Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

Department Socioeconomics

Deciphering Professional Forecasters' Stories – Analyzing a Corpus of Textual Predictions for the German Economy

Ulrich Fritsche

Johannes Puckelwald

DEP (Socioeconomics) Discussion Papers
Macroeconomics and Finance Series
4/2018

Hamburg, 2018

Deciphering Professional Forecasters' Stories – Analyzing a Corpus of Textual Predictions for the German Economy*

Ulrich Fritsche[†]
Johannes Puckelwald[‡]

May 9, 2018

Abstract

We analyze a corpus of 564 business cycle forecast reports for the German economy. The dataset covers nine institutions and 27 years. From the entire reports we select the parts that refer exclusively to the forecast of the German economy. Sentiment and frequency analysis confirm that the mode of the textual expressions varies with the business cycle in line with the hypothesis of adaptive expectations. A calculated "uncertainty index" based on the occurrence of modal words matches with the economic policy uncertainty index by Baker et al. (2016). The latent Dirichlet allocation (LDA) model and the structural topic model (STM) indicate that topics are significantly state- and time-dependent and different across institutions. Positive or negative forecast "surprises" experienced in the previous year have an impact on the content of topics.

Keywords: Sentiment analysis, text analysis, uncertainty, business cycle forecast, forecast error, expectation, adaptive expectation, latent Dirichlet allocation, structural topic model

JEL classification: E32, E37, C49

*Acknowledgments: We thank Volker Daniel, Alexander Kriwoluzky, Christian Pierdzioch, Jörg Döpke, Karsten Müller and Christoph Behrens for helpful comments and Gregor Wiedemann and Andreas Niekler for generous support. We thank the *Deutsche Forschungsgemeinschaft* (DFG) for financial support (project "Macroeconomic Forecasting in Great Crisis" (FR 2677-4-1) within the DFG Priority Program 1859 "Experience and Expectation: Historical Foundations of Economic Behaviour").

†Corresponding author: Ulrich Fritsche, Universität Hamburg, Department Socioeconomics, KOF ETH Zurich, GWU Research Program on Forecasting, email: E-Mail: ulrich.fritsche@wiso.uni-hamburg.de

‡Universität Hamburg, Department Socioeconomics, Welckerstr. 8, 20354 Hamburg

1 Introduction

The past decade has brought an increasing interest in "text as data" for empirical research in economics, finance and accounting (Gentzkow et al., 2017; Loughran and McDonald, 2016). The availability of large – often unstructured and textual – datasets in the information age is indeed a methodological challenge for empirical research in social science and economics. Recent examples for the usage of "text as data" in economics cover topics such as central bank communication (Lüdering and Tillmann, 2016; Tobback et al., 2017; Jegadeesh and Wu, 2017), information economics and "media slant" (Gentzkow and Shapiro, 2010), policy uncertainty (Baker et al., 2016), forecasting financial market results (Tetlock, 2007; Tetlock et al., 2008; García, 2013), animal spirit models (Tuckett et al., 2014), forecasting inflation (Beckers et al., 2017), nowcasting GDP (Thorsrud, 2016) and other topics (see e.g. Gentzkow et al. (2017) for a very recent survey).

In our study, we will bring methods stemming from computer-based corpus linguistics to a corpus of textual representations of German macroeconomic forecasts. The central goal of the paper is to infer information about the formulated topics forming the expectations of forecasters. Our research is closely related to the following strands of literature: First, to the usage of "narrative" methods in macroeconomics (Romer and Romer, 2004, 2010) especially to the content analysis of textual representations of (mainly historical) business cycle expectations (Mathy and Stekler, 2017; Stekler and Symington, 2016; Lundquist and Stekler, 2012). We differ from that literature by using a larger corpus of German textual representations of future-related business cycle texts and by mainly using corpus-linguistic or distributional semantics (bag-of-words) methods instead of traditional qualitative content analysis tools. Second, there is another strand of literature in the field of political economy which analyzes the political positions of business cycle forecasters and the way political viewpoints may influence forecasts (Batchelor and Dua, 1990; Ngo et al., 2015). Our paper also adds to this literature. To the best of our knowledge, models of distributional semantics have not been applied in this field yet.

The data set consists of 564 texts containing professional forecasters' expectations of the German macroeconomic development of the current and the next calender year for 1990 to 2017. The documents were all published in publicly available outlets on a regular basis, with at least bi-annually frequency.

Given that our paper is a first attempt to identify topics contained in business cycle reports, we are aiming at addressing the following questions:

1. What can we infer from a corpus-based analysis of macroeconomic forecast texts regarding sentiment and uncertainty at the time the forecast was published?
2. Do discussed topics vary between boom and bust periods of publication?
3. Are all institutions alike in the way they phrase their expectations in topics?
4. Does the content of terms within topics change after having experienced (growth) forecast errors?

We will use frequency and dictionary-based sentiment analysis and analyze these measures over time. Furthermore, we will use different versions of distributional semantics models, namely the latent Dirichlet allocation (LDA) model (Blei et al., 2003) and structural topic model (STM, see (Roberts et al., 2016)).

The results of our investigation indicate some interesting aspects: First, dictionary-based sentiment analysis is a useful tool for analyzing forecast reports. There is a change of sentiment in the verbally expressed business cycle expectations over time which fits into the boom/bust chronology of the German economy. The result is surprising as the the reaction lags the business cycle. That points to adaptive expectation formation. Business cycle forecasters seem to shift to a more negative tone in textual descriptions of the expected path of the economy after slipping into a recession, and several years of upswing in a row improve the sentiment steadily. Furthermore, a constructed modal word uncertainty index delivers the theoretically expected negative correlation between uncertainty and the respective stance of the output gap in the cycle (Bloom, 2014).

Our estimated topic models reveal several results: First, unrestricted LDA analysis with uncorrelated topic structure point to some variation of topics over time as well as across boom and bust periods. A more subtle investigation using a structural topic (STM) model (Roberts et al., 2017, 2016) indicates the following: First, the bulk of topics seems to be non-distinctive with respect to the institutional differences after controlling for time effects. There are, however, few topics, for which “institution-specific” usage can clearly be identified. Time as a covariate for topic prevalence reveals significant effects of time on the topic usage. Furthermore, for some topics, we can find significant effects of the sign of the experienced past forecast errors on the word usage within a topic. This in turn indicates an experience-expectation nexus where past errors influence the tone in which the expectation on future economic development is phrased. This is an interesting finding which has to be investigated further in follow-up research.

Section 2 explains the methods used in the paper. In section 3 we describe the data set and 4 explains the findings. Section 5 puts the results in context and discusses the need for further research.[1]

2 Methodology

2.1 Frequency and Sentiment Analysis

As a first approach to analyze the content of our corpus, we use frequency analysis. This is a fundamental approach in text analytics where the absolute or relative occurrence of specific terms or concepts are counted and compared across texts. Differences over time or between authors can reveal changes and patterns of topics and key aspects, such as the prevailing sentiment, in forecast reports.

Using the occurrence of single terms in documents as a (crude) measure of meaning follows the *bag-of-word* hypothesis/ assumption in distributional semantics, claiming that words in a similar context refer to a similar meaning (see Turney and Pantel, 2010, p. 142f. and papers

[1]Detailed results for the topic models are available as an online appendix at `https://sites.google.com/site/ulrichfritsche/home/data-code`.

cited therein for the history of distributional analysis in linguistics). The order of words in texts is not relevant in this approach. While it appears to be a rough measure only, it is successfully utilized for economic research, for example by Baker et al. (2016) who constructed an economic policy uncertainty index, which builds on the appearance of few specific terms which might indicate a relation to the spheres of *economy*, *policy* and *uncertainty*.

Beyond looking at single terms we take a deeper look on the occurrence of specific subjects by using word lists. Such lists incorporate terms related to the field of interest and are used to calculate the frequency of appearance of any item of a list in texts. This approach allows for changes and variety in language. To be more precise, we use such lists to detect uncertainty in documents as well as their sentiment.

We utilize a new modality word list based on Tobback et al. (2016), which contains around 300 different terms expressing uncertainty.[2] We analyze the sentiment of documents by identifying the number of positive and negative coined terms used. This method puts terms into two lists according to their tone and is categorized as dictionary analysis (Stone et al., 1966). We make use of the *SentimentWortschatz*, a publicly available German-language word list for sentiment analysis (Remus et al., 2010). The used version (v1.8b) contains 1,650 positive and 1,818 negative words, which sum up to 15,649 positive and 15,632 negative word forms (e.g. *Verrat, Verrats, Verraten*). It also includes weights for the positive and negative terms on a scale from 0 to -1 for negative and 0 to 1 for positive words. The stronger the positive/ negative sentiment according to (Remus et al., 2010), the higher the absolute value of the weight. By calculating the balance between positive and negative words in the documents, we can analyze how the language in our corpus has become more negative or positive over time, and what differences can be found between institutions. This approach is related to other recent research on sentiment analyzes in connection with central bank communication and media perception (e.g. Beckers et al., 2017; Cannon, 2015; García, 2013).

2.2 Topic Models

2.2.1 Unsupervised Identification of Topics: Latent Dirichlet Allocation

In this section, we present the latent Dirichlet allocation (LDA, (Blei et al., 2003)) model to identify topics contained in documents. The LDA is a Bayesian mixed-membership model, a class of models where documents simultaneously belong to several topics and the distribution of topics may vary over documents (Grün and Hornik, 2011, p. 1).[3]

The idea behind LDA stems from a generative model of language/text formation and is related to the "bag-of-word" hypothesis in distributional semantics (see Turney and Pantel, 2010, p. 142f.).

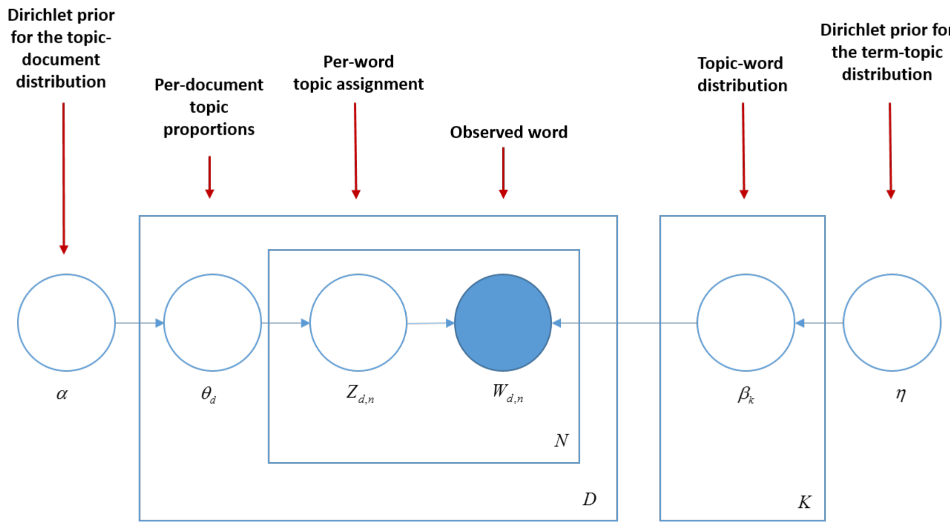
The model is built on the assumption that some (pre-defined) number of topics for the whole collection of documents exist which are in turn modeled as distributions over words. Each document is assumed to be generated in a certain order (Blei et al., 2003; Blei, 2012):

[2] A full word list is available in the online appendix at `https://sites.google.com/site/ulrichfritsche/home/data-code`

[3] To estimate the LDA we make use of the R package (R Core Team, 2018) topicmodels (Grün and Hornik, 2011).

First a distribution over the topics is chosen (β_k); then, for each word, a topic assignment is chosen (θ_d). At last, the word from the corresponding topic is chosen. Figure 1 gives a graphical representation.

Figure 1: Model Representation of LDA



Own representation based on Blei (2012, 2013)

More formally, D, N and K denote the number of documents, terms and topics, respectively. β_k denotes the distribution of terms in each of the K topics and θ_d the distribution of topics in each document. $Z_{d,n}$ denote the per-word topic assignment, $W_{d,n}$ denote the observed word(s). Both $\beta_k|\eta$ and $\theta_d|\alpha$ follow a Dirichlet distribution (with η as a Dirichlet prior for the term-topic distribution and α as Dirichlet prior for the topic-document distribution), $Z_{d,n|\theta_d}$ and $W_{d,n|Z_{d,n},\beta}$ follow a multinomial distribution. The complete model can then be formulated as follows (Grün and Hornik, 2011):

$$p\left(\theta, Z, W, \beta|\alpha, \eta\right) = \left(\prod_{i=1}^{K} p\left(\beta_i|\eta\right)\right)\left(\prod_{d=1}^{D} p\left(\theta_d|\alpha\right)\prod_{n=1}^{N} p\left(Z_{d,n}|\theta_d\right) p\left(W_{d,n}|Z_{d,n}, \beta\right)\right) \quad (1)$$

Equation (1) turns out to be analytically intractable (Grün and Hornik, 2011, p.4). Blei et al. (2003) proposed a variational inference method for the estimation, Griffiths and Steyvers (2004) suggest collapsed Gibbs sampling which has some computational advantages over variational inference (Heinrich, 2009).[4] During the inference process, researchers face a trade-off between two mutually exclusive goals: the estimation procedure could assign terms to as little as possible topics in one document (first goal) or assign high probability to as little as possible terms in one topic (second goal). This is mutually exclusive: As assigning all terms in one document to one topic only implies the missing of the second goal. Conversely, assigning high probability to few terms in each topic implies missing the first goal (Blei, 2013, p.17). A sensible selection of parameters should balance the trade-off. It has to be mentioned that identification of "topics" in LDA models is usually conditional on a given number of topics (K). Furthermore, priors should be selected in a way that

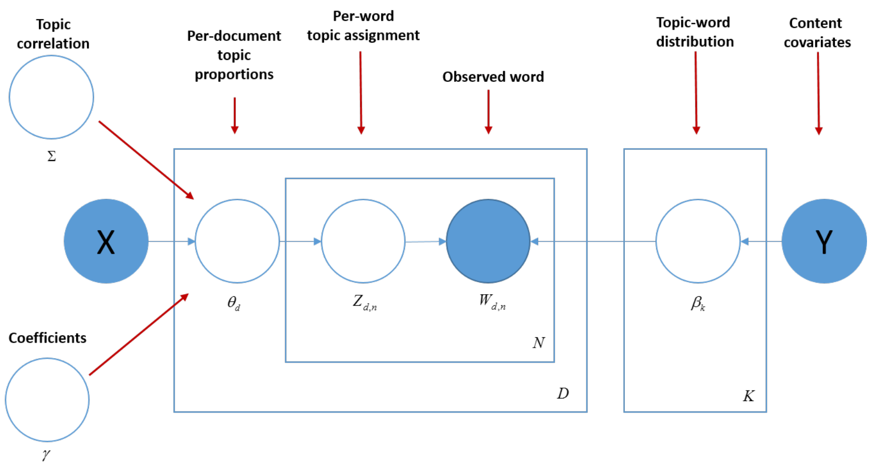
[4]For a detailed exposition of the solution procedure refer to Grün and Hornik (2011); Heinrich (2009).

the estimation results in semantically coherent groups of terms and interpretable topics in documents. Several criteria were proposed in the literature to deal with that issue. We used the criteria proposed by Arun et al. (2010), Cao et al. (2009), Deveaud et al. (2014) and Griffiths and Steyvers (2004) to determine the number of topics. For the Dirichlet prior α we apply a test procedure based on the perplexity of the model and on the topic coherence. The concept of so-called model perplexity was introduced by Blei et al. (2003). The main idea is to use a subset of the documents and omit them from the corpus. Then those "new" documents are presented to the model and the degree of "surprise" of the model confronted with the new data can be calculated. The lower the perplexity, the better the model captures semantic coherence. Topic coherence is based on the co-occurrence of the most frequent terms of topics in documents and was suggested as a measure by Mimno et al. (2011). They argue that it is more in line with human judgment of topic quality than the perplexity measure. Both measures are calculated for different values of α.

2.2.2 Identification of Topics using Covariates: Structural Topic Model

In practice, researchers often know more about a document than its word counts. From a statistical perspective it is worthwhile to include additional covariates and information into the estimation procedure. Among the many proposed extensions of the LDA (Blei and Lafferty, 2009; Eisenstein et al., 2011; Mimno and McCallum, 2012, e.g.), the structural topic model (STM) of Roberts et al., 2016 is of special interest in social science applications. A graphical (plate model) representation of the STM idea is shown in figure 2:

Figure 2: Model Representation of STM



Source: Own representation based on Roberts et al. (2016)

The fundamental structure of the model is based on the LDA model. However, two important aspects are different and give interesting opportunities for a more structural interpretation (Roberts et al., 2016, p. 989): First, the document-topic proportions θ_d are now modelled as a function of covariates and the assumption of orthogonality across topics is also abandoned. That allows for correlation between topics and a topic prevalence which depends on document-specific covariates. To model this, the Dirichlet distribution that controls θ_d has to be replaced by a logistic normal distribution. Furthermore, the β_k of the LDA is now decomposed on the document level into a baseline word distribution m, a topic specific deviation κ_k for each document, a covariate group deviation κ_g and an interaction term. The content-covariate part is embedded in the structure of the model using an exponential family model (as e.g. multinomial logistic model).

As with other models of this type, the exact posterior distribution is intractable and suffers from identification problems (Airoldi et al., 2014). Roberts et al. (2016) propose a solution that is based on a variational expectation-maximization algorithm and uses a Laplace approximation to solve the non-conjugacy problem between the logistic Normal and the multinomial distribution. Examples are given in Roberts et al. (2014) and Lucas et al. (2015).

For our analysis we use (basis (b-) spline functions of) time and institutional affiliation at the document level as covariates for topic prevalence and a factorial variable for positive / negative GDP forecast errors in the calender year previous to the publication date of the forecast as a covariate for the topic content.[5]

3 The Corpus

3.1 Incorporated documents

Our document corpus consists of business cycle forecasts reports for Germany from several institutions. They range from the German economic research institutes (e.g. DIW, ifo, RWI, IfW, ...), the "joint diagnosis" of the six/five "leading" research institutes to publications of Bundesbank, IW Köln and IMK[6] (see table 1 for the details).

Decisive for our choice is that most of these institutions have a long history of economic forecasting and publish forecasts on a bi-annual or higher frequency. These publications provide text data for our analysis with a suitable number of observations both over time and across institutions. We restrict the selection to forecasts published from 1990 onwards to ensure a sensible level of homogeneity in language usage.[7] In total, our corpus consists of 564 documents. The complete forecast reports usually cover more than just a forecasts for Germany but also comprise the description of the recent economic development, international forecasts, policy advice or methodological explanations. Texts were therefore manually fragmented and only the parts covering textual expressions of forecasts for the German macroeconomic development (future development of GDP, its main aggregates, of prices and employment) are used. However, we are not able to distinguish between different forecast horizons, as these parts of the content are frequently not separated in the texts. The documents usually cover outlooks for the ongoing as well as the next calendar year and less frequent for two years ahead. Newer publications of the considered institutions are usually available online. Where digitalized text files are not available, readable documents were obtained by *Optical Character Recognition* (OCR) finished scans of the original paper issues.

[5]STM model estimates and related visualizations were realized using the R package (R Core Team, 2018) stm (Roberts et al., 2017).

[6]IW and IMK are institutionally linked to employer's associations and trade unions, respectively.

[7]Pretests showed inferior topic coherence when older documents were included.

3.2 Pre-processing steps

We apply multiple pre-processing steps to the original texts to make the corpus suitable for the topic model procedures. The goal is to remove content from the documents which does not contain essential information for defining its topics. This has two main effects. Primarily, it enables us to identify meaningful topics, which could otherwise be shrouded. Additionally, it significantly reduces the volume of our data, which speeds up the computationally intensive procedures used.

First, words are removed which occur very frequently but have little meaning with regard to content. We use the word list, referred to as *stopwords*, included in the *tm* package for R. Second, additionally removed words are related to the specifics of the forecast reports. These are mainly names of the institutions and their publications, date related expressions and words related to the structure of tables and figures. A full list can be found in table 1. Finally, words with three or less letters are also deleted.[8]

Our analysis is not concerned with the structure of sentences. For topic models as for all *bag-of-words* approaches only the joint appearance of words in documents is crucial. Therefore, all punctuation is removed completely and all letters are transformed to lower cases. We also remove all numbers from texts as they are not needed for content analysis.

For all manually scanned documents, there is a risk of OCR problems, leading to errors in the identification of words and text structure. We try to fix some of these errors by additional pre-processing steps: At the first step, syllabicated words are reconnected. Then graphical characters and special whitespace are deleted which otherwise produce control characters in the text, leading to UTF-8 related errors.

The corpus is enriched by several document-level meta-data. Specifically, we use institution, date (month and year) of the publication, business cycle reference data from ECRI institute[9] and the forecast error of GDP growth rates in the calendar year preceding the publication date. Regarding the date, we refer to the time the forecast was actually completed if possible. When this information is not available, the publication date is used. A year is classified as a boom year when at least six months relate to a boom phase and as a bust year otherwise. We can use this information to analyze whether turning points in the business cycle are represented in the wording of the preceding forecasts.[10] The reference value to calculate the forecast error is the first publication of the GDP growth rate.[11] We decode the

[8]As a common practice the vocabulary is also often reduced by *stemming*. This reduces different grammatical forms of words to their common base form and thereby leads to smaller and more concise vocabulary. While stemming algorithms work fairly well for the English language, the results for German are not satisfying. We therefore refrain from using this form of preprocessing.

[9]`https://www.businesscycle.com/`

[10] We use the ECRI data in the version of march 2018. As the last data point marks a trough in January 2009, we assume an ongoing boom phase until the end of our time sample.

[11] We use the last forecast, usually made in December, while the first official GDP release is published in the German statistical office's publication series *Wirtschaft und Statistik*, usually in the January or February issue. So the the forecast error which is supposed to have an influence on the language of forecast texts in the year 2001 is the difference between the forecast for 2000 from December 1999 and the first official value for 2000 published at the beginning of 2001. However, some forecasts are published in the beginning of January. Those reports are usually finished in late December and we assume that no significant additional information is included compared to forecasts published in December in the previous year. Therefore we treat those forecasts as if they were published in the previous year when we calculate the forecast errors.

forecast error into a factor variable with level 2 for negative growth forecast errors ($\hat{y} < y$) and level 1 for positive growth forecast errors ($\hat{y} > y$).

4 Empirical Findings

4.1 Results of Frequency and Sentiment Analysis

We start the content analysis for the corpus with a frequency analysis per year over a couple of key terms related to recessions and uncertainty (see figure 3 and 4). To account for varying text lengths and numbers of documents per year, we look at the frequency of occurrence of the terms relative to the total sum of words per year. The light shaded areas in the figures refer to boom years, while darkly shaded years are classified as bust years.

There is a sharp rise in the use of the word "recession" in the bust periods in the early nineties as well as during the financial crises. For both periods the usage of this term peaks at a relative frequency of over 0.3 percent. In the bust phase in the early 2000-years the spike is much less pronounced. Here, however, the usage of "stagnation" is strongest over the whole sample. Noticeable, the term "crisis" strongly coincides with the EU sovereign debt crisis.

We can show that relevant aspects of textual content of business cycle forecasts can already be captured by looking at expressive key terms. The frequency distribution can be directly related to the stance of the business cycle. But it is also obvious, that the rise in frequencies is only pronounced during bust episodes and peaks at their ends. That means, the stronger appearance is presumably connected to a stronger perception of the current economic situation and not to the successful prediction of recessions. This is in line with the evidence from the forecast evaluation literature, which repeatedly finds, that forecasters missed almost all recession (see An et al. (2018) and literature cited therein).

For different terms related to uncertainty we see a less distinct variation over the business cycle. Only for "uncertainty" itself we observe a rise during the bust periods in the beginning of the 2000s and in 2008. What can be seen instead is a clear upward sloping trend in the usage of "uncertainty" and "risks" over time, since around the financial crises. Even though the business cycle has been in a up-swing phase since 2010, uncertainty clearly remained to be a crucial issue for forecasters, if not even became more important. Similar patterns are found with the economic policy uncertainty (EPU) index by Baker et al. (2016) and other uncertainty measures for Germany (e.g. Krebs and Yao (2016)).

When we are trying to identify the verbalization of uncertainty in our texts, looking at the frequency of single terms can be misleading when it is expressed in a versatile way. This is especially true for a more vague concept like "uncertainty". Therefore, we also apply a word list of uncertainty related terms to be able to take a deeper look into the subject.[12] Figure 5 shows the frequency of these terms relative to the total word count. We find a similar pattern as for uncertainty as single term but with a more distinct progression. There are now clear peaks in frequency during bust periods, followed by declines at the end of the

[12] For the full list refer to the online appendix at `https://sites.google.com/site/ulrichfritsche/home/data-code`.

downward cycle. The increased expression of uncertainty happens at the start of the bust episodes. Also we see the positive trend in the occurrence of uncertainty in the forecast texts over time, apparently starting at the beginning of the 2000s. This finding can be related to an environment which has become more complex and unclear, making forecasts more uncertain. On the other hand, forecasters might adapted to the critic in connection with large forecast errors, especially around the financial crises, by covering themselves via a more cautious wording.

When we compare the uncertainty measure seen in figure 5 with the development of the EPU index for Germany compiled by Baker et al. (2016), we see a similar pattern. The index also spikes during the two bust periods, has two additional sharp rises afterwards and shows a positive trend. Though, the peaks of the EPU index seem to be located one to two years earlier in several cases. This impression is supported by examining the cross-correlation between the uncertainty measures. Figure 7 shows the highest correlation of 0.73 between the frequency of uncertainty terms and the EPU index one year ahead, even though it is only slightly higher than the contemporaneous correlation of 0.71.

The second analysis using a word list aims at the sentiment of the forecast reports. Identifying more positive or more negative wording, respectively how the tone of texts changes, is an important insight for bringing together the quantitative and qualitative information on forecasts. As the sentiment is formed by diverse formulations, looking at a single word is not constructive in this regard. Figure 6 shows the results for the weighted and unweighted balance between the sum of positively and negatively classified words relative to the total word count.[13] The unweighted balance shows positive values, indicating an overall higher frequency of positive terms. When weights are included however, the balance is constantly in the negative area. This means that the negative terms used have a stronger sentiment in total.

The balance shows a clear pattern in regard to the business cycle. There is a decline in the year before and during bust phases with a recovery afterwards. Only for the year 1994 the tone of the forecast texts became more positive in the last year of a recession. The weighted sentiment frequency shows a more accentuated development, indicating that accounting for the nuances of the sentiment is advised. For the sentiment of the corpus, we see a similar pattern of the uncertainty measure at the beginning of a bust period. The institutions change their wording to a more negative sentiment in the year the business cycle starts to decline. But on the other side, in two out of three bust episodes the tone gets more positive only when the economy is already recovering. The level of uncertainty tends to decrease already in the last year of the bust.

The main result of the frequency analysis is, that both the sentiment and the uncertainty measures fail to act as a leading indicator for recessions. Forecasters do not seem to change their wording to a more negative or uncertain tone until the beginning of a downturn. In this regard, expectations of forecasters could be classified as more adaptive than anticipatory. On the other hand, the degree of uncertainty in the diction of forecasts seem to decrease in the first year after a bust phase. Of course, whether this change originates from the perception of a forthcoming upswing or the “stable” recession state is not yet clear. Compared with the

[13] The balance is calculated as the relative (weighted) sum of positive terms minus the sum of negative terms.

sentiment analysis, it can be stated that for the weighted frequency balance of positive and negative expressions only for the bust phase from 1991 to 1993 a change to a more positive language can be observed in the same pattern. For the other two bust periods the negative tendency carries on for an additional year. This is also in line with an adaptive behavior. Finally, the comparison with the EPU index also suggests a slower adaption to the economic environment. The EPU peaks are usually leading the ones of our the frequency measure. Forecasters follow the wording of newspapers rather than leading it.

If not as an early indicator, the frequency measures might be useful to identify the current state of the business cycle. After all, there are characteristic and fitting patterns for the boom and bust phases visible. However, there are also clear exceptions from this patterns. The increase in uncertainty between 2014 and 2017 is larger than those presumably related to the busts in the early 2000s and 2008. And for the weighted sentiment measure for the same time period and also for the years after 1994 there is a strong change to a more negative tone in the texts. In these cases there is no relation to recession. Hence, we need to be careful when using the measures for structural interpretations.

4.2 Results of LDA Analysis

After preprocessing our corpus as described, we estimate a LDA topic model before we turn to the extended structural topic model in the next section. Following Grün and Hornik (2011), we use Gibbs sampling with 1000 iteration steps for the estimation of the model after a burn-in phase of 1000 iterations. The first task is to find optimal values for the initial parameters for the number of topics (K), the prior for the topic-document distribution (α) and the prior for the term-topic distribution (η).

Regarding the search for an optimal K, figure 8 gives an overview for the four metrics used. We test a topic range from ten to 100 topics, what seems suitable for the size of the corpus. Even though the estimated optima are close to each other, there are some differences. The statistic by Deveaud et al. (2014) suggests a number around 40 while the ones from Arun et al. (2010) and Griffiths and Steyvers (2004) report values between 50 and 60 being optimal. The highest number of topics of around 70 is suggested by the metric by Cao et al. (2009). Evaluating the different values shows, that more topics do not lead to more well-defined topics but to more that are hard to interpret. Therefore, we estimate our topic model with $K = 40$ topics.

Figure 9 and 10 show the results for the search of an optimal α. It should produce a minimal model perplexity and a maximal topic coherence. Tests were estimated for different α values ranging from 0.01 to 2. The perplexity test suggests an α of 0.01 while the topic coherence has its maximum at 0.03. Since a lower α shifts more probability to single words within the topics, we use the lower value of 0.01 for our estimation. The prior η is set to 0.1, as suggested by Griffiths and Steyvers (2004).

As topic modeling is an unsupervised classification tool which does not have a structural interpretation at the first glance, we can only ex post rationalize the results. Some of the estimated topics are to some extent structurally interpretable. Figures 14 to 17 illustrate four topics via so called wordclouds. They visualize the topics as a cloud of the 50 most probable words defining the topic. The example topics show how the model, without any further meta-data input, identified recognizable economic topics from recent years. It can also be seen that some terms play a role for different topics. "migration of refugees" for example is one of the defining topics in our so called refugee topic. But it also is one of the most likely topics in the "UK topic", as it played an important role in this discussion on Brexit. The exercise illustrates how topics are not just related to single terms but to their joint appearance.

The topic distribution changes substantially when we divide the corpus according to the meta data. In figure 12 we look at the institution specific topic distribution. Even though one may think that all report on the same brought topic - the expectation about the business cycle over the next 1 to 2 years - it shows that the institutions might have a distinct style and focus in their texts. As a couple of topics are related to economic events, the topic distribution should also change over time. Figure 13 shows which topics were important over the years. In the early 90s topic 11 is especially important, which captures the subject of East and West Germany economy and the German reunification. It can also be seen how the share of the distribution of single topics tend to change smoothly over the years. This illustrates how topics become more or less important in the economic discussion and possibly also change in the wording and the style of forecast reports. On the other hand, there are also sudden events, like the financial crisis, represented for example by a sudden increase in the share of topic 36. Finally, the topic distribution also changes over the business cycle. In figure 11 we see a less distinct difference than between the years or institutions. Nevertheless, we see that in bust phases certain topics become less relevant, while others are more dominant. All in all, we can identify meaningful economic topics in forecast reports. Moreover, topics in the explanations of the business cycle forecasts vary with the economic circumstances.

In the next step, we want to test the influence of institutional affiliation and time on topic prevalence (θ) more rigorously using a structural topic model (Roberts et al., 2016). Furthermore, we investigate, to what extend experienced forecast errors influence the word-topic distribution (β).

4.3 Results of STM Analysis

The choice regarding the number of topics is essential in topic modelling (Grimmer and Stewart, 2013). In our case, according to the criteria proposed in Roberts et al. (2017) – namely held-out-likelihood (Wallach et al., 2009), residuals (Taddy, 2012) and lower bound – the model was estimated setting the number of topics to $k = 40$. This, furthermore, guarantees some kind of comparability with the LDA model specification. To check the appropriateness of the selected k, we calculated information criteria for a couple of models with different numbers of topics. For details refer to figure 18.

The STM model is estimated in the following specification: topic prevalence is modeled as a b-spline or basis spline function (Boor, 1993) of the covariates "time" and "institution" and topic content is estimated as a function of the sign of GDP growth forecast errors of the respective institution observed in the year before the forecast was made (see section 3 for details of the variable construction). We operationalize past "negative" or "positive" forecast surprises using this binary proxy. Since the covariate "past growth forecast" error was not available for all institutions at all times, the number of documents for the STM model is only

455.

We test for differences in topic prevalence with respect to the institutional affiliation. To avoid confusing results, we decided to present only results of a test of the respective institutions against the “joint diagnosis” of the five/six leading research institutes. The idea behind this test can be justified by assuming that the negotiation process of the “joint diagnosis” leads to a “consensual speech” and the respective institutions can be judged by testing if the prevalence of certain topics differs with respect to the so-defined “consensus”. The results are shown in figures 19 and 20.

The results imply that for the great majority of topics the prevalence does not differ with respect to institutions but that certain “institution-specific” topics can be identified. As an example for the Bundesbank, topic 24 seems to grasp a lot of the institution-specific differences in language usage of the Bundesbank relative to the joint diagnosis texts. This specific topic 24 does not show time variation in the prevalence but strong institution-specific dependence. In general, however, institution-specific differences do not play a large role for the topic prevalence. For the great majority of topics there are no significant differences in prevalence compared to the “joint diagnosis” texts.

Let us turn to some estimated topics. We select some topics to illustrate the identification of the structural model and argue that sensible results can be expected from the usage of this type of models for the detection of argumentation structures. Specifically, we select topics 16, 11, 20 and 36 which are related to interesting events of the last couple of years.

Topic 16 is a pre-crisis/ boom topic, related to the increase of the value-added tax in Germany in 2007 (see figure 23). The wordcloud (subfigure a) indicates that words like “value-added tax” on the one hand appear conjointly with “price increases”, “inflation” but also with terms indicating “dynamism”, “investment” and “boom”. The topic prevalence has a clear time-dependency (subfigure b). The topic appears around 2006/2007 – in line with the discussions of a value-added tax increase before and around January 1^{st} 2007. The topic content analysis (subfigure c) by covariate shows that the usage of terms within the topic is related to past forecast error experiences: in case of negative surprises (left side of subfigure c, terms in red), the terms “inflation” together with “dearness”, “consumer price index” and “prices” appear whereas under the experience of positive forecast surprises (right side, terms in blue) forecasters tend to use terms like “investment”, “boom”, “dynamism”, “employment” and “increase” appear.

Topics 11 and 20 are topics related to crisis, recession and demand-management policies (see figures 24 and 25). Topic 11 seems to precede topic 20 somewhat but both are clearly related to the recent financial crisis and the related business cycle drop thereafter (see subfigures b in both figures). The word-topic distribution in both cases shows a systematic relationship to experienced recent growth forecast errors. In case of “negative surprises” terms like “recession”, “banks” (topic 11) or “short-time work”, “decline”, “firms” and “considerable” (topic 20) show up, in case of positive surprises, we find terms like “recovery”, “increase”, “expenditures” or “labor market” together with “expect” and “production”.

Topic 36 is related to the “refugee” debate in Germany (see figure 26). The wordcloud (subfigure a) contains terms like “refugee migration”, “refugee”, “expenditures”, “calculations” which indicates that the economic consequences of the “refugee crisis” as a common

German narrative framed this historical episode after 2015 are discussed in the texts. The effect of time on topic prevalence confirm that the topic is indeed strongly related to the last years of the sample (subfigure b). Interestingly, negative forecast surprises are connected with terms as "uncertainty", "expenditure" whereas positive forecast surprises is connected with more "technical" terms as "calculations", "change" and modal words as "may". This indicates that under the impression of a positive forecast surprise, there is a higher tendency to interpret the consequences of the "refugee migration" in a more rosy way, whereas under negative surprises there are more concerns and uncertainties.

Our results can only give some hints about the argument structures in forming expectation topics. Nevertheless, it is interesting that the STM model delivers reasonable and to some extent economically interpretable results for a corpus of documents describing business cycle expectations. The advantage of the STM approach in contrast to descriptive text-mining statistics clearly lies in the possibility to test the significance of covariates impacting the topic prevalence and topic content. In our model, we find significant time effects on topic prevalence, some significant institutional effects on prevalence and topic content effects for selected topics of the sign of the past growth forecast.

5 Discussion

This paper brings together a corpus of forecast report texts for the German economy from 1990 to 2017 from ten different research institutions with modern text analysis procedures from computer-based corpus linguistics. First, we use frequency analysis for single terms of interest as well as for world lists from special dictionaries to investigate content, sentiment and the presence of uncertainty in the corpus. Second, we estimate latent Dirichlet allocation (Blei et al., 2003) and structural topic models (Roberts et al., 2016) to discover the underlying topics of the documents and how they vary with regard to several factors. Above all we are interested in how topics differ over time, across institutions and with respect to the forecast performance.

Our results show that the applied methods give economically meaningful results and are, thus, indeed useful and help to structure and analyze forecast reports with regard to the interpretation of the economic situation. Single term and dictionary-based frequency analysis show patterns which fit the German business cycle chronology since the 1990s. The sentiment of the texts becomes more negative in bust phases and turns more positive afterwards. Also, terms indicating uncertainty tend to be used more often when the economy is experiencing a bust. These patterns suggest that the frequency measures are suitable to be applied to evaluate the economic situation. But it is also shown that caution is advised when doing so. The forecasters seem to adjust the wording for their outlooks not in a forward-looking way with regard to the business cycle but adaptively. This is true for both sentiment as well as for uncertainty measures. For the latter we see that the economic policy uncertainty index by Baker et al. (2016) shows a comparable but shifted pattern (and high correlation) to the frequency measure. Furthermore, the up- and down-swings in sentiment and uncertainty are not exclusively related to the business cycle. Especially after the financial crises, negative and uncertainty terms occasionally went up even though Germany experienced an ongoing upswing.

Also, the topic model approach proofs to be suited for our corpus. Already for the LDA model, without any initial input on the content of the documents, we see that topics are identified with a clear link to certain economic events or forecast themes. While these topics are precisely defined, their occurrence varies with respect to the business cycle, the historic development and between institutions. They capture political and economic events which change over the years. Finally, there appear to be differences between the main focus of institutions when discussing the economic development in Germany.

While the LDA model is more descriptive, the structural topic model enables us to use the information about the background of the forecasts as additional data in the estimation of the topic model and to test the significance of differences in the resulting topic distributions. We find that most of the topics show no institution specific usage after controlling for time effects. This is to some extend expected, as several general topics should appear in every economic forecast. However, some topics are significantly over- or underused by single institutions. These topics are of special interest, as they can reveal special interest or political agendas. Putting these differences in context with the theoretical and political orientation of institutions is a field for future research.

The importance of topics over time also have been tested formally in the STM. For example, the topic describing the financial crisis is not significantly used aside from 2008/2009. For "event topics" this test is of interest to evaluate whether forecasters judge a specific event as important or not.

Finally, for some topics we find effects of the sign of experienced past forecast errors on the word usage within a topic. This, in turn, indicates an experience-expectation nexus which has to be investigated further in follow-up research. Specifically, a larger corpus with more meta-data might help to investigate deeper what factors influence the usage of certain arguments in the texts.

This is a first attempt to use distributional semantics technologies for the texts produced by professional forecasters. There is still a way to go. We can, however, show that the applied methods can provide useful information for future research on the field of forecast evaluation. Besides quantitative forecasts it is becoming more and more possible to access their qualitative background. Bringing together the two spheres can provide further information for the appraising of the more subtle tendencies and uncertainties related to the latest forecast numbers.

References

Airoldi, Edoardo M., David Blei, Elena A. Erosheva, and Stephen E. Fienberg, *Handbook of Mixed Membership Models and Their Applications*, 1st ed., Chapman & Hall/CRC, 2014.

An, Zidong, João Tovar Jalles, and Prakash Loungani, "How Well Do Economists Forecast Recessions?," *IMF Working Paper*, 2018, (18/39).

Arun, R., V. Suresh, C. E. Veni Madhavan, and M. N. Narasimha Murthy, "On Finding the Natural Number of Topics with Latent Dirichlet Allocation: Some Observations," in Mohammed J. Zaki, Jeffrey Xu Yu, B. Ravindran, and Vikram Pudi, eds., *Advances in Knowledge Discovery and Data Mining*, Springer Berlin Heidelberg Berlin, Heidelberg 2010, pp. 391–402.

Baker, Scott, Nicholas Bloom, and Steven Davis, "Measuring Economic Policy Uncertainty," *The Quarterly Journal of Economics*, 2016, *131* (4), 1593–1636.

Batchelor, Roy and Pami Dua, "Forecaster Ideology, Forecasting Technique, and the Accuracy of Economic Forecasts," *International Journal of Forecasting*, 1990, *6* (1), 3–10.

Beckers, Benjamin, Konstantin Kholodilin, and Dirk Ulbricht, "Reading between the Lines: Using Media to Improve German Inflation Forecasts," Discussion Papers of DIW Berlin 1665, DIW Berlin, German Institute for Economic Research 2017.

Blei, David M., "Probabilistic Topic Models," *Commun. ACM*, April 2012, *55* (4), 77–84.

— , "Probabilistic Topic Models: Origins and Challenges," 2013 Topic Modeling Workshop at NIPS December 2013.

— **and John D. Lafferty**, *Text Mining: Classification, Clustering, and Applications*, Chapman and Hall/CRC Press,

— **, Andrew Y. Ng, and Michael I. Jordan**, "Latent Dirichlet Allocation," *Journal of Machine Learning Research*, 2003, *3*, 993 – 1022.

Bloom, Nicholas, "Fluctuations in Uncertainty," *Journal of Economic Perspectives*, 2014, *28* (2), 153–76.

Boor, Carl De, "B (asic)-Spline Basics," in Les Piegl, ed., *Fundamental Developments of Computer-Aided Geometric Modeling*, Academic Press, 1993, pp. 27–49.

Cannon, San, "Sentiment of the FOMC: Unscripted," *Economic Review-Federal Reserve Bank of Kansas City*, 2015, p. 5.

Cao, Juan, Tian Xia, Jintao Li, Yongdong Zhang, and Sheng Tang, "A Density-based Method for Adaptive LDA Model Selection," *Neurocomputing*, 2009, *72* (7), 1775 – 1781. Advances in Machine Learning and Computational Intelligence.

Deveaud, Romain, Éric SanJuan, and Patrice Bellot, "Accurate and Effective Latent Concept Modeling for ad hoc Information Retrieval.," *Document numérique*, 2014, *17* (1), 61–84.

Eisenstein, Jacob, Amr Ahmed, and Eric P. Xing, "Sparse Additive Generative Models of Text," in "Proceedings of the 28th International Conference on International Conference on Machine Learning" ICML'11 Omnipress USA 2011, pp. 1041–1048.

García, Diego, "Sentiment during Recessions," *Journal of Finance*, 2013, *68* (3), 1267–1300.

Gentzkow, Matthew and Jesse Shapiro, "What Drives Media Slant? Evidence From U.S. Daily Newspapers," *Econometrica*, 2010, *78* (1), 35–71.

— **, Bryan T. Kelly, and Matt Taddy**, "Text as Data," NBER Working Papers 23276, National Bureau of Economic Research, Inc 2017.

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

Grimmer, Justin and Brandon M. Stewart, "Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts," *Political Analysis*, 2013, *21* (3), 267–297.

Grün, Bettina and Kurt Hornik, "topicmodels: An R Package for Fitting Topic Models," *Journal of Statistical Software, Articles*, 2011, *40* (13), 1–30.

Heinrich, Gregor, "Parameter Estimation for Text Analysis," Technical report, Fraunhofer IGD 2009.

Jegadeesh, Narasimhan and Di (Andrew) Wu, "Deciphering Fedspeak: The Information Content of FOMC Meetings," Technical Report, SSRN 2017.

Krebs, Tom and Yao Yao, "Labor Market Risk in Germany," *IZA Discussion Paper Series*, 2016, (9869).

Lüdering, Jochen and Peter Tillmann, "Monetary Policy on Twitter and its Effect on Asset Prices: Evidence from Computational Text Analysis," MAGKS Papers on Economics, Philipps-Universität Marburg, Faculty of Business Administration and Economics, Department of Economics (Volkswirtschaftliche Abteilung) 2016.

Loughran, Tim and Bill McDonald, "Textual Analysis in Accounting and Finance: A Survey," *Journal of Accounting Research*, 2016, *54* (4), 1187–1230.

Lucas, Christopher, Richard A. Nielsen, Margaret E. Roberts, Brandon M. Stewart, Alex Storer, and Dustin Tingley, "Computer-Assisted Text Analysis for Comparative Politics," *Political Analysis*, 2015, *23* (2), 254–277.

Lundquist, Kathryn and Herman Stekler, "Interpreting the Performance of Business Economists During the Great Recession," *Business Economics*, 2012, *47* (2), 148–154.

Mathy, Gabriel and Herman Stekler, "Expectations and Forecasting during the Great Depression: Real-time Evidence from the Business Press," *Journal of Macroeconomics*, 2017, *53* (C), 1–15.

Mimno, David, Hanna M Wallach, Edmund Talley, Miriam Leenders, and Andrew McCallum, "Optimizing Semantic Coherence in Topic Models," in "Proceedings of the Conference on Empirical Methods in Natural Language Processing" Association for Computational Linguistics 2011, pp. 262–272.

Mimno, David M. and Andrew McCallum, "Topic Models Conditioned on Arbitrary Features with Dirichlet-multinomial Regression," *CoRR*, 2012, *abs/1206.3278.*

Ngo, Ha Quyen, Niklas Potrafke, Marina Riem, and Christoph Schinke, "Ideology and Dissent among Economists: The Joint Economic Forecast of German Economic Research Institutes," CESifo Working Paper Series 5393, CESifo Group Munich 2015.

R Core Team, *R: A Language and Environment for Statistical Computing* R Foundation for Statistical Computing, version 3.4.3 ed. 2018.

Remus, R., U. Quasthoff, and G. Heyer, "SentiWS – A Publicly Available German-language Resource for Sentiment Analysis," in "Proceedings of the 7th International Language Resources and Evaluation (LREC'10)" 2010.

Roberts, Margaret E., Brandon M. Stewart, and Dustin Tingley, *stm: R Package for Structural Topic Models* 2017. R package version 1.3.0.

__ , __ , and Edoardo M. Airoldi, "A Model of Text for Experimentation in the Social Sciences," *Journal of the American Statistical Association*, 2016, *111* (515), 988–1003.

__ , __ , Dustin Tingley, Christopher Lucas, Jetson Leder-Luis, Shana Kushner Gadarian, Bethany Albertson, and David G. Rand, "Structural Topic Models for Open-Ended Survey Responses," *American Journal of Political Science*, 2014, *58* (4), 1064–1082.

Romer, Christina and David Romer, "A New Measure of Monetary Shocks: Derivation and Implications," *American Economic Review*, 2004, *94* (4), 1055–1084.

__ and __ , "The Macroeconomic Effects of Tax Changes: Estimates Based on a New Measure of Fiscal Shocks," *American Economic Review*, 2010, *100* (3), 763–801.

Stekler, Herman and Hilary Symington, "Evaluating Qualitative Forecasts: The FOMC Minutes, 2006–2010," *International Journal of Forecasting*, 2016, *32* (2), 559–570.

Stone, Philip J, Dexter C Dunphy, and Marshall S Smith, "The General Inquirer: A Computer Approach to Content Analysis.," 1966.

Taddy, Matt, "On Estimation and Selection for Topic Models," in Neil D. Lawrence and Mark Girolami, eds., *Proceedings of the Fifteenth International Conference on Artificial Intelligence and Statistics*, Vol. 22 of *Proceedings of Machine Learning Research* PMLR La Palma, Canary Islands 21–23 Apr 2012, pp. 1184–1193.

Tetlock, Paul C., "Giving Content to Investor Sentiment: The Role of Media in the Stock Market," *Journal of Finance*, 2007, *62* (3), 1139–1168.

__ , Maytal Saar-Tsechansky, and Sofus Macskassy, "More Than Words: Quantifying Language to Measure Firms' Fundamentals," *Journal of Finance*, 2008, *63* (3), 1437–1467.

Thorsrud, Leif, "Nowcasting using News Topics. Big Data versus Big Bank," Working Paper 2016/20, Norges Bank 2016.

Tobback, Ellen, Hans Naudts, Walter Daelemans, Enric Junqué de Fortuny, and David Martens, "Belgian Economic Policy Uncertainty Index: Improvement through Text Mining," *International Journal of Forecasting*, 2016, *34* (2), 355 – 365.

__ , Stefano Nardelli, and David Martens, "Between Hawks and Doves: Measuring Central Bank Communication," Working Paper Series 2085, European Central Bank 2017.

Tuckett, David, Paul Ormerod, Robert Elliott Smith, and Rickard Nyman, "Bringing Social-Psychological Variables into Economic Modelling: Uncertainty, Animal Spirits and the Recovery from the Great Recession," Technical Report 2014.

Turney, Peter D. and Patrick Pantel, "From Frequency to Meaning: Vector Space Models of Semantics," *J. Artif. Int. Res.*, January 2010, *37* (1), 141–188.

Wallach, Hanna M., Iain Murray, Ruslan Salakhutdinov, and David Mimno, "Evaluation Methods for Topic Models," in "Proceedings of the 26th Annual International Conference on Machine Learning" ICML '09 ACM New York, NY, USA 2009, pp. 1105–1112.

Appendix: Tables and Figures

Table 1: List of included institutions and publications

Institution	Dates	Publication schedule	Source
German central bank (Deutsche Bundesbank)	2007 - 2016	bi-annually	Monatsberichte der Deutschen Bundesbank
German Institute of Economic Research (DIW Berlin)	1990 - 2016	bi-annually (1990-2004), quarterly (2005-2016)	Wochenbericht des DIW
Joint forecast of the German economic research institutes (Gemeinschaftsdiagnose)	1990 - 2016	bi-annually	various publications
Hamburg Institute of International Economics (HWWI) (HWWA until 2006)	1990 - 2017	bi-annually (1990-2006), quarterly (2005-2017)	Konjunktur von morgen (1990-1996), Wirtschaftsdienst (1997-2016), occasionally press releases from 2015
ifo Institute Munich	1990 - 2016	bi-annually	Monatsberichte des ifo Instituts für Wirtschaftsforschung (1990-2000), ifo Schnelldienst (2001-2016)
Kiel Institute for World Economics (IfW)	1990 - 2017	bi-annually (1990-1991), quarterly (1992-2017)	Die Weltwirtschaft (1990-2005), Kieler Diskussionsbeiträge (2006-2014), Kieler Konjunkturberichte (2015-2017)
Macroeconomic Policy Institute (IMK)	2005 - 2016	quarterly	IMK Report
German Economic Institute (IW Köln)	1995 - 2016	annually (1995-2006), bi-annually (2007-2016)	IW Trends
Institute for Economic Research Halle (IWH)	1997 - 2017	bi-annually (1997-2000), quarterly (2001-2017)	Wirtschaft im Wandel (1997-2012), Konjunktur aktuell (2013-2017), occasionally press releases
Rhine-Westphalia Institute for Economic Research (RWI)	2006 - 2017	bi-annually (2006-2012), quarterly (2013-2017)	RWI-Konjunkturberichte

Figure 3: Relative frequency of recession terms

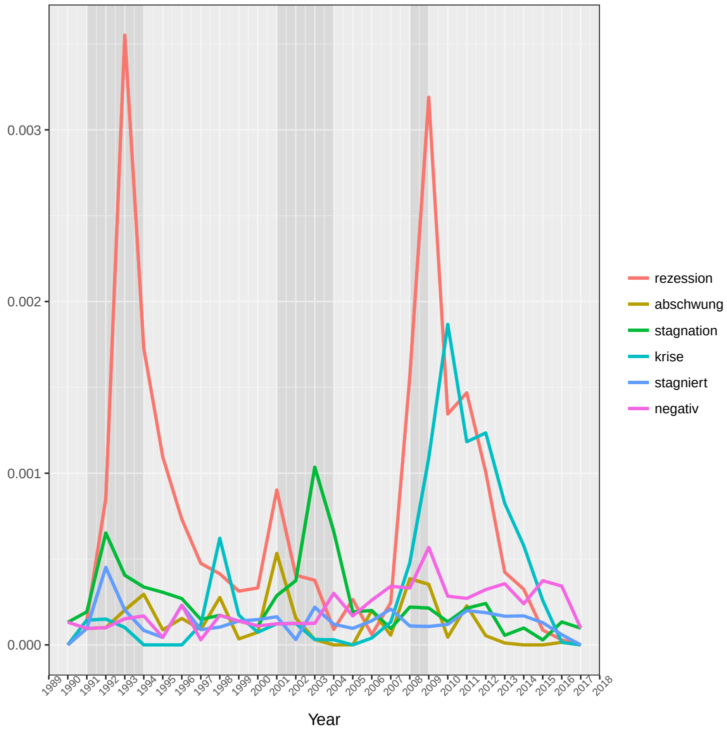


Figure 4: Relative frequency of uncertainty terms

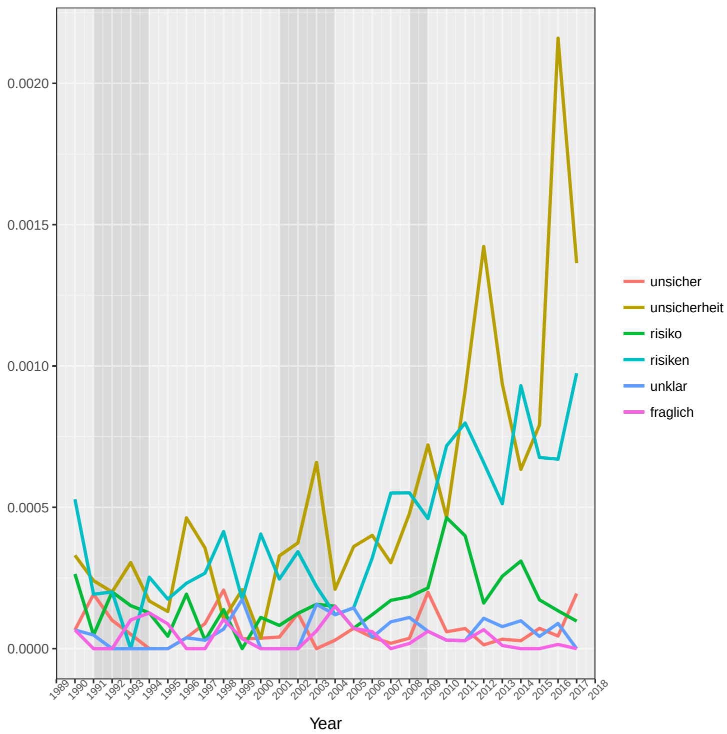


Figure 5: Relative frequency of uncertainty terms and economic policy uncertainty index by Baker et al. (2016)

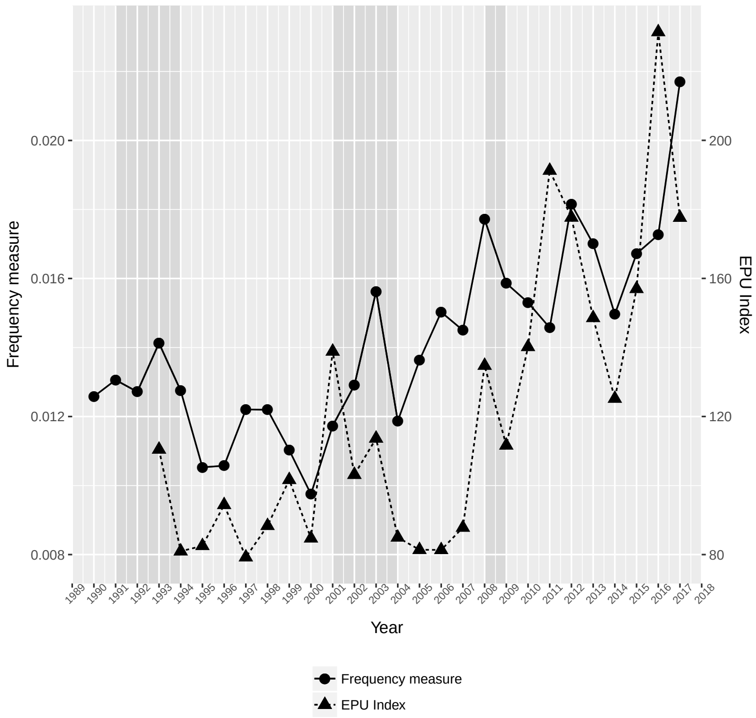


Figure 6: Balance of relative frequency of positive and negative terms

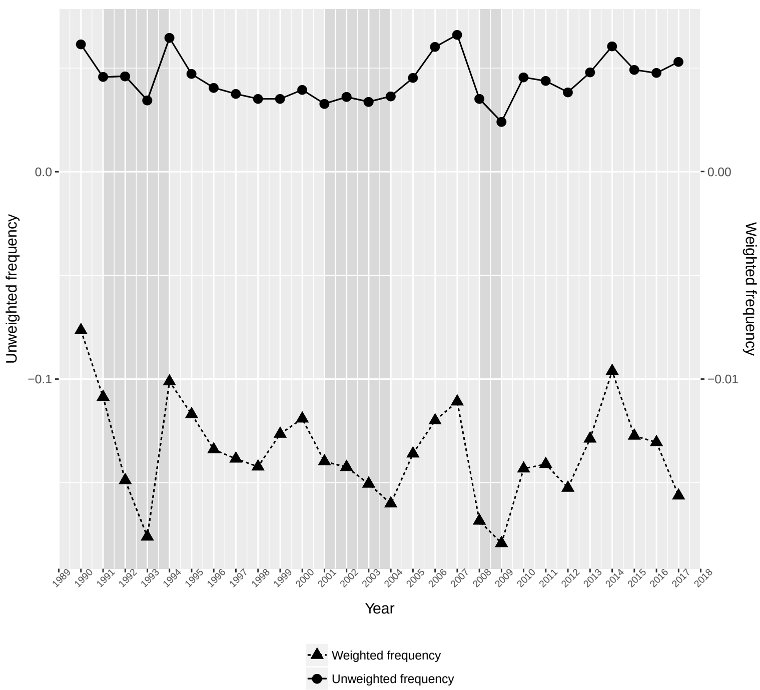


Figure 7: Cross-correlation between the frequency of uncertainty terms and leads and lags of the economic policy uncertainty index

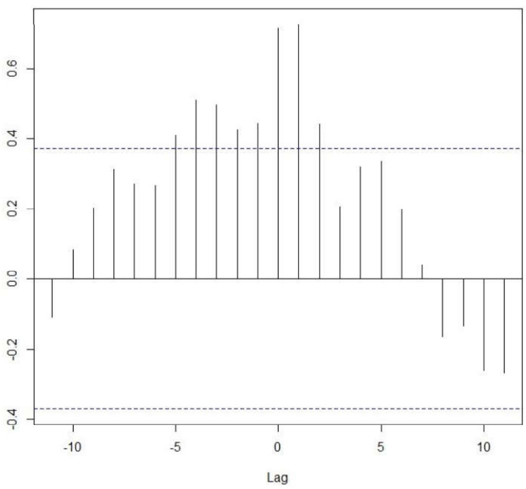


Figure 8: Results of determining K, “FindTopicsNumber” function (ldatuning package)

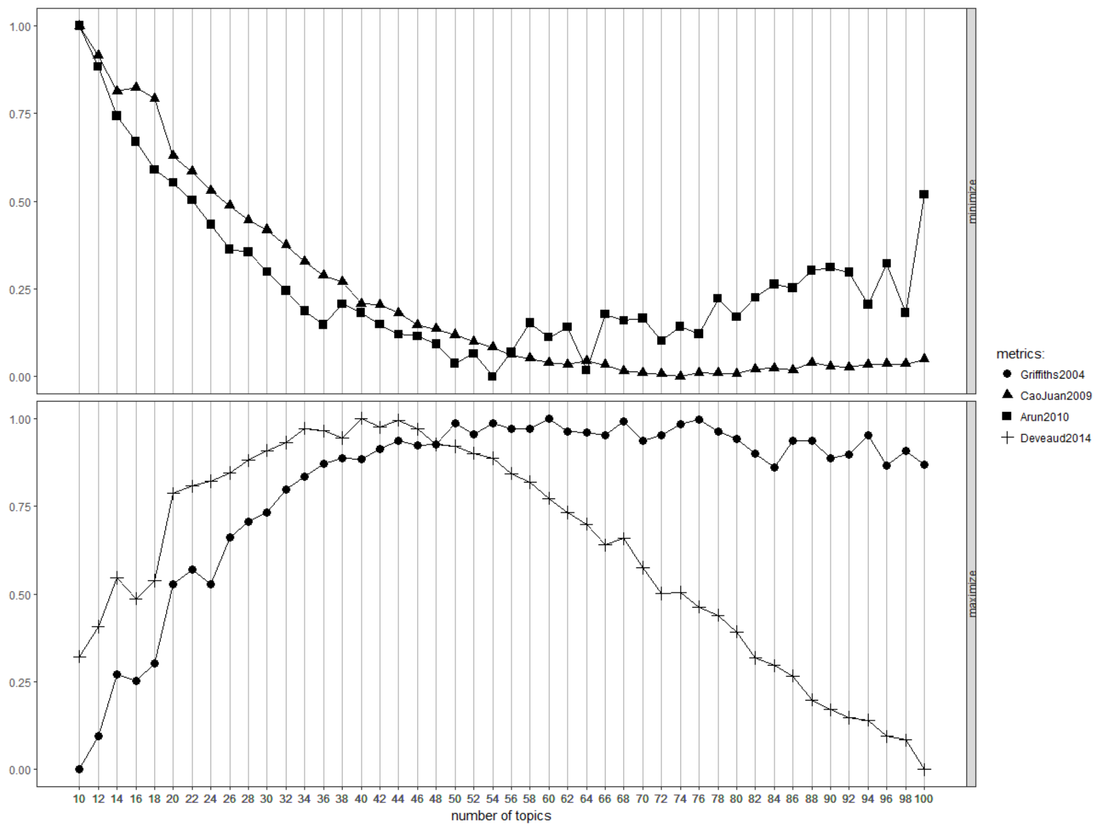


Figure 9: Model perplexity for different α, “perplexity” function (topicmodels package)

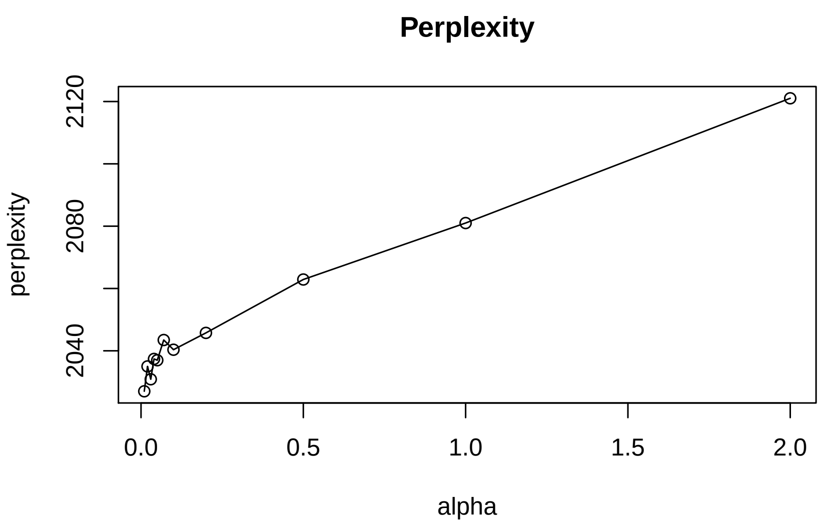


Figure 10: Topic coherence for different α

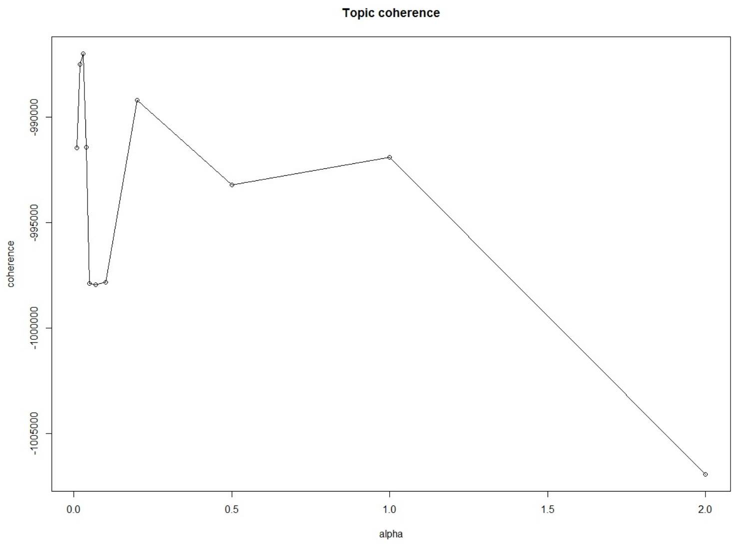


Figure 11: Topic distribution in boom and bust periods (LDA)

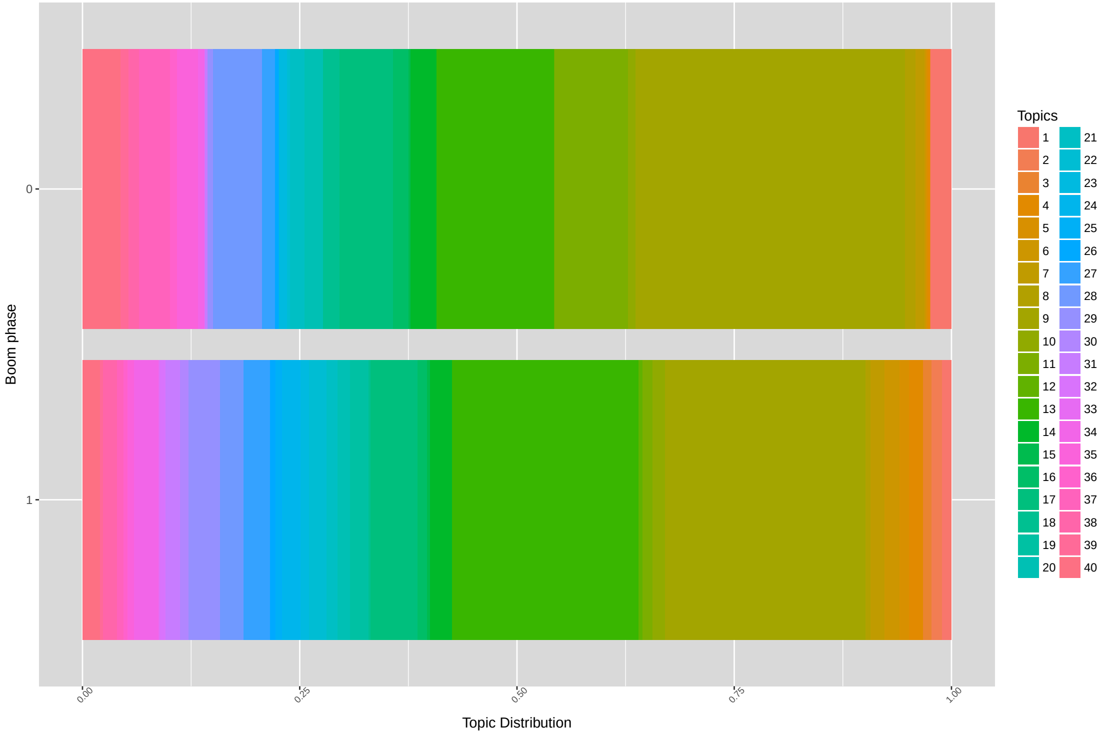


Figure 12: Topic distribution over institutions (LDA)

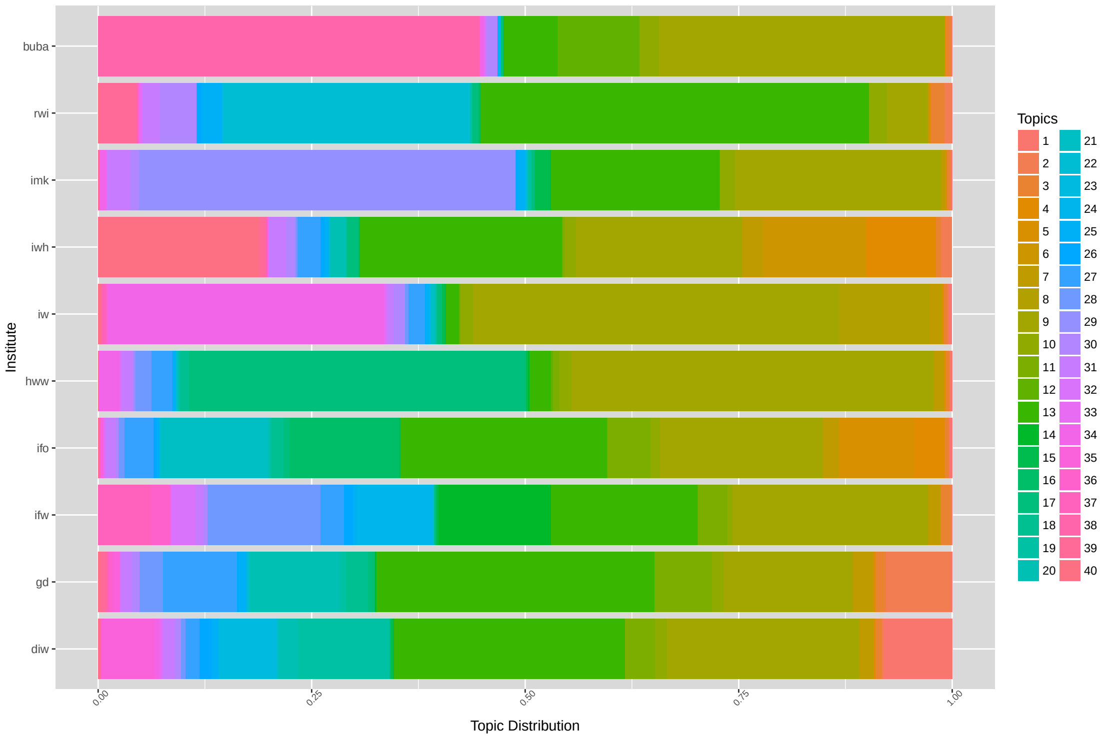


Figure 13: Topic distribution over time (LDA)

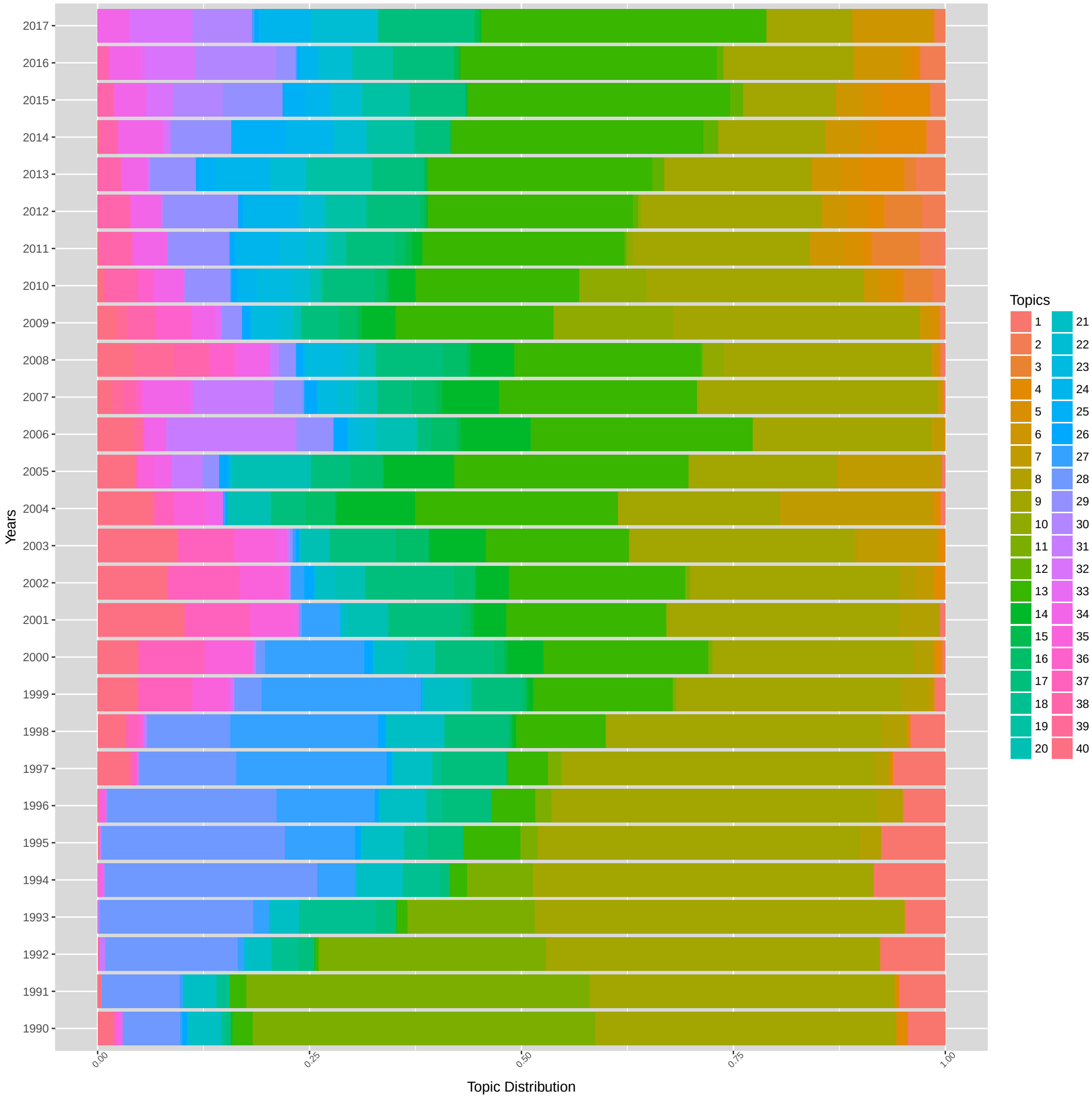


Figure 14: LDA topic 3: Refugees topic

Figure 15: LDA topic 2: Crises topic

Figure 16: LDA topic 4: UK trade and “Brexit” topic

Figure 17: LDA topc 31: Labour market reform topic

Figure 18: Results of "SearchK" function (stm package)

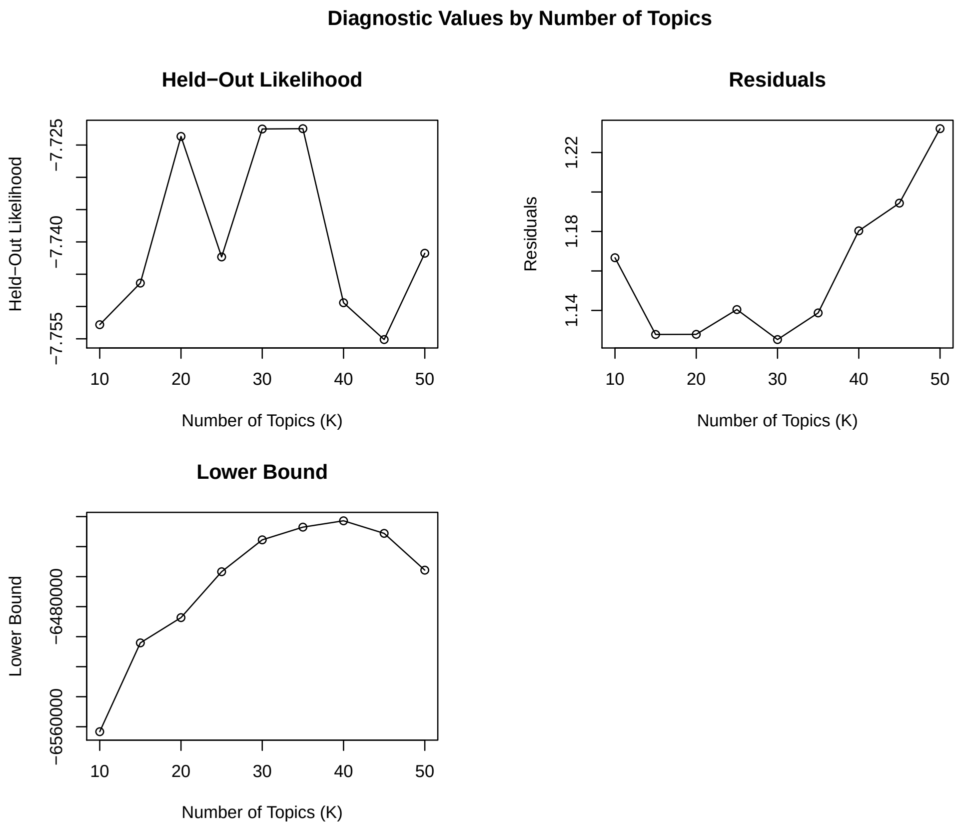
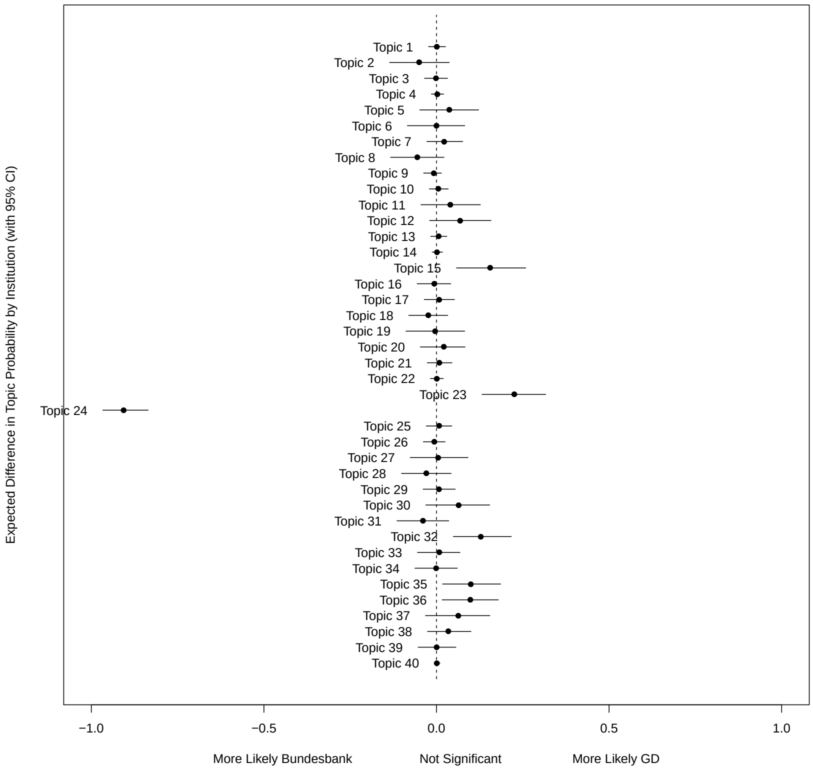
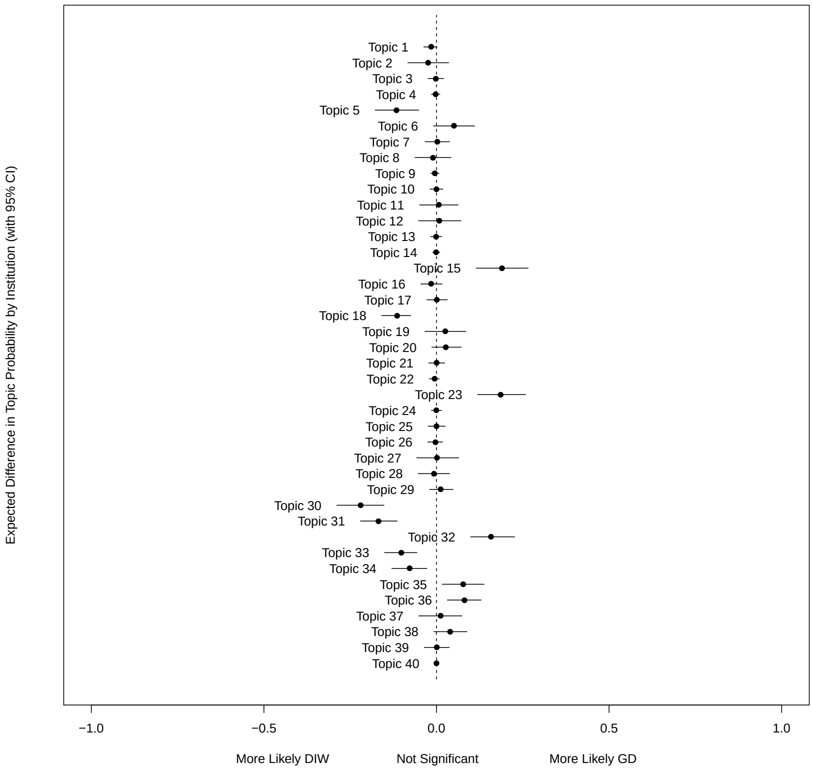


Figure 19: Effect of Institution on Topic Prevalence (STM), Reference = “GD” 1/4

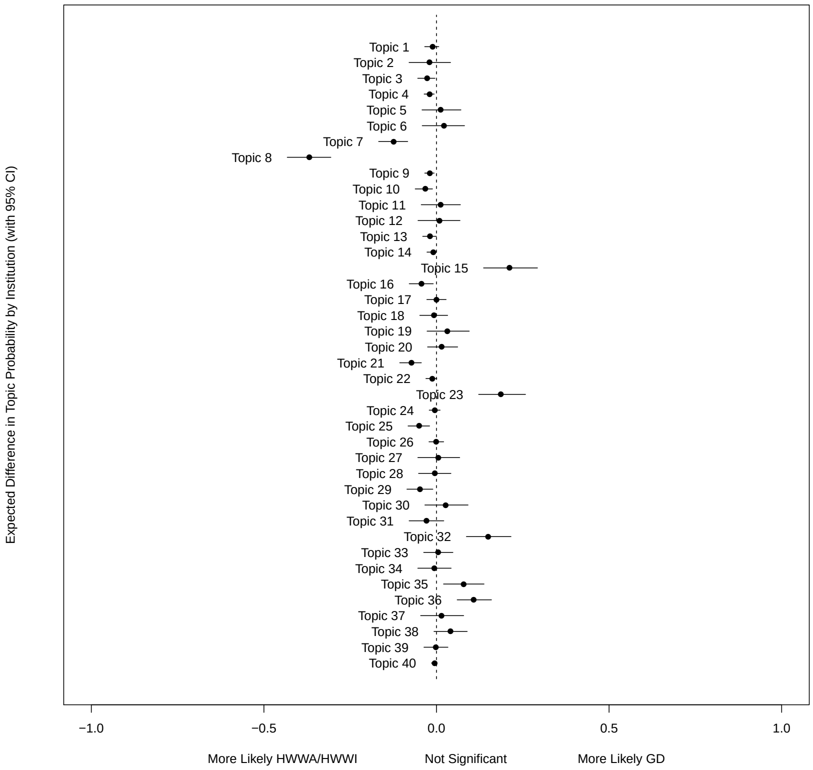


(a) Bundesbank vs. GD

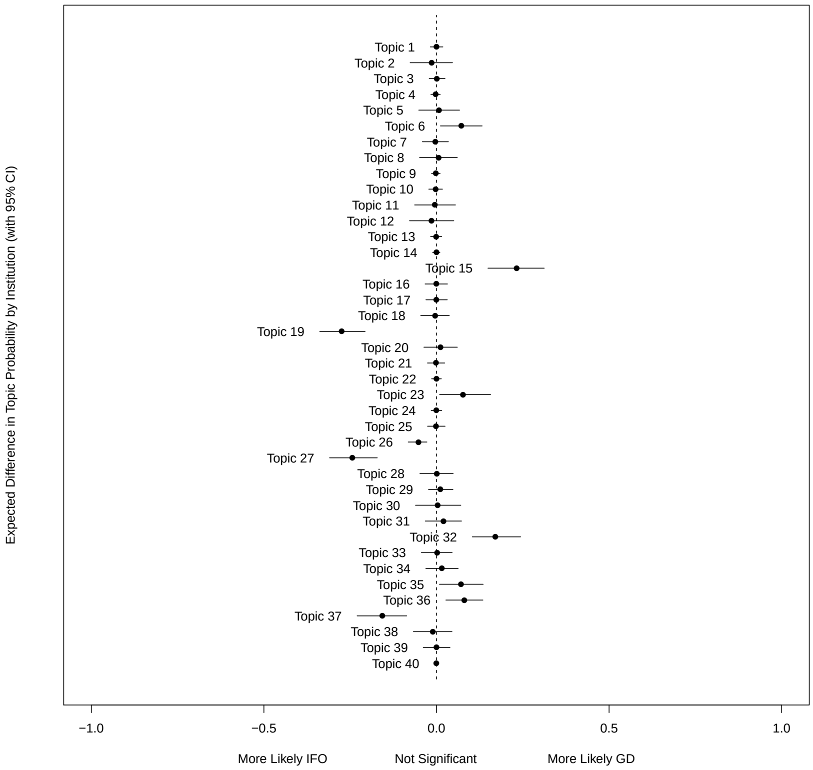


(b) DIW vs. GD

Figure 20: Effect of Institution on Topic Prevalence (STM), Reference = "GD" 2/4

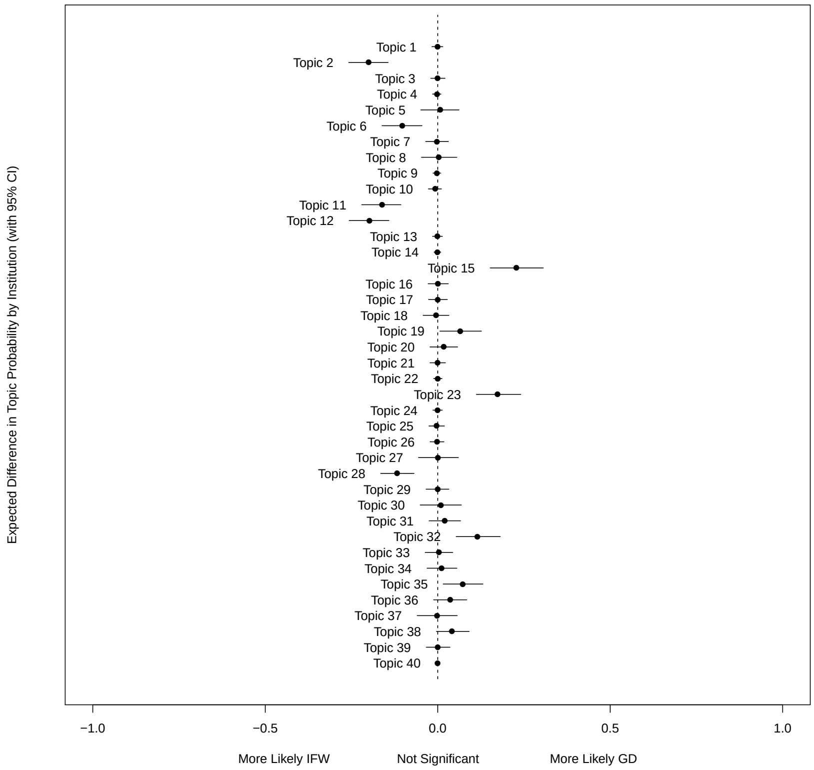


(a) HWWA/HWWI vs. GD

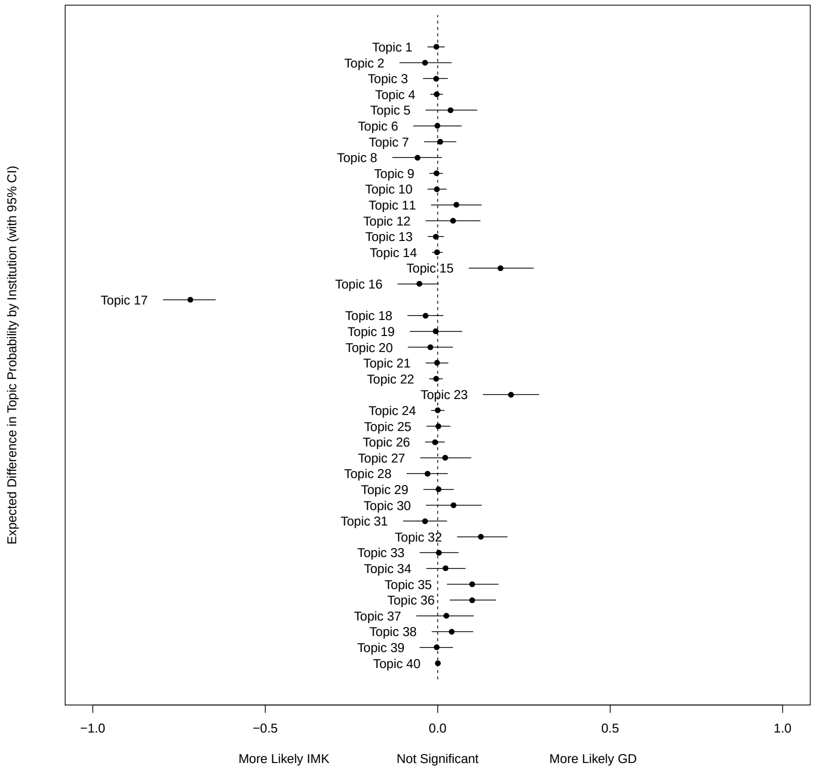


(b) ifo vs. GD

Figure 21: Effect of Institution on Topic Prevalence (STM), Reference = "GD" 3/4

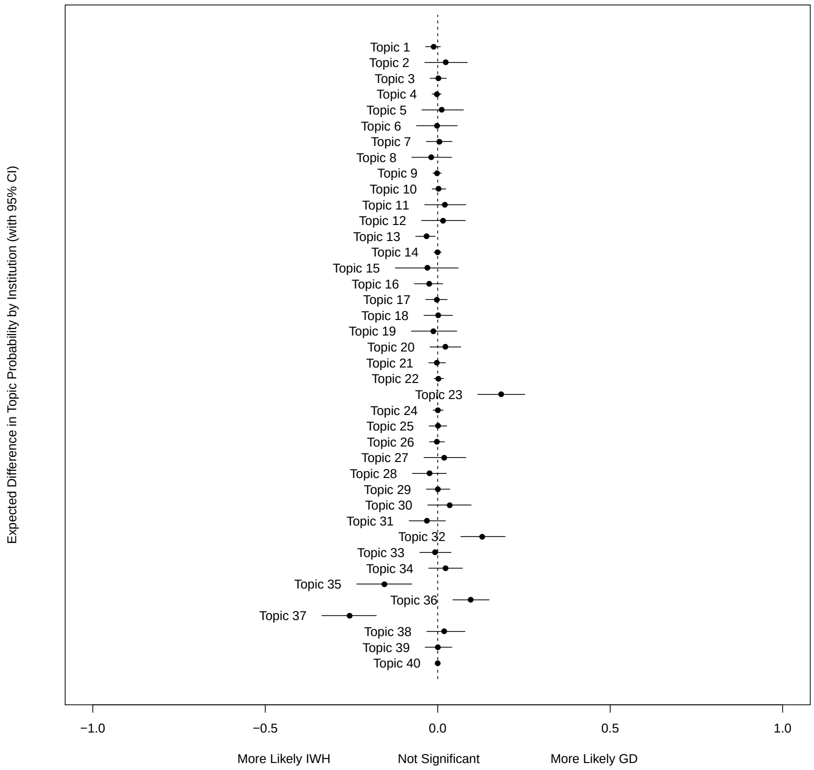


(a) IfW vs. GD

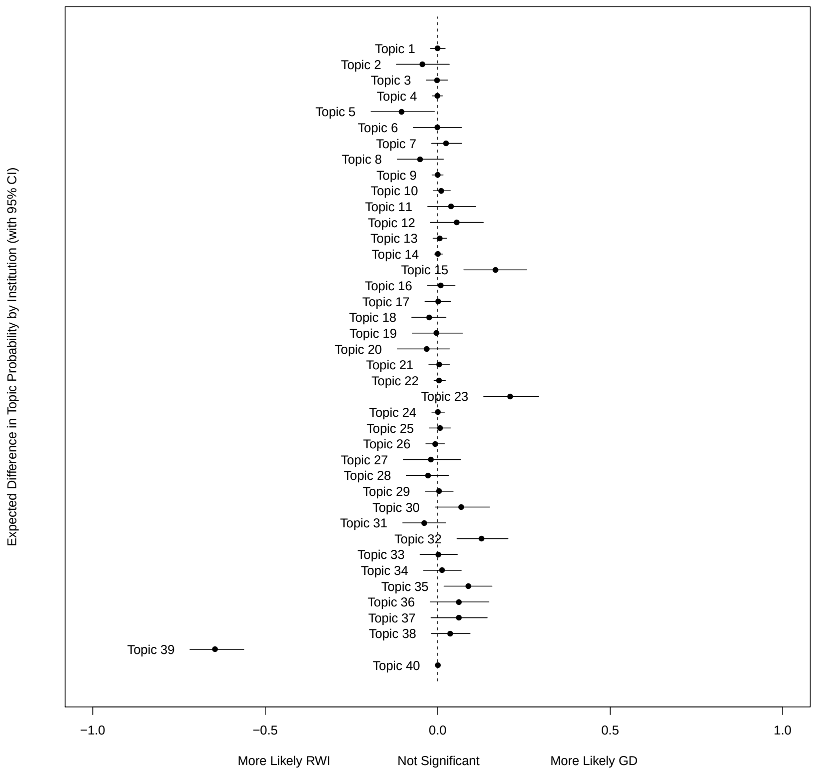


(b) IMK vs. GD

Figure 22: Effect of Institution on Topic Prevalence (STM), Reference = “GD” 4/4



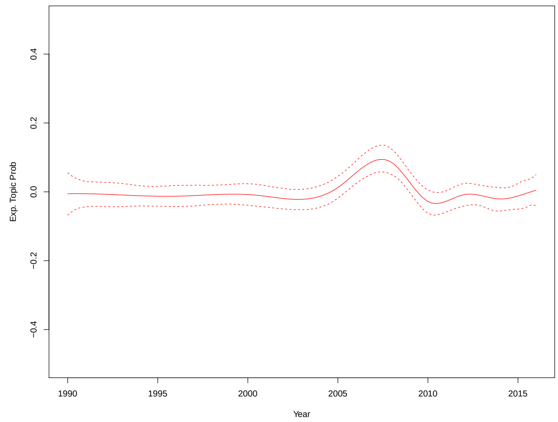
(a) IWH vs. GD



(b) RWI vs. GD

Figure 23: STM topic 16: "*Vorkrise/ Hochkonjunktur*" (Pre-Crisis/ Boom)

(a) Wordcloud

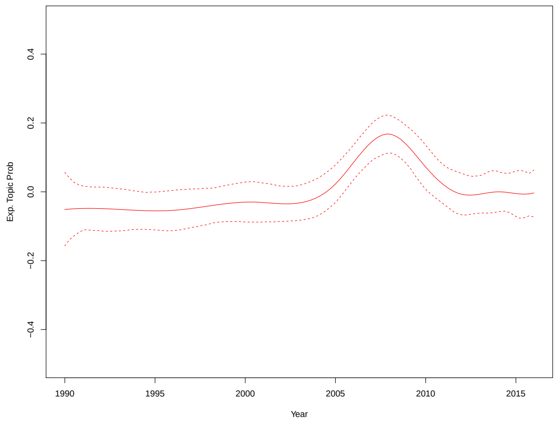


(b) Prevalence over time

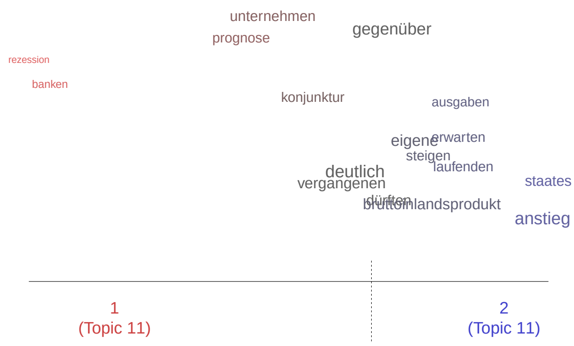
(c) Content by covariate

Figure 24: STM topic 11: "*Banken/ Krise*" (Banking/ Crisis)

(a) Wordcloud



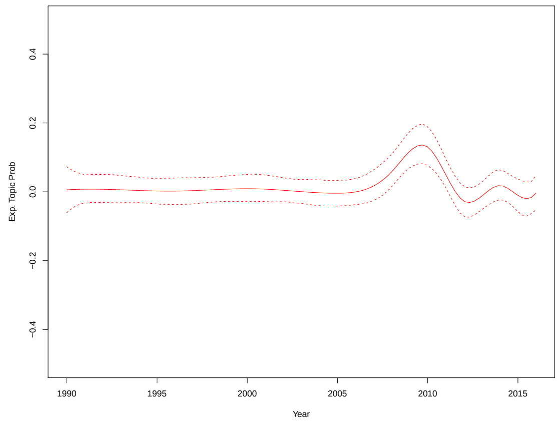
(b) Prevalence over time



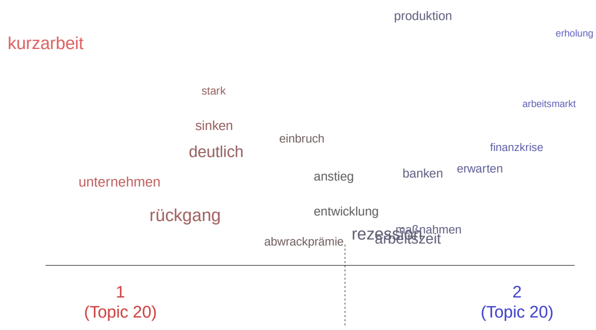
(c) Content by covariate

Figure 25: STM topic 20: "*Kurzarbeit/ Rezession*" (Short-time work/ Recession)

(a) Wordcloud



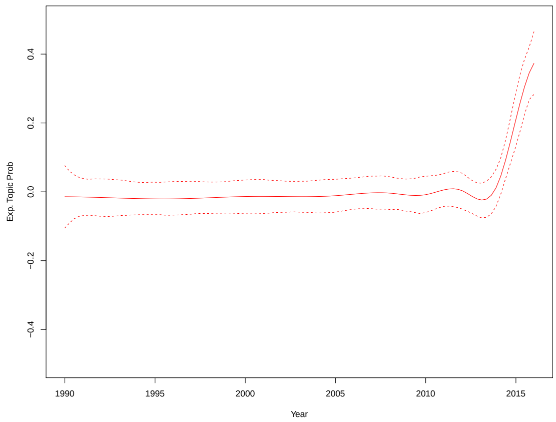
(b) Prevalence over time



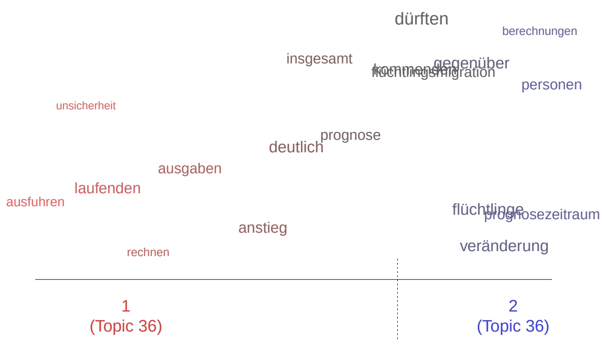
(c) Content by covariate

Figure 26: STM topic 36: "*Flüchtlingsmigration*" (Refugee Migration)

(a) Wordcloud



(b) Prevalence over time



(c) Content by covariate