. First, we sum the topic proportions of these five categories by month. This gives us a raw aggregation of the fraction of news articles describing EPU per country. Second, we divide each raw aggregation by the total number of news articles containing the word "today". Figure 3 shows the quarterly EPU indices computed for the four largest economies in the euro area (blue line) and the BBD-EPU index obtained by Baker, Bloom, and Davis (2016) (red line). Overall, the time series produced by grouping the EPU topics retrieved by the LDA algorithm and the BBD-EPU indices are fairly similar (correlations of 0.69 for Germany, 0.78 for France, 0.67 for Italy and 0.86 for Spain).

There are three particular episodes where EPU picked up in the four major euro area economies. The first peak occurred in the first quarter of 2003 with the invasion of Iraq. The second peak corresponds to the European sovereign debt crisis between 2010 and 2012 when the risk premiums of several EU countries reached historically high levels. Finally, the third peak is found around the Brexit vote in the third quarter of 2016. For Germany and France we find high uncertainty peaks during and after the Brexit referendum. This is not entirely surprising since these two countries have stronger trade links with the United Kingdom.[11] For Italy and Spain, the EPU indices display the highest level during the sovereign debt crisis, in particular when the Spanish government requested financial assistance to re-

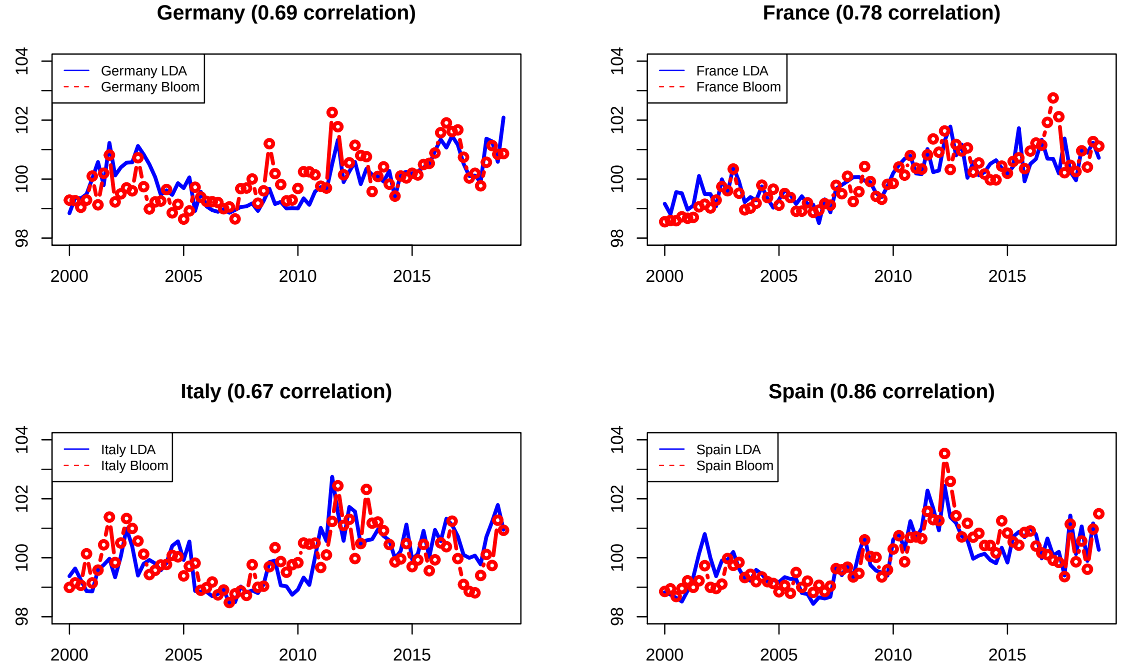
[11] For example, UK imports in 2016 totalled £75.1bn with Germany, £37.6bn with France, £28.0bn with Spain, and £22.6bn with Italy. See https://www.ons.gov.uk/businessindustryandtrade/internationaltrade/articles/whodoestheuktradewith/2017-02-21

Table 1: Most relevant words representing given by the LDA for each category

height	**Germany** Articles = 28,941	**France** Articles = 31,434	**Italy** Articles = 74,144	**Spain** Articles = 54,550
Monetary	ezb, notenbank, geldpolit, prozent, zentralbank, fed, europa, euro, stark, zins, inflation, draghi	taux, économ, euro, monétair, bce, banqu, inflat, baiss, ralent, croissanc	banc, bce, spread, monetar, deb, drag, tass, central, eurozon, titol, inflazion	tipos, bce, monetaria, inflación, draghi, euro, interés, banco, economía
Fiscal	rent, riest, gewerkschaft, arbeitgeb, hartz, iv, metall, ig, tarifvertrag, zeitarbeit	fiscal, impôt, dépens, financ, budget, milliard, tax, retrait, défic, publiqu, réform, prélev	fiscal, manovr, bilanc, pubblic, spes, tagl, deficit, padoan, commission	gobierno, ley, medidas, pensiones, fiscal, reforma, impuestos, presupuestos, déficit
Political	spd, cdu, merkel, koalition, grun, csu, fdp, kanzlerin, schaubl, partei, minist	ministr, président, sarkozy, gouvern, chef, franc, macron, réform, elys	renz, pd, salvin, premier, vot, part, elettoral, leg, polit, palazz, president, leghist	pp, rajoy, psoe, cataluña, partido, elecciones, voto, gobierno, presidente
Geopolitical	russland, russisch, iran, ukrain, putin, sanktion, syri, israel, iran, arabi, krim, irak, barrel, konflikt	militair, iran, armé, arab, iranien, syr, turqu, sécur, irak, guerr, terror, immigr, migr, réfugi, russ, ukrain	terror, lib, sir, iran, arab, iraq, guerr, militar, russ, cines, sanzion, jihad, saud, tunis, sunn, curd	irán, siria, turquía, saudí, guerra, ejército, irak, militar, arabia, refugiados, islámico
Trade / Manufacturing	china, usa, global, trump, weltwirtschaft, zoll, strafzoll, iwf, weltweit, import, protektionismus	produit, agricultur, commerc, lait, viand, omg, industriel, export, producteur, automobile, véhicul, psa	trump, aut, fiat, diesel, automobilist, produtt, industr, settor, export, competit, pmi, manifattur, merc, paes	china, rusia, mundial, pekín, aranceles, comercio, unidos, comerciales, ventas, diésel, fabricantes, seat
European Regulation	eu, brexit, britisch, london, pfund, austritt, brussel, binnenmarkt, votum, parlament, komission	européen, europ, union, ue, brex, grec, bruxel, britainn, allemagn, pay, irland, euro, commiss, referendum, zon	europe, ue, german, tedesc, union, grec, merkel, migrant, bruxelles, brexit, vot, referendum, popul, part	europea, ue, bruselas, grecia, unión, comisin, comunitario, eurozona, socios, brexit, referéndum
Domestic Regulation	regier, kommission, nutzungsrecht, schaubl, rechststaat, justiz, dat, kund, internet, ausbild, fluchtling, arbeit	syndicat, text, cgt, salari, syndical, tribunal, jurid, commiss, emploi, enterpris, travail, embauch	pag, pension, red, gentilon, univers, pdl, scuol, sindac, contratt, sindacal, lavor, sentenz, tribunal	justicia, tribunal, supremo, deuda, bancos, crisis, rescate, laboral, sindicatos, ugt, universidades
Energy	energi, strom, gas, erneuerbar, klimaschutz, rwe, bio, offshor	énerg, électr, edf, gaz, nucléair, pétroli, baril, réacteur, carbon, alstom	ambiental, carbon, energ, climat, elettr, inquin, petrol, gas, baril, petrolifer	energía, climático, emisiones, carbn, gases, electricidad, contaminación

capitalise its banking sector (third quarter of 2012), and the financial turmoil that pushed up the Italian spread leading to the resignation of Berlusconi in favour of the technocrat Mario Monti in Italy (fourth quarter of 2011).

Figure 3: Evolution of EPU indices produced using LDA and Bloom's EPU indices for the four biggest EU economies



Notes: Quarterly time series for the period Q1:2000 - Q1:2019. Each time series is normalised to mean 100 and 1 standard deviation. BBD-EPU indices are obtained from http://www.policyuncertainty.com

3.1 EPU sub-indices

We now describe in more depth the aggregate EPU index and its sub-indices for each country. To make the event validation easier, we display the monthly frequency of each index (as opposed to quarterly frequency as in Figure 3). To show the weights of each sub-index (relative importance), we do not standardise them but display their raw magnitude multiplied by a factor of 100. For example, when a sub-index i reaches 0.1 in a particular month t this would mean that the sum of all topic-article proportions in that given month divided by the total number of news containing the word "today" is 0.1%.[12]

[12] Our research shows that around 15% of all news articles contains the word "today"

Economic policy uncertainty in Germany

Figure 4 depicts the main sources of policy uncertainty that Germany has been exposed to in recent years. As can be seen, the German EPU index effectively identifies several episodes: the 2002 Federal election, the Iraq war, the sovereign debt crisis and the Brexit vote. Not surprisingly, the spike in EPU uncertainty during the 2002 Federal election is recorded by the political uncertainty index. The 2002 election was heavily influenced by the poor economic performance of Germany (the country was in a recession), the introduction of the euro, and the opposition campaign against taxes (particularly on fuel). In 2003, we see an increase in geopolitical and monetary policy uncertainty. The rise in geopolitical uncertainty coincides with the beginning of the Iraq war (March 2003), while the rise in the monetary policy uncertainty index can be attributed to two events. First, the Iraq war put upward pressure on oil prices, creating doubts regarding the monetary policy stance that the ECB should pursue in a context of subdued growth. On the other hand, the clarification of the ECB's monetary policy strategy was interpreted by some observers as a sign of a disappointing ECB performance.[13] In addition, we also observe spikes in the monetary policy uncertainty index from the beginning of the sovereign debt crisis until Mario Draghi's famous quote "whatever it takes" (WIT) in July 2012; in 2015, when the ECB expanded its asset purchase programme to include bonds issued by euro area central governments, agencies and European institutions as part of its non-standard policy measures; and finally, the extension of the ECB's asset purchase programme to the corporate sector in March 2016.

The sub-index fiscal uncertainty index describes national regulation and it shows the most prominent spike during the Hartz reforms (H-R) and, to a lesser extent, coinciding with the presentation of the "refugee integration law" in May 2016 . The H-R aimed at making new types of jobs easier to create (minijobs and midijobs) and changed welfare benefits, in particular unemployment benefits.[14] The "refugee integration law" presented in May 2016 also aimed at integrating refugees by creating 100,000 "one euro jobs" and train-

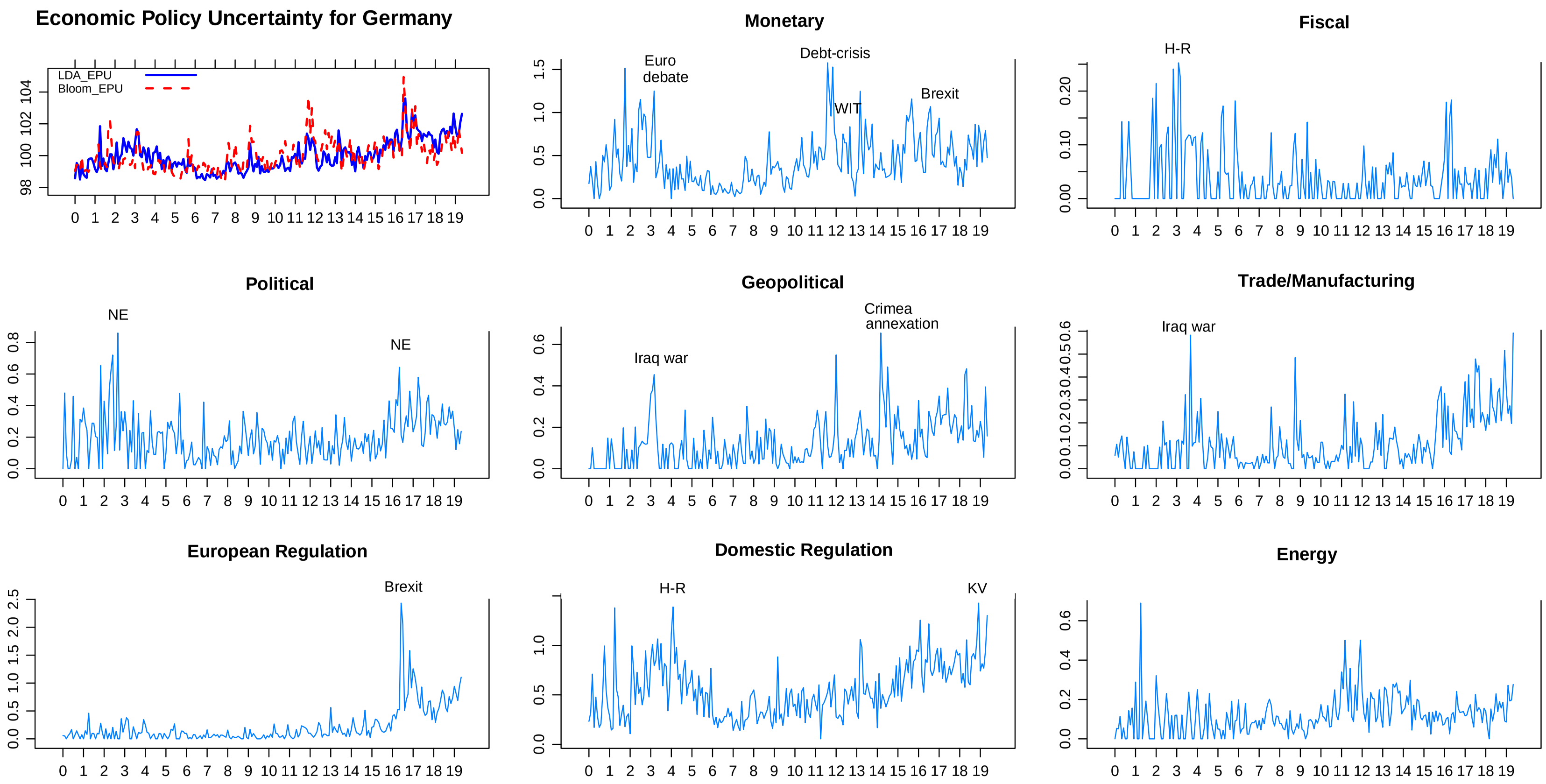
[13]See https://www.ecb.europa.eu/press/key/date/2003/html/sp031120.en.html

[14]They were implemented in several steps: Hartz I-III between January 2003 and 2004, and Hartz IV in January 2005.

ing courses. While the regulation uncertainty index also rose during the Hartz reforms, it also reacts to other regulatory reforms; the major peak in this index took place in Q2:2018 during the coalition agreement (*Koalitionsvertrag*). The deal between the CDU/CSU and SPD included measures to cap the pension contribution rate at 20% and set a floor on replacement rates at 48% of average salaries until 2025. These measures were viewed with scepticism by the IMF.[15] The geopolitical uncertainty sub-index captures the tensions between Russia and the EU, which started as a result of the annexation of Crimea in March 2014.

[15] see https://www.ipe.com/countries/germany/imf-questions-german-coalition-government-pension-measures/10025630.fullarticle

Figure 4: Evolution of German Economic Policy Uncertainty and its individual categories



Notes: **WIT** refers to Mr Draghi's quote "whatever it takes"; **H-R** refers to the Hartz reforms; **KV** refers to Koalitionsvertrag (coalition agreement); and **NE** refers to national elections.

Economic Policy Uncertainty in France

The EPU for France (Figure 5) has been shaped by four main episodes: i) the Iraq war (March 2003); ii) the sovereign debt crisis (2010-2012); iii) the Brexit vote (June 2016); and iv) the presidential election run-off between Macron and Le Pen (April-May in 2017). The first two episodes were the most prominent in terms of the history of the index.

The sub-indices showing the greatest rise in uncertainty during the Iraq conflict were the geopolitical uncertainty sub-index and, to a lesser extent, the monetary policy uncertainty sub-index. The highest peak in the geopolitical uncertainty sub-index occurred in February 2011, around the time the Syrian civil war began. It should be noted that France played an active role during the Syrian civil war and insisted later that year that the Syrian president Bashar Assad should step down.[16]

The second episode of high uncertainty corresponds to the EU sovereign debt crisis (2010-2012). This episode is well captured by the fiscal uncertainty sub-index and partly by the monetary uncertainty sub-index. Although France did not have levels of debt as high as in other European countries, France's credit default swaps escalated by 300% between January 2010 and November 2011. Furthermore, the winner of the 2012 general elections, François Hollande, promised to eliminate France's budgetary deficit (around 7%) by cancelling enacted tax cuts and exceptions to the wealthy and raising the top tax bracket rate to 75% for those with an income over EUR 1 million. For this reason, it is not surprising to also see peaks in the political uncertainty index during this period (May 2012).[17]

Additional spikes in the fiscal policy uncertainty index occurred during the national election of April-May 2017. The policies proposed by the two candidates Macron and Le Pen could not have been more different, which explains the rise in uncertainty. While Le Pen proposed to take France out of the euro, increase welfare benefits, implement a quota to cut immigration by 80%, and introduce more regulated labour reform and protectionism, Macron advocated for free trade, reform of the labour market to make it more flexible,

[16] See for example https://www.theguardian.com/world/2015/nov/14/france-active-policy-syria-assad-isis-paris-attacks-air-strikes

[17] See for example https://www.wsj.com/articles/SB10001424052970204369404577206623454813632?mod=googlenews$_wsj$

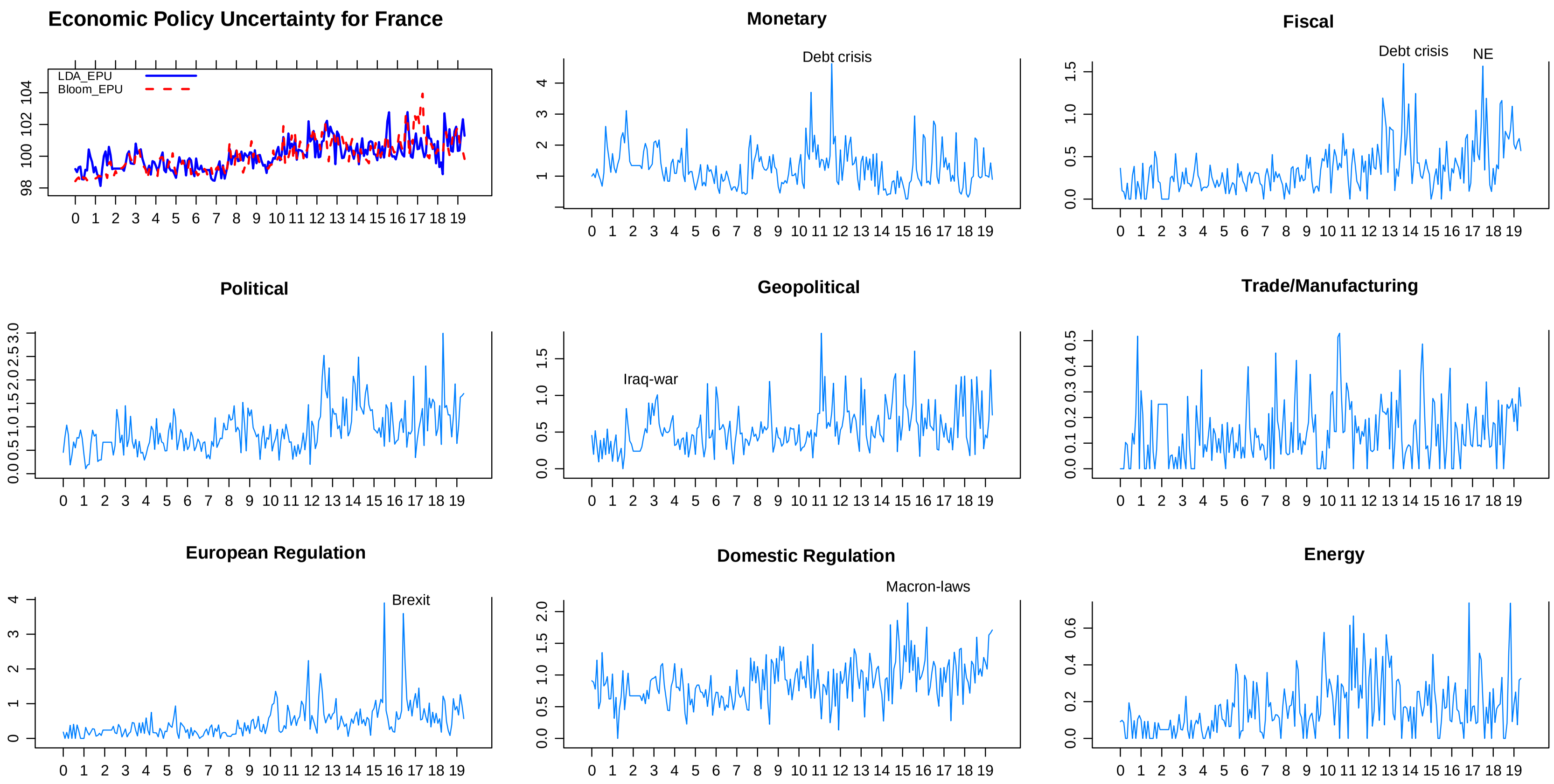
pro-immigration policies, less spending and pro-EU policies. [18] It is worth noting that the French EPU index shows an abrupt peak coinciding with the so-called Macron laws (enacted in August 2015 when Macron was Minister of the Economy and Finance). These laws set in motion an ambitious project to promote growth and employment.[19] Not surprisingly, the domestic regulation uncertainty sub-index captures this event as the highest peak of the index.

Macron's popularity plummeted in 2018, illustrated by the series of protests that were conducted by trade union and left-wing activists during the second half of 2018. In May of that year, several thousand people across France protested against Macron's reforms of the public sector. The political uncertainty index displays the highest peak during this month. In October 2018, Macron announced that the carbon tax would be increased in 2019, triggering the "yellow-vest" protests the following month. This time period coincides with the highest peak of the energy uncertainty index.

[18] see https://www.ft.com/content/fb0ac974-2909-11e7-9ec8-168383da43b7

[19] see https://www.gouvernement.fr/en/law-on-economic-growth-and-activity.

Figure 5: Evolution of French economic policy Uncertainty and its individual categories



Notes: NE refers to national elections.

Economic Policy Uncertainty in Italy

The Italian EPU index is shaped by several events: the stock market downturn of July-Sept 2002; the sovereign debt crisis, which reached several peaks coinciding with the EU Commission's deficit target ultimatum (August 2011); Berlusconi's resignation and replacement by the technocratic cabinet led by Mario Monti (Nov 2011); the Monti-Fornero reforms (June 2012); the Italian constitutional referendum (December 2016); the Italian banking crisis (July 2016); and the 2018 national elections and government coalition agreement between the Five Star Movement and Lega Nord.

The main difference between our index and that of Baker, Bloom, and Davis (2016) (2016) (BBD-EPU) can be observed in the month following the general election of February 2013 when the anti-establishment party, the Five Star Movement, became the third largest party with a 25.5% share of the votes. While it certainly is an episode of high uncertainty, given their unconventional measures proposed, it is hard to see it as the greatest uncertainty episode in Italy's historical EPU index, as in the BBD-EPU index. The highest peak in our index occurs during Berlusconi's resignation and replacement by Monti in November 2011. In this month the monetary, fiscal, and political and domestic regulations uncertainty sub-indices all increased.

Monti undertook several reforms in the country, including the well-known Monti-Fornero reforms (June 2012). The Monti-Fornero reforms, which aimed to raise cash and reassure markets of the commitment to spending discipline, stopped indexing pensions for inflation above a certain income level and gradually increased the retirement age to 67.[20] These reforms are captured by the domestic regulation sub-index.[21] The domestic regulation sub-index also peaked during Italy's banking crisis, which started in July 2016 when Monte dei Paschi di Siena failed the European Banking Authority's stress test. It also peaked during the constitutional referendum held on 4 December 2016. Regarding the banking crisis, the Italian government announced that Monte dei Paschi would be helped via a EUR 8.8 billion government fund for "precautionary recapitalisation". Talks concerning a bailout of Veneto Banca and Banca Popolare di Vicenza soon followed, and Italy's high debt-to-GDP ratio

[20] See https://www.ft.com/content/db0a1d22-3363-11e8-b5bf-23cb17fd1498

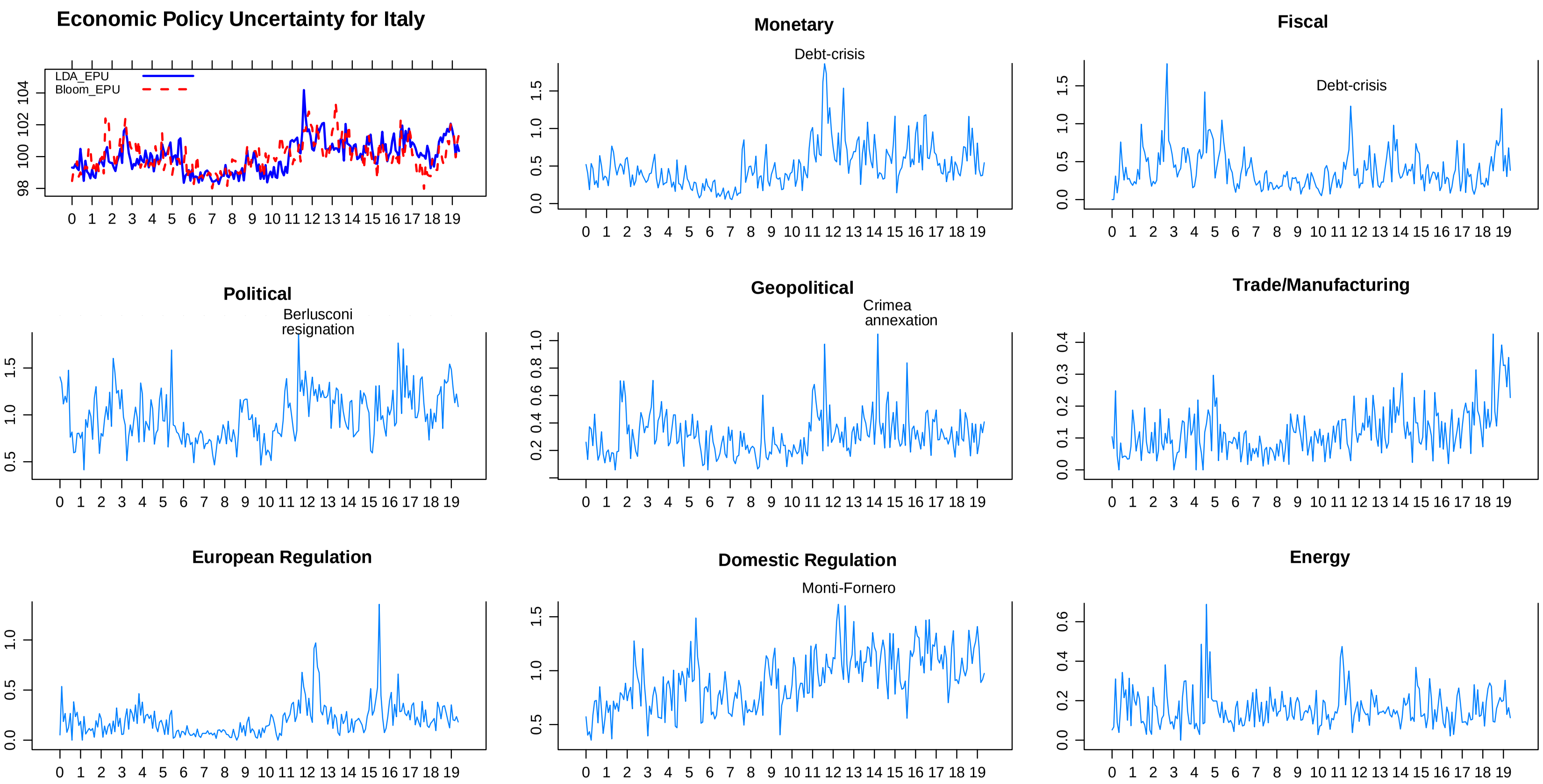
[21] See https://www.economist.com/europe/2012/03/24/montis-labour-law-tangle

second only to Greece among euro area countries raised concerns that a worsening of Italy's banking problems could trigger a sovereign debt crisis (Hodson (2017)).

The constitutional referendum held in Italy on 4 December 2016 represented an ambitious project. Voters were asked whether they approved a constitutional law amending the Italian Constitution to reform the composition and powers of the Italian parliament, as well as the division of powers between the state, the regions and administrative bodies. The proposed reform was rejected with 59% of the votes. Not surprisingly, the political uncertainty sub-index rose during this event.

In 2018 disagreements between the EU and the Italian coalition government (the Five Star Movement and Lega Nord) increased, particularly with respect to fiscal issues. For example, at the end of September 2018 the governing coalition announced its 2019 budget, which increased deficit spending to 2.4 percent of GDP. This triggered a response by the European Commission. These events are captured by the fiscal uncertainty sub-index, which shows major spikes in December 2018. Further large spikes are also visible in the geopolitical uncertainty sub-index during the Syrian civil war, although these are not as pronounced as in the case of France. Interestingly, we also observe increases in the energy uncertainty sub-index in 2011 (February to March 2011), most likely as a consequence of the Libyan revolution. Libya, a former Italian colony, had always been a central focus of Rome's foreign policy and one of the largest suppliers of oil and natural gas to Italy. In March 2014, the geopolitical uncertainty index rose once again, most likely as a consequence of the annexation of Crimea and the second Libyan civil war.

Figure 6: Evolution of Italian economic policy Uncertainty and its individual categories



Economic policy uncertainty in Spain

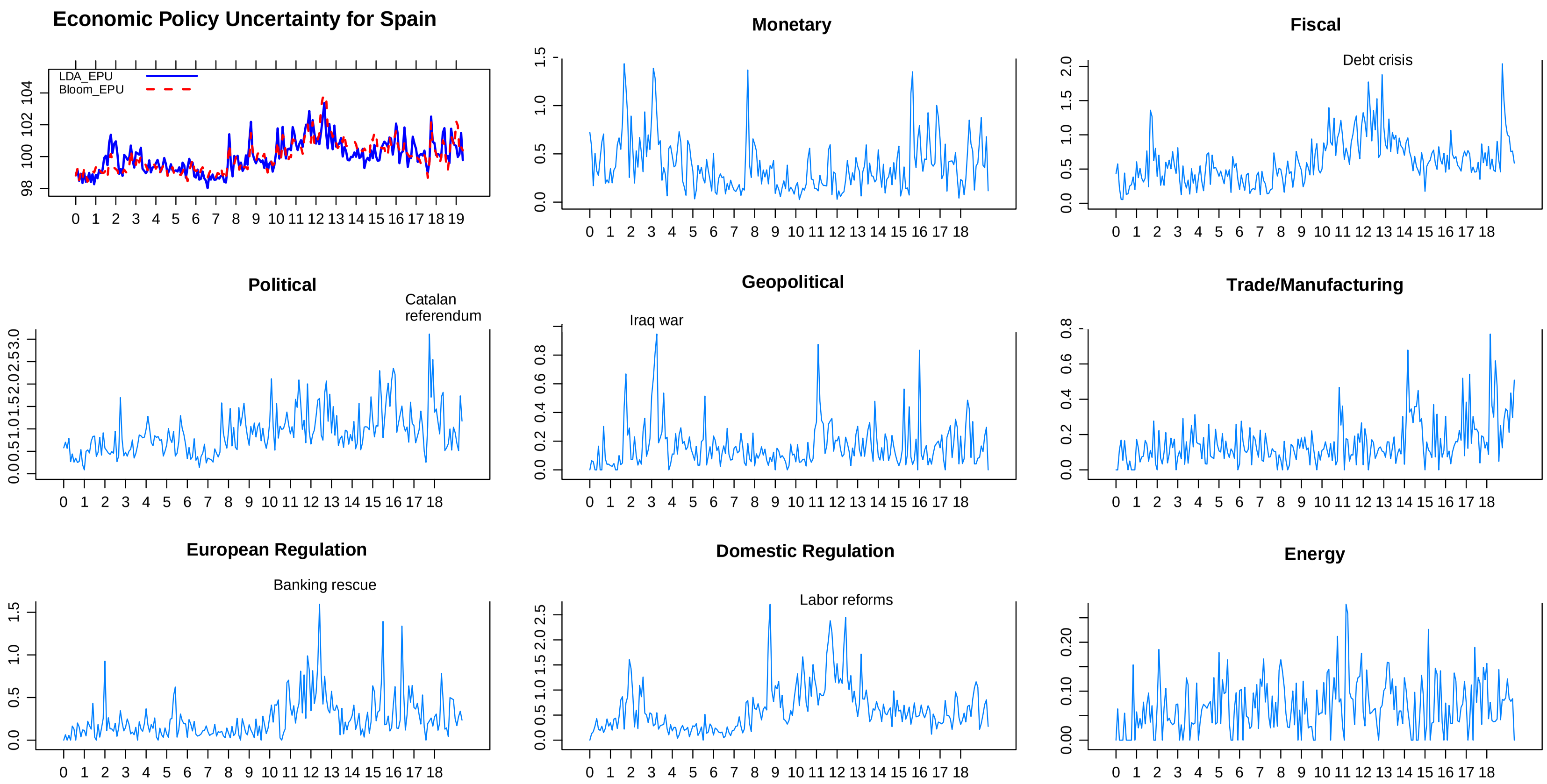
The Spanish EPU index is shaped by three main events: i) the Iraq war at the beginning of 2003; ii) the sovereign debt crisis (2010-12) with its pinnacle during the period of banking recapitalisation (June 2012); and iii) the Catalan independence referendum (Oct 2017).

The high level of uncertainty during by the Iraq war (March 2003) is reflected in the geopolitical uncertainty and monetary policy uncertainty sub-indices. As mentioned before, there were some concerns regarding an increase in oil prices and the possible interventions of the ECB. While there is no doubt that the Iraq war was a major source of uncertainty, some people questioned the BBD-EPU index for Spain for presenting it as the highest uncertainty point in the history of the index. For example, while revising the index by incorporating new keywords and new media sources, Ghirelli, Pérez, and Urtasun (2019) found the highest point of the index to occur during the period of banking recapitalisation (June 2012). This is also in line with our own index.

During this time, Spain experienced a sovereign debt crisis (2010-12) with the Spanish risk premium reaching all-time highs. We can observe three sub-indices rising during this period, namely those relating to fiscal policy, monetary policy and domestic regulation uncertainty. All indices peaked when the Spanish government requested financial assistance from the EU for banking recapitalisation (June 2012).[22] Another important episode recorded by the fiscal and domestic regulation uncertainty sub-indices is the Spanish labour reform of September 2010. This reform was an early step towards tackling the crisis. It included measures such as the suspension of collective agreements (making it possible for employers and workers to suspend collective agreements in case of economic downturns); a reduction in the compensation payments for layoffs; and cheaper dismissals for companies facing losses. Finally, the Catalan crisis sparked a debate on autonomous regional powers both in Catalonia and at the national level. Consequently, the Catalan referendum held in October 2017 was accompanied by an increase not only in the political uncertainty sub-index, but also in uncertainty regarding domestic regulation.

[22]The European Stability Mechanism provided Spain with up to EUR 100 billion in assistance, although in the end it only needed EUR 41.3 billion. Two disbursements were made, in December 2012 and February 2013 respectively.

Figure 7: Evolution of Spanish Economic Policy Uncertainty and its individual categories



Notes: NE refers to national elections.

4 EPU and economic activity

4.1 Model Specification and Identification

Following the standard approach in the literature, we investigate the relationship policy uncertainty and investment in a structural vector autoregression (VAR) framework.

We follow the procedure of Baker, Bloom, and Davis (2016) and specify a VAR using the natural logarithm of EPU, the quarter-on-quarter growth rate of the stock market index, the shadow short term interest rate (SSR) for the euro area[23], the quarterly growth rate of real investment in machinery and equipment as a proxy for business investment and the quarterly growth rate of real GDP. Including the stock market index mitigates concerns of endogeneity because stock markets are forward-looking and stock prices react to all sources of information (Baker, Bloom, and Davis (ibid.)). The data for each stock market index comes from Datastream, while the rest of the data is obtained from Eurostat.

The VAR is run at quarterly frequency. The estimation period is Q1 2000-Q1 2019. We estimate the model as the pth-order VAR:

$$y_t = B_1 y_{t-1} + ...B_p y_{t-p} + u_t \tag{2}$$

$$u_t \sim N(0, \Sigma), \tag{3}$$

where y_t denotes a $q \times 1$ vector of endogenous variables, u_t a $q \times 1$ vector of errors, and $B_1..., B_p$, and Σ represent matrices of suitable dimensions containing the unknown parameters of the model, coefficients of lagged endogenous variables ($B_1..., B_p$), and the covariance matrix (Σ). Since the VAR model is estimated using quarterly data, we follow the common practice in the literature and include three lags. To overcome possible "overfitting" issues we employ Bayesian estimation techniques. Note that "overfitting" might be an issue given our relatively short sample period, i.e. quarterly data and 19 years of

[23]Following the common practice in the literature, we use the shadow short rate (SSR) (see Meinen and Röhe (2017)). The SSR aims to measure the accommodation in monetary policy when the short rate is at the zero lower bound (ZLW). The SSR is obtained from Leo Krippner's website at the https://www.rbnz.govt.nz/research-and-publications/research-programme/additional-research/measures-of-the-stance-of-united-states-monetary-policy/comparison-of-international-monetary-policy-measures

observations. In this respect, we use an independent normal-inverse Wishart prior, assuming that $\beta \equiv vec(c, \gamma, B_1, ..., B_p)$ is normally distributed and that Σ has an inverse Wishart distribution with scale S and ν degrees of freedom:

$$\beta \sim N(b, H) \tag{4}$$

$$\Sigma \sim IW(S, \nu) \tag{5}$$

The prior for β is the Minnesota-type. Specifically, i refers to the dependent variable in the ith equation, j to the independent variable in that equation, and l to the lag number. We then assume that the prior distribution for β is defined such that $E[(B_l)_{ij}] = 1$ for $i = j$ and $l = 1$ and 0 otherwise, while all other elements in b are set to zero. The diagonal elements of the diagonal matrix H are defined as $(\frac{\lambda_1}{l^{\lambda_3}})^2$ if $i = j$ and $(\frac{\sigma_i \lambda_1 \lambda_2}{l^{\lambda_3} \sigma_j})^2$ if $i \neq j$. The prior parameters σ are specified using ordinary least squares (OLS) estimates of univariate AR(1) models. More specifically, σ_i and σ_j denote the standard deviations of error terms from the OLS regressions. Given that our dependent variable is in growth rates, we do not include either a trend or a constant.

The hyperparameters λ_1 to λ_4 are set in accordance with standard values commonly used in the literature.[24] For the inverse Wishart distribution prior, the degrees of freedom v amount to $T + q + 1$, with T denoting the sample length. The scale parameter S is a q diagonal matrix with diagonal elements σ_i^2. Lastly, a Gibbs sampling approach is employed to generate draws of β and Σ from their respective marginal posterior distribution. In this respect, we simulate 10,000 draws and discard the first 10% as a burn in.

To calculate the impulse-response function (IRF), as in Baker, Bloom, and Davis (2016) the structural shocks are identified using a Cholesky decomposition based on the following variable ordering: EPU, stock price index, shadow short rate, investment in machinery and equipment and GDP.

[24]That is, we set hyperparamenters $\lambda_1 = 0.2$, $\lambda_2 = 0.5$, $\lambda_3 = 1$, and $\lambda_4 = 100$

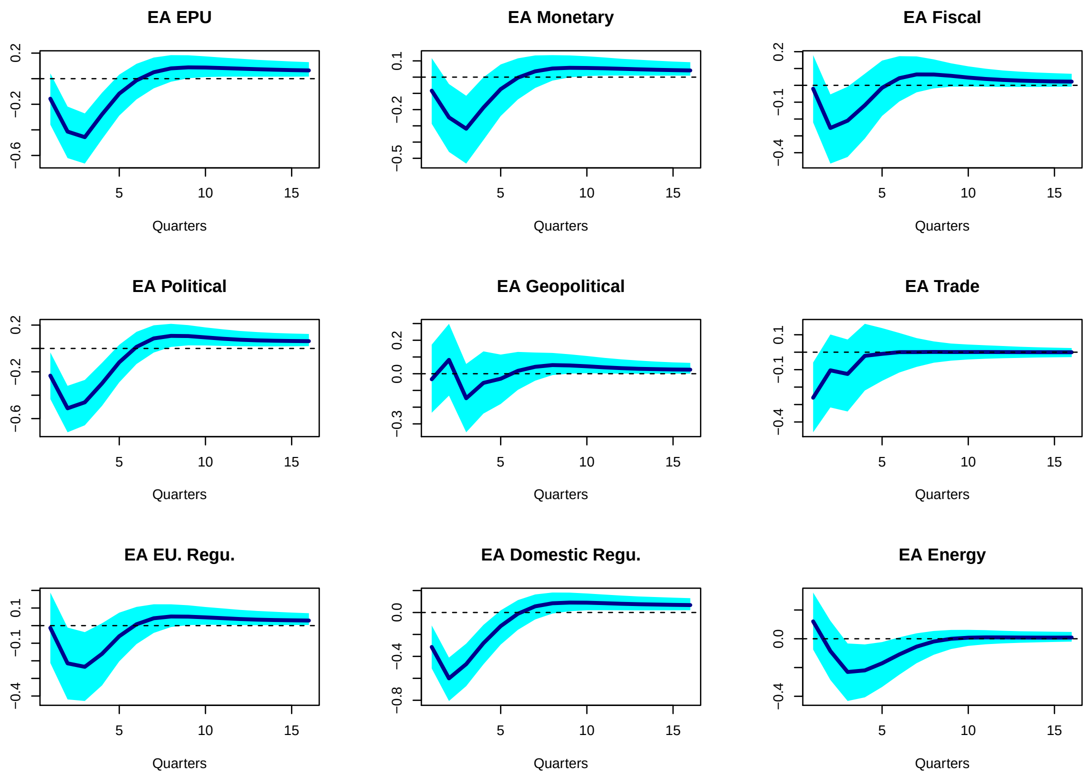
4.2 Results

Figure 8 displays the impulse responses of investment in machinery and equipment for the euro area to shocks in the different EPU sub-indices. Note that the aggregate index at the euro area level is the weighted sum of the different country components. For example, to obtain the euro area monetary policy uncertainty index, we sum each of the monetary policy uncertainty indices of the four countries. Similarly, we construct aggregate indices for the eight sub-indices and the aggregate EPU index.

Overall, we observe a strong and significant impact of increases in EPU uncertainty on business investment proxied by investment in machinery and equipment in the euro area. This significant negative impact lasts four quarters and rebounds after the fifth quarter. This is consistent with the idea that once uncertainty is resolved, firms increase investments to satisfy pent-up demand (Gulen and Ion (2015)). In addition, we observe that only some uncertainty sub-indices have a particularly detrimental effect on investment in the euro area. These are domestic regulation, political, monetary and fiscal uncertainty. In contrast, we find that geopolitical, trade and energy uncertainty barely have any significant negative effects on investment. The relationships between uncertainty and investment that we see at the aggregate (euro area) level might be heterogeneous at the country level. For this reason, we then run the same VAR exercise feeding data at the country level. Figures 9, 10, 11, 12 show the impulse response functions (IRFs) for each EPU category (and aggregate) for Germany, France, Italy and Spain respectively. The top left panel of these figures shows the aggregate effect of EPU on investment. The blue line represents the dynamics of investment in response to one shock of our EPU index on investment while the red line reflects the dynamics of investment using the BBD-EPU indices (the old version of the BBD-EPU in the case of Spain).[25] Altogether, the responses of investment to overall policy uncertainty are negative. However, the impact and significance of our index seems higher than the BBD-EPU indices for all countries except for Germany. It is worth noting that in the BBD-EPU indices, only the indices for Germany and Italy are statistically significant. This highlights the value added of our method when constructing uncertainty indices.

[25] Note that for Spain, we use the original uncertainty index: https://www.policyuncertainty.com/europe_monthly.html

Figure 8: Impulse-response functions of machinery and equipment investment in the euro area (EA) to shocks in EPU index and its components



Notes: SVAR-estimated impulse response functions for machinery and equipment investment to a positive EPU shock (one standard deviation). The SVAR is estimated using Bayesian methods and the shocks are identified using the Cholesky decomposition with the variables in the following order: log(EPU), Δ(log(EuroStoxx price index)), shadow short rate (SSR), Δ(log(M&E)) and Δ(log(GDP)), where Δ indicates first differences or quarterly growth rates. Fit to quarterly data from Q1:2000 - Q1:2019. The blue bands represent the 68% confidence interval.

Regarding each individual category, we observe some interesting heterogeneity. For example, the results display a particularly strong effect of the trade uncertainty sub-index on Germany's investment while not for the other countries. This is not entirely surprising given that, as the biggest exporter of the euro area, we would expect Germany to be especially vulnerable to trade disputes. Regarding the political uncertainty index, we observe the opposite effect: this matters for all countries except for Germany. This is plausible since France, Italy and Spain have suffered prolonged periods of political instability. Regarding the fiscal uncertainty index, we observe that it is only relevant for France while not for Germany, Spain or Italy. This is a bit puzzling given that Spain and Italy have undergone significant episodes of fiscal distress. Nonetheless, much of the uncertainty registered during this period is captured by the monetary or domestic regulation uncertainty sub-indices. As such, we observe a particularly strong effect of monetary policy uncertainty on Italy's investment.

In addition, domestic regulation shows a strong impact only for Italy and Spain, the two countries that went through banking rescues and major fiscal and labour reforms. Furthermore, we observe that the European regulation uncertainty index has only a negative effect in Italy and Spain although it is not statistically significant in the case of Spain. Finally, we observe that the geopolitical and energy uncertainty indices show no or only a negligible impact on investment in all countries.

5 Robustness checks

To assess whether news articles and in particular the set chosen in our set-up are valid for measuring uncertainty, we draw a comparison with uncertainty indices that roughly represent ground truths. By ground truths we mean an accurate and alternative index with which we can compare our indices. This is the case for financial uncertainty, represented by implied volatility indicators such as the VIX for the United States, VFTSE for the United Kingdom, and the VSTOXX for Europe.

Implied volatility indices are based on stock market data, are forward-looking and are often referred to as the "investor fear gauge" (Whaley (2000)). Most importantly, implied volatility indices are often used as a proxy for financial uncertainty (see, for example, Baker, Bloom, and Davis (2016) and Gulen and Ion (2015)). We compare the European implied volatility index, VSTOXX, to a financial uncertainty index computed by adding all those finance-related topics retrieved by the LDA. To this end, we select those topics that are characterised by the following words:

- **German financial uncertainty:** *dax, prozent, akti, punkt, bors, analyst, anleg,* *leitindex, index, rendit, anleg, fond, anleih, investment*

- **French financial uncertainty:** *bours, indic, cac, investisseur, march, séanc, street, wall, valeur, semain, point, actionnair, group, capital, fusion*

- **Italy financial uncertainty:** *bors, rialz, dollar, wall, street, listin, titol, fed, merc, azionar, investitor, mediagroup, banc, carig, soc, azion, mps*

- **Spain financial uncertainty:** *bolsa, inversores, ibex, puntos, mercado, dólares, wall, street, banco, entidad, bankia, millones, entidades, cajas, bbva*

Panel (a) of Figure 13 shows the evolution over time of the index computed by aggregating the topics above and the European implied volatility index, the VSTOXX. Overall, we see a strong resemblance between the two indices, with a 0.61 correlation. The first major spike reported by both indices took place at the time of the 9/11 terrorist attacks, which produced a shock in the financial markets' liquidity worldwide (Posner and Vermeule (2009)). Note that this spike is more abrupt in the index computed by aggregating topics than the VSTOXX. While news reported in the media is cumulative over a whole month, the index reported by the VSTOXX is an average over the whole month. The early decision of the Federal Reserve to provide liquidity, thereby enabling payments to firms and individuals, calmed the markets a few days after the terrorist attacks.

The most prominent spike in the VSTOXX index corresponds to the beginning of the recent financial crisis. Here we observe an interesting phenomenon; while the VSTOXX shows a major spike, it is just above average for the one computed using LDA. It should be noted that we have pre-selected those news articles describing economic uncertainty. We think the explanation lies in the fact that at the beginning there was no clear idea of whether this financial shock would have substantial effects on the real economy. In support of this argument, we observe that in the next major spike, i.e. the one which occurred during the sovereign debt crisis of August 2011, our index increases more abruptly than the VSTOXX. This seems to support the evidence that the index produced by aggregating finance-related topics (within the economy uncertainty spectrum) is somehow more tuned towards the real economy rather than purely financial events.

In addition, we compare the European trade uncertainty sub-index computed by aggregating those topics under the trade/manufacturing category with the world trade uncertainty index developed by Ahir, Bloom, and Furceri (2018). We acknowledge that this latter index is less close to being a ground truth given that it is computed at the global level rather than at the European level. Despite these differences, we observe some resemblance in the form of a 0.55 correlation. Most notably, both indices show a strong upward trend from mid-2018 onward when the China-US trade disputes emerged.

6 Conclusions

This paper presents a novel model-based bottom-up approach to estimate economic policy uncertainty in the euro area and its sub-components. In our approach, we run an unsupervised machine learning algorithm on news articles describing overall economic uncertainty on the German, French, Spanish and Italian newspapers. This allows us to endogenously extract individual uncertainty components and to assess their weight on the overall EPU. In this sense, we find that while the fiscal policy uncertainty component was quite significant for Spain and Italy when the sustainability of public finances was an important issue, it barely featured in the case of Germany and France.

Using the distinct measures unveiled by the algorithm, we document heterogeneity in the relationship between aggregate investment in equipment and machinery and our EPU sub-indices. While investment for France, Italy and Spain reacts heavily to political uncertainty, Germany's investment is more sensitive to trade uncertainty. In addition, Spanish and Italian investment is highly tuned towards domestic regulation uncertainty.

Our results have two main implications. First, they suggest that when building text-based economic policy uncertainty measures, even with press media using a language other than English, it is useful to use techniques beyond word counting. In this respect, we have shown how using a continuous bag of words model makes it possible to retrieve those articles relevant to economic uncertainty for each country, while LDA can be useful when categorising the individual components of EPU. Second, our results highlight the heterogeneity in the relationship between different types of uncertainty and the real economy. Regulators and politicians should then be aware of which type of uncertainty is materialising since, depending on the source, they will be more or less detrimental to the real economy.

References

Ahir, Hites, Nicholas Bloom, and Davide Furceri (2018). "The world uncertainty index". *SSRN 3275033*.

Atalay, Enghin, Phai Phongthiengtham, Sebastian Sotelo, and Daniel Tannenbaum (2017). "The evolving US occupational structure". *Technical Report*.

Azqueta-Gavaldón, Andrés (2017). "Developing news-based economic policy uncertainty index with unsupervised machine learning". *Economics Letters* 158, pp. 47–50.

— (2020). "Causal inference between cryptocurrency narratives and prices: Evidence from a complex dynamic ecosystem". *Physica A: Statistical Mechanics and its Applications* 537, p. 122574.

Azzimonti, Marina (2018). "Partisan conflict and private investment". *Journal of Monetary Economics* 93, pp. 114–131.

Baker, Scott R, Nicholas Bloom, and Steven J Davis (2016). "Measuring economic policy uncertainty". *The quarterly journal of economics* 131.4, pp. 1593–1636.

Bayer, Christian, Ralph Lütticke, Lien Pham-Dao, and Volker Tjaden (2019). "Precautionary savings, illiquid assets, and the aggregate consequences of shocks to household income risk". *Econometrica* 87.1, pp. 255–290.

Bernanke, Ben S (1983). "Irreversibility, uncertainty, and cyclical investment". *The Quarterly Journal of Economics* 98.1, pp. 85–106.

Blei, David M, Andrew Y Ng, and Michael I Jordan (2003). "Latent dirichlet allocation". *Journal of machine Learning research* 3.Jan, pp. 993–1022.

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

Ghirelli, Corinna, Javier J Pérez, and Alberto Urtasun (2019). "A new economic policy uncertainty index for Spain". *Economics Letters*.

Griffiths, Thomas L and Mark Steyvers (2004). "Finding scientific topics". *Proceedings of the National academy of Sciences* 101.suppl 1, pp. 5228–5235.

Gulen, Huseyin and Mihai Ion (2015). "Policy uncertainty and corporate investment". *The Review of Financial Studies* 29.3, pp. 523–564.

Hansen, Stephen, Michael McMahon, and Andrea Prat (2017). "Transparency and deliberation within the FOMC: a computational linguistics approach". *The Quarterly Journal of Economics* 133.2, pp. 801–870.

Hassan, Tarek A, Stephan Hollander, Laurence van Lent, and Ahmed Tahoun (2019). "Firm-level political risk: Measurement and effects". *The Quarterly Journal of Economics* 134.4, pp. 2135–2202.

Hodson, Dermot (2017). "Eurozone Governance in 2016: The Italian Banking Crisis, Fiscal Flexibility and Brexit (Plus Plus Plus)". *JCMS: Journal of Common Market Studies* 55, pp. 118–132.

Hoffman, Matthew, Francis R Bach, and David M Blei (2010). "Online learning for latent dirichlet allocation". *Advances in neural information processing systems*, pp. 856–864.

Jens, Candace E (2017). "Political uncertainty and investment: Causal evidence from US gubernatorial elections". *Journal of Financial Economics* 124.3, pp. 563–579.

Meinen, Philipp and Oke Röhe (2017). "On measuring uncertainty and its impact on investment: Cross-country evidence from the euro area". *European Economic Review* 92, pp. 161–179.

Mikolov, Tomas, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean (2013). "Distributed representations of words and phrases and their compositionality". *Advances in neural information processing systems*, pp. 3111–3119.

Posner, Eric A and Adrian Vermeule (2009). "Crisis governance in the administrative state: 9/11 and the financial meltdown of 2008". *University of Chicago Law Review* 76, p. 1613.

Rehurek, Radim and Petr Sojka (2010). "Software Framework for Topic Modelling with Large Corpora". *IN PROCEEDINGS OF THE LREC 2010 WORKSHOP ON NEW CHALLENGES FOR NLP FRAMEWORKS*, pp. 45–50.

Saltzman, Bennett and Julieta Yung (2018). "A machine learning approach to identifying different types of uncertainty". *Economics Letters* 171, pp. 58–62.

Tobback, Ellen, Stefano Nardelli, and David Martens (2017). "Between hawks and doves: measuring central bank communication". *ECB Working Paper, No 2085.*

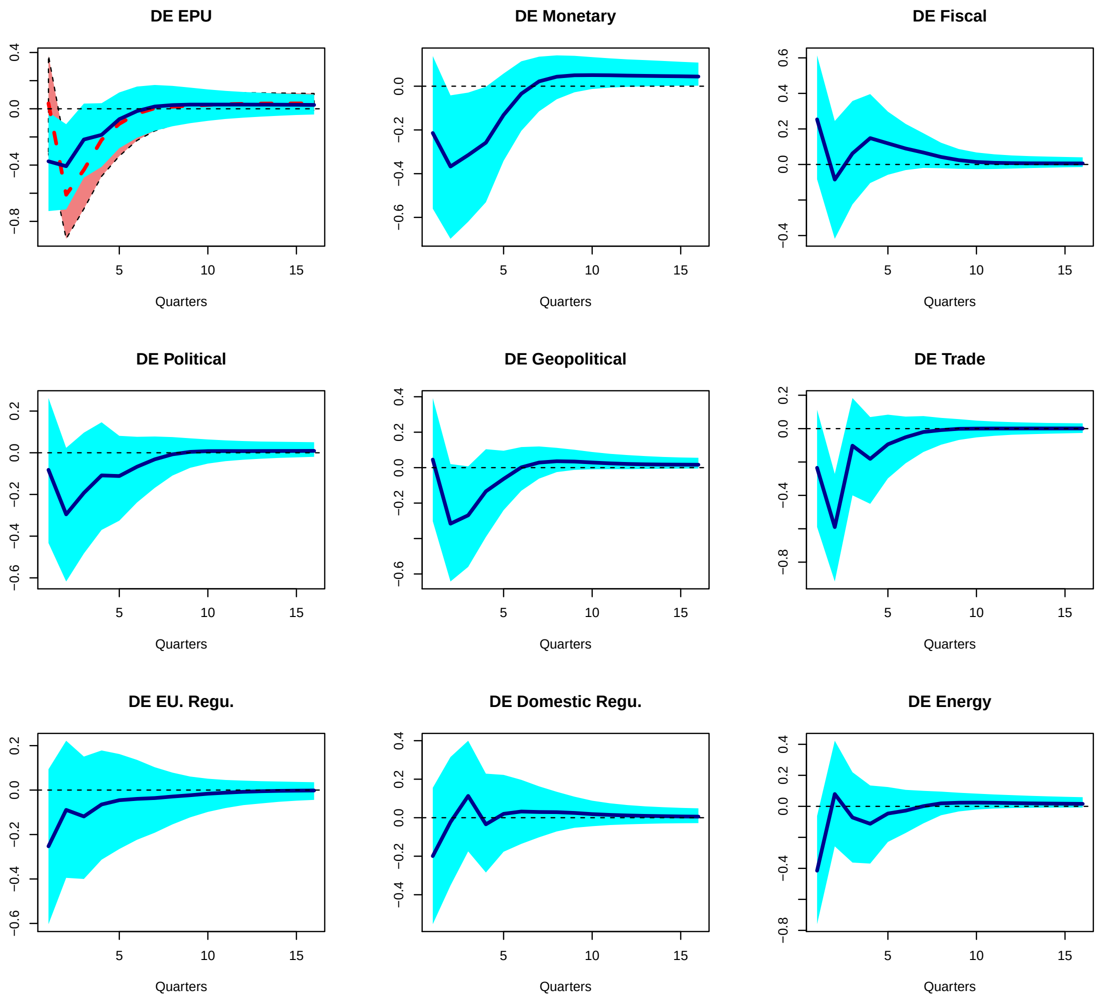
Whaley, Robert E (2000). "The investor fear gauge". *The Journal of Portfolio Management* 26.3, pp. 12–17.

Appendix A: word2vec results

- **wirtschaft:** (0.61) konjunktur; (0.59) volkswirtschaft; (0.56) ökonomie
- **unsicherheit:** (0.73) verunsicherung, (0.63) ungewissheit
- **économie:** (0.40) conjoncture
- **incertitude:** (0.53) flou, (0.52) inquiétud
- **economia:** (0.38) congiunturali
- **incertezza:** (0.56) instabilitá, (0.49) preoccupazione
- **economía:** (0.58) economico
- **incertidumbre:** (0.65) inquietud, (0.55) desconfianza

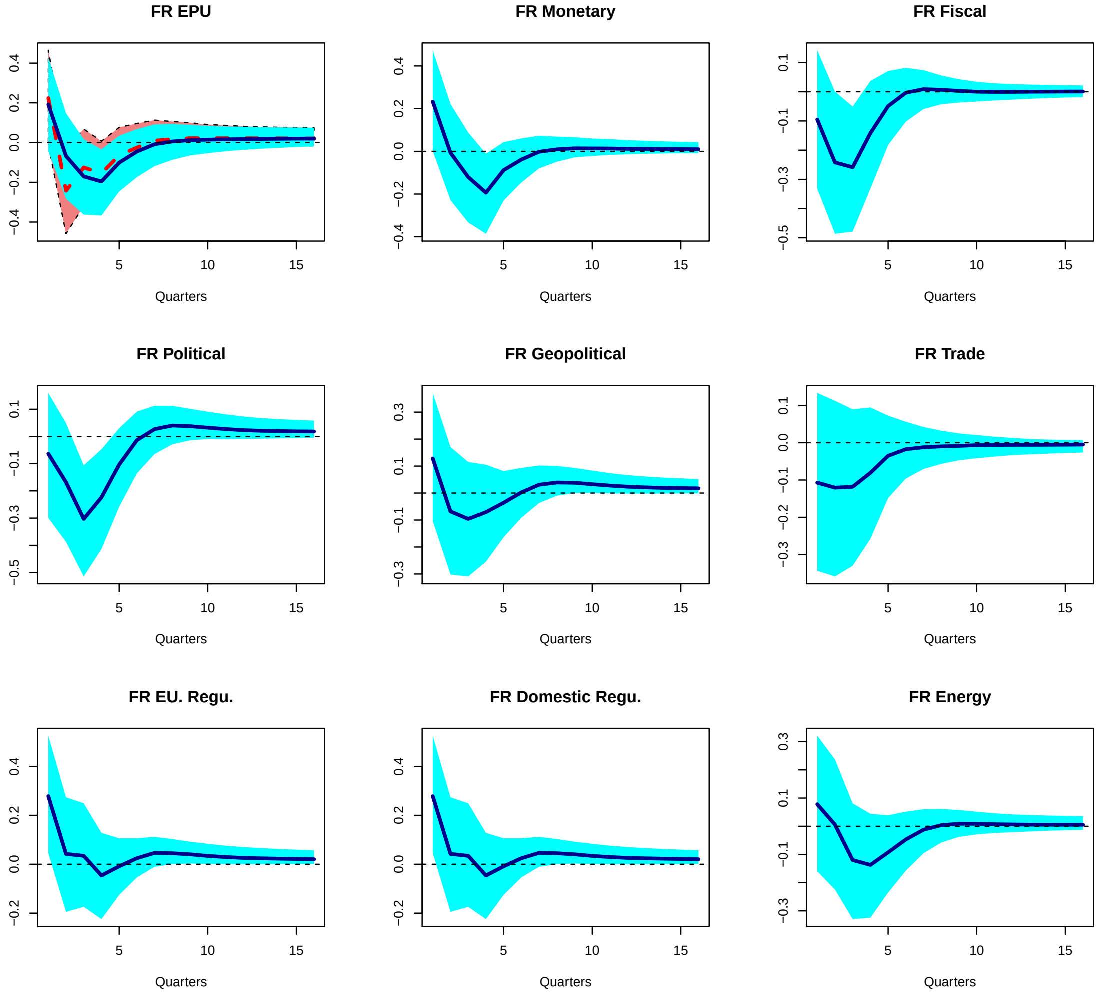
Appendix B: IRFs at country level

Figure 9: IRFs of investment in machinery and equipment to shocks in EPU and components for Germany



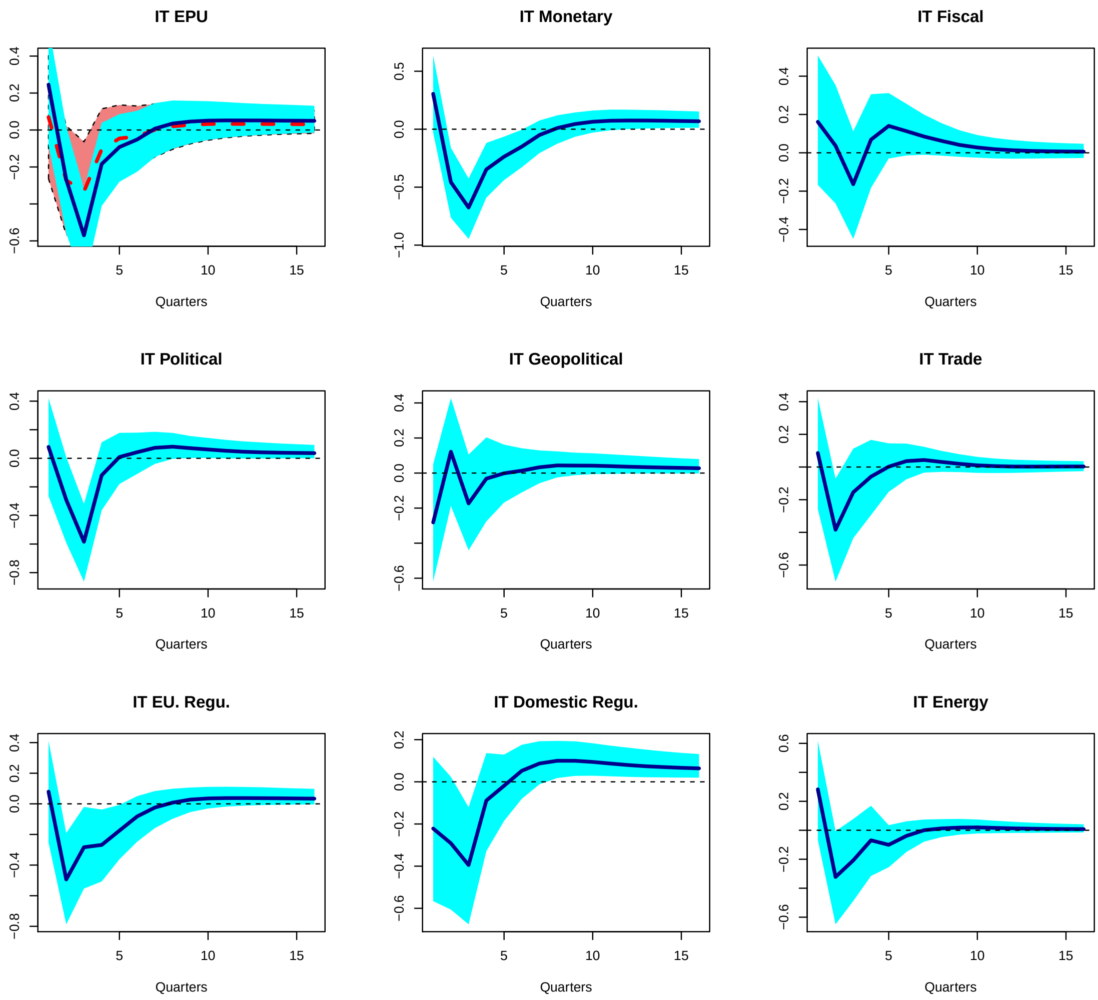
Notes: SVAR-estimated impulse response functions for machinery and equipment investment to a positive EPU shock (one standard deviation). The SVAR is estimated using Bayesian methods and the shocks are identified using the Cholesky decomposition with the variables in the following order: log(EPU), Δ(log(DAX stock price index)), shadow short rate (SSR), Δ(log(M&E)) and Δ(log(GDP)), where Δ indicates first differences or quarterly growth rates. Fit to quarterly data from Q1:2000 - Q1:2019. The blue bands represent the 68% confidence interval. The red line and bands represent the IRF computed using the Baker et al. (2016) aggregate EPU (BBD).

Figure 10: IRFs of investment in machinery and equipment to shocks in EPU and components for France



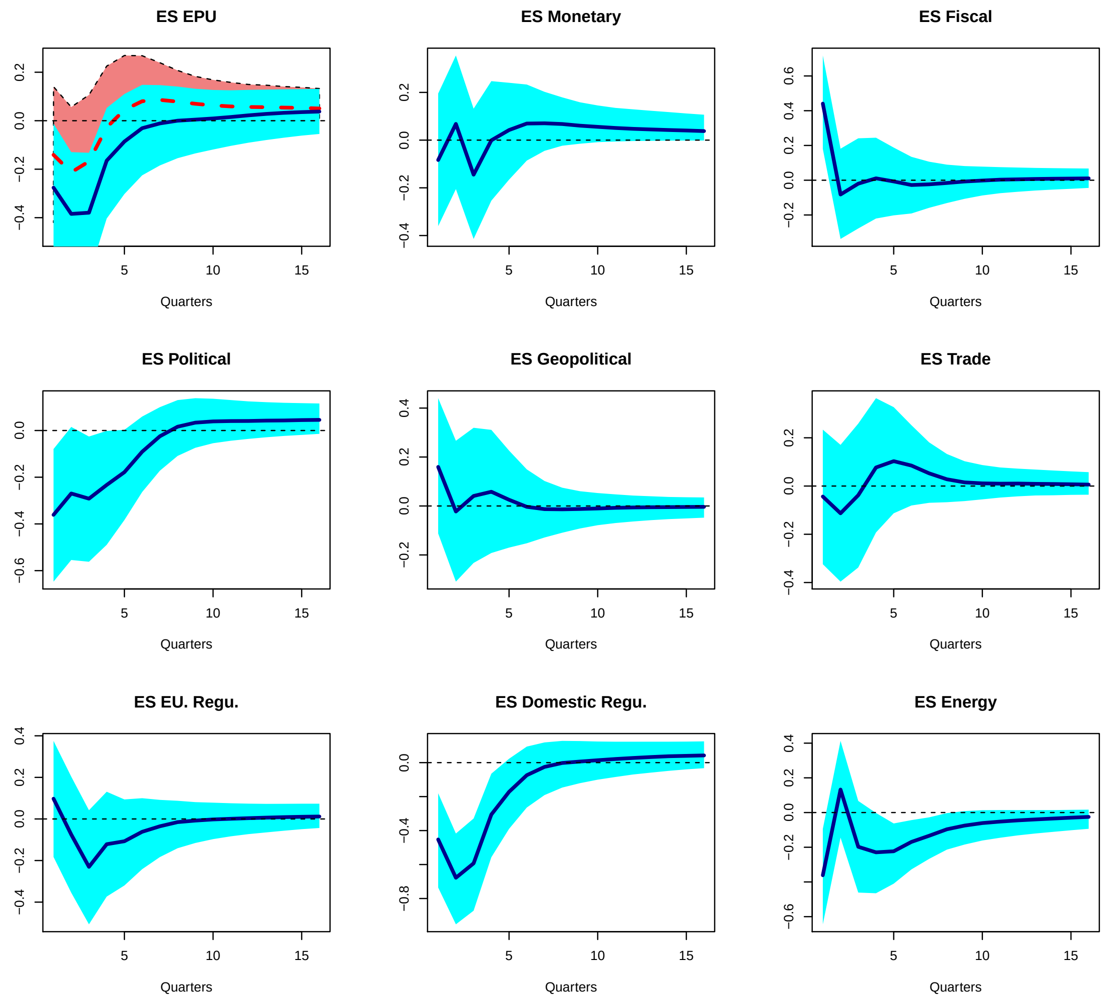
Notes: SVAR-estimated impulse response functions for machinery and equipment investment to a positive EPU shock (one standard deviation). The SVAR is estimated using Bayesian methods and the shocks are identified using the Cholesky decomposition with the variables in the following order: log(EPU), Δ(log(CAC40 stock price index)), shadow short rate (SSR), Δ(log(M&E)) and Δ(log(GDP)), where Δ indicates first differences or quarterly growth rates. Fit to quarterly data from Q1:2000 - Q1:2019. The blue bands represent the 68% confidence interval. The red line and bands represent the IRF computed using the Baker et al. (2016) aggregate EPU (BBD).

Figure 11: IRFs of investment in machinery and equipment to shocks in EPU and components for Italy



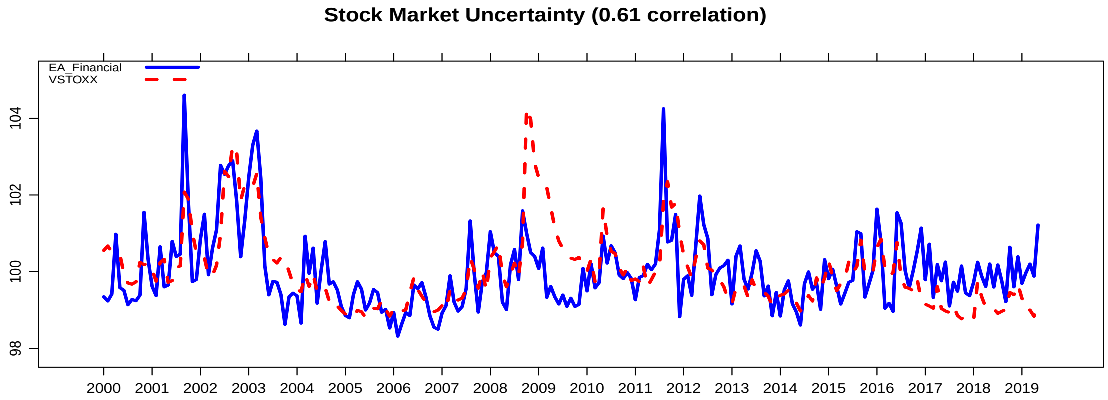
Notes: SVAR-estimated impulse response functions for machinery and equipment investment to a positive EPU shock (one standard deviation). The SVAR is estimated using Bayesian methods and the shocks are identified using the Cholesky decomposition with the variables in the following order: log(EPU), Δ(log(Italian stock price index)), shadow short rate (SSR), Δ(log(M&E)) and Δ(log(GDP)), where Δ indicates first differences or quarterly growth rates. Fit to quarterly data from Q1:2000 - Q1:2019. The blue bands represent the 68% confidence interval. The red line and bands represent the IRF computed using the Baker et al. (2016) aggregate EPU (BBD).

Figure 12: IRFs of investment in machinery and equipment to shocks in EPU and components for Spain

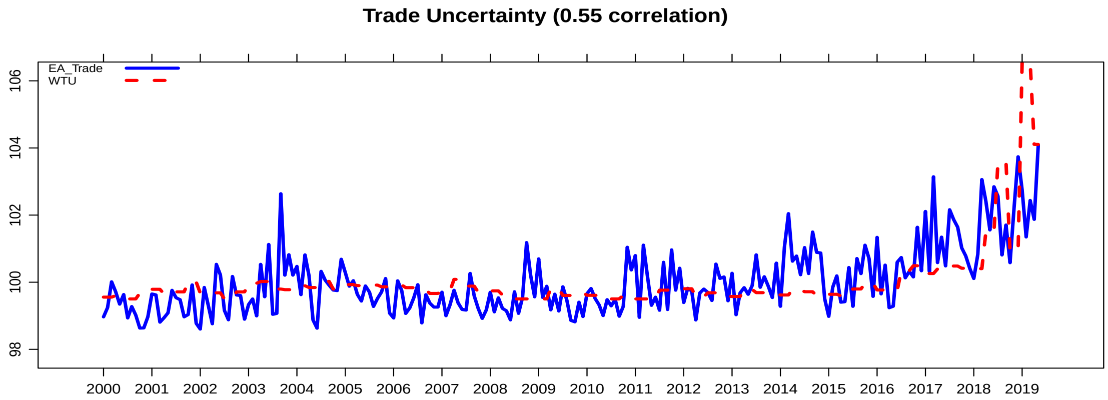


Notes: SVAR-estimated impulse response functions for machinery and equipment investment to a positive EPU shock (one standard deviation). The SVAR is estimated using Bayesian methods and the shocks are identified using the Cholesky decomposition with the variables in the following order: log(EPU), Δ(log(IBEX35 stock price index)), shadow short rate (SSR), Δ(log(M&E)) and Δ(log(GDP)), where Δ indicates first differences or quarterly growth rates. Fit to quarterly data from Q1:2000 - Q1:2019. The blue bands represent the 68% confidence interval. The red line and bands represent the IRF computed using the Baker et al. (2016) aggregate EPU (BBD).

Figure 13: Additional uncertainty indices



(a) Financial Uncertainty and VSTOXX



(b) Trade Uncertainty and WTU

Notes: For comparison purposes, all series are standardised to mean 100 and 1 standard deviation. Panel (a) compares the financial uncertainty index computed by aggregating those finance-related topics per country and the Eurostoxx implied volatility index (VTOXX). Panel (b) compares the trade uncertainty computed by aggregating those trade/industry-related topics and the world trade uncertainty index available at: https://www.policyuncertainty.com/wui_quarterly.html

Acknowledgements

We would like to thank Joao Sousa, Diego Rodriguez Palenzuela, Michele Lenza, Dimitris Korobilis, Federico Maria Ferrara, Thomas Gottron, Raluca Gui and participants of the Directorate General-Economics seminar at the ECB and the workshop on Big Data and Economic Forecasting organized by the European Commission for their helpful comments and suggestions. All remaining errors are our own.

Andrés Azqueta-Gavaldón
European Central Bank, Frankfurt am Main, Germany; email: andres.azqueta_gavaldon@ecb.europa.eu

Dominik Hirschbühl
European Central Bank, Frankfurt am Main, Germany; email: dominik.hirschbuhl@ecb.europa.eu

Luca Onorante
European Central Bank, Frankfurt am Main, Germany; email: luca.onorante@ecb.europa.eu

Lorena Saiz
European Central Bank, Frankfurt am Main, Germany; email: lorena.saiz@ecb.europa.eu

Postal address 60640 Frankfurt am Main, Germany
Telephone +49 69 1344 0
Website www.ecb.europa.eu

PDF ISBN 978-92-899-4002-3 ISSN 1725-2806 doi:10.2866/885526 QB-AR-20-011-EN-N