HWWI Working Paper Series

Quantifying Priorities in Business Cycle Reports: Analysis of Recurring Textual Patterns around Peaks and Troughs

Alexander Foltas

Hamburgisches WeltWirtschaftsInstitut

HWWI Working Paper No. 5/2025

Authors:

Alexander Foltas (alex.foltas@gmx.de)
Helmut-Schmidt-Universität Hamburg

Imprint

Publication Series: HWWI Working Paper Series, ISSN 2750-6355
Responsible editor: Michael Berlemann

Hamburg Institute of International Economics (HWWI)
Scientific Director: Prof. Dr. Michael Berlemann
Managing Director: Dr. Dirck Süß
Mönkedamm 9 | 20457 Hamburg | Germany
Phone: +49 40 340576-0 | Fax: +49 40 340576-150
info@hwwi.org | www.hwwi.org

Quantifying Priorities in Business Cycle Reports: Analysis of Recurring Textual Patterns around Peaks and Troughs

Alexander Foltas[1]

[1] *Helmut-Schmidt-University Hamburg, Holstenhofweg 85, 22043 Hamburg, Germany.*

Abstract

This paper proposes a novel method to uncover shifting thematic priorities in textual business cycle reports and links them to macroeconomic fluctuations. To this end, I leverage qualitative business cycle forecasts published by leading German economic research institutes from 1970-2017 to estimate the proportions of latent topics. These topics are then aggregated into broader macroeconomic subjects using a supervised approach. By extracting the cyclical components of these subjects' proportions, I derive dynamic measures of forecasters' thematic priorities.

Correlating the cyclic components with key macroeconomic indicators reveals consistent patterns across economic expansions and contractions. Around economic peaks, forecasters emphasize inflation-related over recession-related topics. I thus propose that forecasters' failure to predict recessions may stem from a tendency to underestimate growth risks and overestimate inflation risks during periods of contractionary monetary policy. Around troughs, forecasters prioritize investment-related topics over general growth considerations.

Keywords: Macroeconomic forecasting; Evaluating forecasts; Recession forecasting; Topic Modeling; Natural language processing; Judgemental forecasting

1 Introduction

Professional business cycle forecasts are informed economic outlooks that typically comprise numerical point forecasts and accompanying textual reports that describe the forecasters' information set, expectations, and assumptions. Most of the studies investigating professional forecasts primarily focus on numerical point forecasts while largely overlooking the accompanying reports, even though the quantitative part is most likely only a parsimonious reduced-form representation of a forecaster's information set.

This narrow focus on the analysis of quantitative forecasts has been widened in recent years due to advances in computational linguistics.

Multiple studies in recent years use Natural Language Processing (NLP) methods to quantify textual forecasts and analyze their informational value. These studies show that texts contain valuable information that is often inefficiently incorporated into point forecasts through quantifying forecasters' tonality (Jones, Sinclair, and Stekler 2020; Clements and Reade 2020) or content of their narratives (Foltas 2022; Foltas and Pierdzioch 2022a, 2022). Other studies analyze the predictive power of sentiment indices derived from forecast reports (Hollrah, Sharpe, and Sinha 2020; Müller 2022) and construct rankings of forecasters based on their sentiment's predictive power (Rybinski 2021).

I contribute and expand upon this previous research by proposing a novel approach that leverages textual forecasts to uncover forecasters' priorities of macroeconomic subjects. Macroeconomic subjects are principal themes or concepts guiding analyses and discussions within the field of macroeconomics, such as the labor market, inflation, or recessions. Given the limited length of business cycle reports, forecasters must choose which subject to discuss in detail and which to address more briefly. I assume that forecasters strive to maximize the informational value of their reports by adjusting their priorities based on their perceptions of various subjects' importance in the current economic situation. For example, in a high inflationary environment, I would expect forecasters to provide a more in-depth analysis of inflation, resulting in a higher proportion of the forecast report dedicated to this subject.

Measuring and comparing the shares of macroeconomic subjects within forecast reports through quantitative textual analysis uncovers forecasters' priorities, enabling researchers to infer subjects forecasters regard as most important at a given time. In this regard, I apply the Word2Sense-LDA topic model (Foltas 2022), which uncovers latent topics in documents by identifying recurring word patterns on business cycle reports of three German forecasting institutes spanning the period from 1970 to 2017. Word2Sense-LDA, an extension of the widely-used latent Dirichlet allocation (LDA) topic model (Blei, Ng, and Jordan 2003), is specifically optimized for small, thematically-specialized corpora with widely divergent publication dates.

The quantification of forecast reports based on priorities represents a substantial methodological enhancement of previous quantifications of business cycle reports using the Word2Sense-LDA model (Foltas 2022; Foltas and Pierdzioch 2022a, 2022). Firstly, I employ supervised aggregation of estimated topics with similar themes, leveraging the aggregated topics' proportions to measure the share of broader macroeconomic subjects within the reports. This approach accounts for the findings of Blair, Bi, and Mulvenna (2020), which suggest that aggregating similarly themed topics enhances topic coherence. Secondly, I account for the fact that the topical compositions of forecast reports substantially change over longer periods due to advances in economic theory, structural economic shifts, and the elongation of forecast reports (Foltas 2022) by decomposing topic proportions into trend and cyclic components. The cyclic proportions of the topics exclude structural factors and thus effectively measure forecasters' subject priorities related to the business cycle.

Thirdly and most crucially, I replace the commonly used coherence-based hyperparameter calibration with a novel approach that utilizes the institutes' quantitative business cycle forecasts. Topic models require predefining the values of a set of hyperparameters to infer latent topics statistically. Given that these hyperparameters greatly influence the resulting topics, researchers typically optimize them by maximizing coherence scores, which aim to measure the interpretability of topics by humans. To quantify forecasters' priorities, I leverage the fact that forecasters publish quantitative business cycle forecasts alongside their business cycle reports. As numeric and textual forecasts should be consistent, I optimize the hyperparameters to maximize the fit subject priorities provide when modeling the corresponding point forecasts.

Quantified priorities reveal information fundamentally different from that captured by previous sentiment-based quantification approaches. Most sentiment-based methods reduce forecasts to a univariate score, typically reflecting overall optimism or pessimism. In contrast, a priorities-based quantification of forecasting reports generates multiple fine-grained variables that indicate the importance forecasters attach to specific subjects. This approach enables more differentiated conclusions and provides a nuanced understanding of how forecasters process information. To this end, I correlate forecasters' subject priorities with leads and lags of the cyclical components of various economic variables to investigate whether forecasters' qualitative assessments precede or follow economic fluctuations. The analysis focuses in particular on subject priorities around economic turning points.

Many researchers investigating quantitative forecasts find that forecasts often fail when they are most critically needed, as they do not accurately predict business cycle turning points and typically identify them in hindsight (e.g., Fildes and Stekler 2002; Heilemann and Stekler 2013; Döpke and Fritsche 2006; Stekler 2007). To the best of my knowledge, the only study investigating modern qualitative forecasts around recessions is by Stekler and Symington (2016). The authors investigate the FOMC protocols around the Great Recession and find that the committee not only failed to predict the recession but also was late at recognizing it. The committee correctly saw downside risks for the economy throughout 2007 but evaluated inflation risks to be of greater concern than growth risks until October 2008, after the collapse of Lehman Brothers. Only in December 2008, the committee declared that the USA had been in a recession for 12 months. However, the committee successfully predicted the recovery.

While analyses of quantitative forecasts and textual sentiments show forecasters' failure to forecast turning points, both approaches insufficiently explain the reasons for this systematic failure by themselves. I aim to shed light on potential reasons for these failures or even find previously overlooked textual warning signs by uncovering the subjects forecasters considered as most urgent in the preface and aftermath of turning points.

I organize the remainder of this study as follows. Section 2 conceptually clarifies the notions of topics, themes, and subjects, and outlines the framework for quantifying forecasters' priorities. Section 3 presents the data and describes the data transformation process. Section 4 introduces the Word2Sense-LDA model and details the approach for hyperparameter tuning and topic aggregation. Section 5 analyzes subjects directly related to specific business cycle phases, while Section 6 briefly discusses subjects that forecasters prioritize independently of the growth cycle. Finally, Section 7 provides concluding remarks.

2 Quantifying textual subject priorities

2.1 Topics, themes and subjects

Topic models are a class of statistical algorithms designed to discover the thematic structure within a collection of documents by identifying repeated word patterns and grouping these patterns into topics. Topics are multinomial probability distributions ϕ over a vocabulary of terms.[1] Topics are considered coherent when their most probable terms appear in similar contexts, which allows humans to decipher the topics' themes. In the context of forecast reports, a theme might be, for example, "inflation."

Most topic model algorithms are variations or improvements of LDA developed by Blei, Ng, and Jordan (2003) (Churchill and Singh 2022), which assumes that documents are probabilistically generated as bag-of-words that disregard word order and consider only the frequencies of words. The generative process creates documents by sampling its distribution over topics (denoted as θ). Then, the process generates each word in the document by sampling a topic from its θ distribution and a term from the word distribution ϕ of the drawn topic. For a given corpus, LDA and its variations allow for inferring the topics' ϕ distributions for a pre-set number of topics, as well as the θ proportions for each document. Additionally, LDA enables the calculation of a conditional probability distribution (γ) for terms assigned to each topic. I elucidate the technical details of LDA and the Word2Sense-LDA modification applied in this study in Section 4.1.

I define subjects as principal themes that allow for grouping and organizing related themes within a larger discourse or body of information. Subjects are not a result of measured semantical patterns but instead grounded in theory and motivated by the purpose of textual analysis. By interpreting the themes of topics, researchers can group related topics into subjects and calculate their proportions in documents by aggregating the proportions of the assigned topics. As such, subjects serve as a valuable tool for researchers to reduce the number of topics, which may be practical for analysis.

For instance, a topic model might infer two topics with the themes "exports" and "imports." If a researcher desires to investigate the overall relevance of foreign trade within the corpus, it can be suitable to classify both topics as part of the foreign trade subject and aggregate their

1. The topic modeling literature uses the concepts of "terms" and "words" inconsistently. This study defines each linguistic expression within a dictionary of a corpus as a term and its observations in the data as words.

probability distributions. This aggregation not only facilitates analysis by reducing complexity but also enhances topic coherence, as demonstrated by Blair, Bi, and Mulvenna (2020). I calculate the probabilities and proportions of the subject-based aggregated topic S comprising $1 + \ldots + N$ assigned topics as follows:

$$\phi_S = \frac{\phi_1 + \ldots + \phi_N}{N}, \tag{1}$$

$$\gamma_S = \gamma_1 + \ldots + \gamma_N, \tag{2}$$

$$\theta_S = \theta_1 + \ldots + \theta_N. \tag{3}$$

2.2 The framework

I assume that the forecast report of institute i at time t consists of $K = K^e + K^n$ latent topics with K^e representing economic topics connected to macroeconomic subjects and K^n noise topics that measure forecasters' discussions of matters unrelated to the macroeconomy. To simplify, each economic topic equals an economic subject. I assume that a macroeconomic topic's relative proportion $\theta_{k^e,i,t}$ measures the related subjects' overall priority at a given point in time.

In addition, I decompose forecasters' subject priorities $\theta_{k^e,i,t}$ into trend $\theta^T_{k^e,i,t}$ and cyclic $\theta^C_{k^e,i,t}$ components:

$$\theta_{k^e,i,t} = \theta^T_{k^e,i,t} + \theta^C_{k^e,i,t}. \tag{4}$$

As shown by Foltas (2022), topic proportions exhibit trend characteristics. To correlate priorities with business cycle fluctuations, I assume that subject proportions trend component $\theta^T_{k^e,i,t}$ depends on structural factors and do not comprise information regarding the forecaster's perceptions of the current business cycle stage. These factors include, for instance, the macroeconomic structure, economic paradigms, and the resources of the forecasting institutes. As structural factors slowly change over time, they explain the topic proportions' trend paths.

Cyclical components $\theta^C_{k^e,i,t}$ measure the deviation of forecasters' subject priorities from their trend level. These components provide valuable insights into forecasters' attributed importance of subjects within the business cycle.

For instance, when forecasters anticipate the unemployment rate will remain within a reasonable range around its trend level, there are no business cycle-related reasons to prioritize discussions concerning unemployment. In such cases, I would expect the cyclical components of unemployment-related topics to be around zero or below. However, when forecasters worry about upcoming rises in unemployment, they might want to provide more detailed textual explanations for their concerns, offering a rationale for their forecasts or discussing uncertainties. This increased emphasis on unemployment-related topics causes their proportions to exceed their trend levels. Therefore, quantified cyclical components serve as valuable supplementary information, allowing researchers to identify which specific subjects forecasters consider to be important within the current stage of the business cycle.

A qualitative forecast is quantitatively represented by the vector of its economic topic's cyclical components $\theta^C_{K^e,i,t}$. Accordingly, the vector of forecasters' point predictions $\hat{X}_{i,t+1}$ can be decomposed into the institutes' trend forecasts $\hat{X}^T_{i,t+1}$ and their anticipated near-term trend deviations $\hat{X}^C_{i,t+1}$:

$$\hat{X}_{i,t+1} = \hat{X}^T_{i,t+1} + \hat{X}^C_{i,t+1}. \tag{5}$$

Both vectors of cyclic components $\theta^C_{K^e,i,t}$ and $\hat{X}^C_{i,t+1}$ are different representation forms of the same underlying information set and, as such, should be consistent in the sense that both representations reflect each other's embedded information. Thus, both vectors of cyclic variations should correlate, allowing the estimation of macroeconomic forecasts using recovered priorities of related topics, and vice versa.

Consistency, however, does not require that both representations necessarily are perfectly correlated. Firstly, both vectors are reduced-form representations of the forecaster's information set. Secondly, quantitative forecasts may be subject to behavioral biases, such as forecaster herding, which can introduce discrepancies between the qualitative and quantitative representations. Thirdly, imperfect natural language processing methods will most likely imperfectly recover forecasters' priorities.

Therefore, I do not expect a perfect fit when mapping the elements of the qualitative representation of a forecaster's information set onto associated quantitative forecasts. However, an optimally calibrated natural language processing model should be highly successful in uncovering and grouping semantic nuances associated with specific quantitative forecasts. Thus, the explanatory power priorities achieved when modeling quantitative forecasts serve as a suitable measure for the quality of their quantification.

3 Data

3.1 Forecasts

I study a corpus of texts comprising 349 business cycle reports ranging from 1970 to 2017 as published by three major German research institutes German Institute for Economic Research (DIW), Ifo Institute for Economic Research (ifo), and Kiel Institute for the World Economy (IfW). Table 1 presents the descriptive summary statistics of the corpus divided by institutes. During the 1970s and 1980s, the institutes published on a half-yearly basis or less. In the subsequent period, the DIW and IfW switch their publication frequency to quarterly reports. Furthermore, the reports' average length increases over time, indicating more detailed qualitative elaborations. The changes in forecast reports' publication frequency and level of detail provide the necessity for sample splits as robustness checks.

Table 1. Corpus summary statistics

	DIW		ifo		IfW	
Period	N	w	N	w	N	w
1970s	11	2179	20	2300	20	2275
1980s	18	1850	19	2548	20	2397
1990s	20	1624	20	3499	36	2694
2000s	25	1708	19	3602	40	2680
2010s	33	2346	16	3698	32	3555

Notes: N: Number of Documents. w: Average number of words per document after preprocessing.

Table 2. Publication months of forecast reports.

Institute	01	02	03	04	05	06	07	08	09	10	11	12
	Calendar month											
DIW	40	0	4	6	0	19	18	0	6	5	1	8
ifo	5	3	2	0	0	2	25	4	0	0	0	20
IfW	0	1	25	0	0	44	4	0	25	1	0	48

Table 2 shows the publication months of the forecast reports by institutes and reveals that the institutes mainly publish in December, January, June, and July. I split the sample into half-yearly winter and summer forecasts for the analysis in Section 5. A subdivision into quarterly data would lead to many quarters with missing values, while annual forecasts would result in too small sample size and information losses. Winter half-year ranges from September to February, and the summer half-year ranges from March to August, which minimizes the half-years with missing data.

For the hyperparameter optimization in Section 4.2, I utilize each forecast report's point predictions for growth (pGDP), inflation (pCPI), and unemployment rates (pUN) to optimize the consistency of textual and numerical forecast representations. The data is publicly available on the IWH Forecasting dashboard (Heinisch *et al.* 2023). The summer reports use the current year's predictions (t0). The winter reports use the next year's predictions (t1) from September to December or the current year's (t0) in January and February.

3.2 Macroeconomic variables

As there is no official dating of German recessions, I use the dating of peaks and troughs by the Economic Cycle Research Institute (ECRI). The ECRI recession periods are widely used in the literature (e.g., Kholodilin 2005; Döpke, Fritsche, and Pierdzioch 2017) as they track swings in the production cycle nearly perfectly. There are five recessions in the sample: The first oil-price shock (1973-1975), the second oil-price shock (1980-1982), one after the German Reunification (1991-1994), one after the dotcom boom (2001-2003), and the Great Recession (2008-2009).

Furthermore, I employ other macroeconomic variables to provide a broader picture of forecasters' priorities in different economic phases, which are presented in Table 3. These are the rates of change of the three key variables GDP, inflation, and unemployment. Furthermore, I employ the aggregated investment, government consumption, the balance of trade, and the short-term interest rate, as these variables relate to macroeconomic subjects.

3.3 Data transformation

The raw documents require multiple pre-processing steps to create economic topics with high informational value. Firstly, I remove all parts not concerning qualitative analyses of German macroeconomic conditions. These include sections regarding other economies, methodological discussions and explanations, microeconomic analyses concerning specific sectors, and general policy advice. Subsequently, I convert letters into lower-case forms and remove all numbers and punctuation marks.

Furthermore, I filter expressions without economic meaning out of the corpus. These stop words include articles, prepositions, pronouns, conjunctions, and the most frequently used nouns, verbs, and adjectives without specific informational value. Examples are "percent," "average," "image," and "seasonally adjusted." Although these words are important for reading comprehension, they appear in any context, thus offering no insights about the forecast report concerns while lowering topic quality. Furthermore, I remove all terms that appear less than five times in the corpus, which is crucial for the German language that heavily relies on compound words. These measures reduce

Table 3. Macroeconomic variables

Variable	Acronym	Description	Source
Recession Dummy	ECRI	Period following the peak through the trough (trough method)	ECRI
Growth rate	GDP	Year-on-year rate of change of the quarterly GDP; price and seasonally adjusted	FRED
Consumer Prices	CPI	Logarithmic year-on-year rate of change of the monthly consumer price index; calendar and seasonally adjusted	BUBA
Unemployment rate	UN	Monthly registered unemployed in percent; calendar and seasonally adjusted	BUBA
Investment	GFCF	Quarterly gross fixed capital formation in percent of GDP; seasonally adjusted	FRED
Government Expenditure	GOV	Government Final Consumption Expenditure in percent of GDP; seasonally adjusted	FRED
Balance of Trade	TRADE	Quarterly exports minus imports in percent of GDP; seasonally adjusted	FRED
3 months interest rate	IR	Monthly average of 3 months money market rate	BUBA

Notes: All indicators are for Germany; ECRI - Economic Cycle Research Institute; BUBA - German Central Bank; FRED - Federal Reserve Bank of St.Louis.

the number of terms in the vocabulary from around 50000 to 14500.

All topic proportions (after potential aggregations), numerical forecasts, and macroeconomic variables χ are standardized into χ_S using their mean values $\overline{\chi}$ and standard deviation σ_χ to allow better comparability and interpretability:

$$\chi_S = \frac{\chi - \overline{\chi}}{\sigma_\chi}, \tag{6}$$

Furthermore, I extract the cyclical component for all variables χ_C by subtracting their trend components χ_T from their standardized form:

$$\chi_C = \chi_S - \chi_T. \tag{7}$$

The trend component is estimated using the locally estimated scatterplot smoothing (LOESS) regression with a span of 0.9.[2] From this point onwards, when I mention topic proportions, numeric forecasts, or macroeconomic variables, I refer to their cyclic components if not otherwise specified.

In order to allow institutes to be compared, I calculate half-yearly values for topic proportions and macroeconomic variables before extracting cyclic components in Section 5. Herefore, I average macroeconomic variables for each half-year and determine recession half-years using a majority vote. The topic proportions are adjusted to the mid-date of the half-year using a linear approximation, regarding that the publication date is the most crucial factor for forecasting accuracy (Döhrn and Schmidt 2011; Köhler and Döpke 2023). If an institute has multiple forecasts in a given half-year, I average their topic proportions. This simplification is necessary because of the infrequent forecast report publications. The data averaging might lead to inaccuracies but should still offer reasonable precision for determining recurring textual patterns.

2. Various methods were tested to estimate the trend component. Firstly, I used a simple mean and a simple linear regression. However, neither method achieves stationarity of the cyclic components in all cases. The HP-Filter led to relatively similar results compared to a LOESS regression while needing significantly more computational time, which was scarce in this study. A cubic regression also resulted in very similar outcomes.

4 Model construction

4.1 Word2Sense-LDA

LDA task is to decipher each topic's term distribution ϕ, the conditional probabilities of terms being assigned to them γ, and each document's topic distribution θ using a document-term matrix that counts the word frequency of each document in the corpus. As the posterior distributions are intractable to compute, they require statistical inference methods such as variational inference or collapsed Gibbs sampling. For their application, it is necessary to pre-set the number of topics K and two Bayesian priors: α, which increases the topics per document, and the multinomial parameter β, which increases the number of words with high ϕ probabilities per topic.

Although LDA is a powerful tool, it has shortcomings. In this study's context, LDA's most serious flaw is the inability to handle samples stretching longer periods (Blei and Lafferty 2006). Written language and textual patterns evolve, and LDA cannot handle this evolution appropriately. For instance, the words "gross national product" and "gross domestic product" describe two closely related concepts, which is why each topic's ϕ values should positively correlate. Instead, LDA infers strongly negative correlated ϕ probabilities for both terms as they seldom appear with high frequency in the same document. For this study's thematically narrow corpus with forecast reports across multiple decades, a document's θ proportion primarily depends on the document's publication date, thus resembling a time dummy.

Researchers address LDA's inability to handle language evolution by splitting the corpus into time chunks (e.g. Blei and Lafferty 2006; Dieng, Ruiz, and Blei 2019) or indirectly overcoming the issue by embedding words in a vector space (e.g. Dieng, Ruiz, and Blei 2020; Thompson and Mimno 2020). While these improved topic models seem promising, they require substantially more documents than this study utilizes. Foltas (2022) proposes the Word2Sense-LDA that replaces LDA's document-term matrix with a co-occurrence matrix that counts the times terms co-occur within a skip-gram window of n. Applying collapsed Gibbs sampling over a co-occurrence matrix leads to topics that correctly relate two words with similar meanings from different periods, as long as they appear in similar contexts. As this LDA extension does not require large corpus sizes, hence it suits this paper's corpus.

I use the R package "textmineR" (Jones and Doane 2023), which relies on collapsed Gibbs sampling to infer the posterior distributions.

4.2 Hyperparameter tuning

Word2Sense-LDA requires optimizing the number of topics K, the skip-gram window's size n, and the Bayesian priors α and β. A topic model's hyperparameters considerably impact the quality and characteristics of their resulting topics (Ke, Montiel Olea, and Nesbit 2022), which led to the development of various evaluation metrics, with methods aiming to maximize the topic's semantic coherence becoming more prevalent in recent years (Churchill and Singh 2022). Pointwise mutual information (PMI) is the most common metric within the coherence-based evaluation approaches. It measures the closeness of a topic's most probable words by comparing the number of times both words co-occur in the corpus relative to their total frequency. When a topic's top words exhibit high PMI with each other, it likely has an underlying theme that humans can easily comprehend.

As this study aims to quantify textual forecasts, the primary optimization goal is not to maximize topic coherence but to optimally model the underlying information. As discussed in Section 2.2, quantified textual forecasts should be well suited to explain the corresponding numeric forecast to increase the likelihood of an accurate representation. Thus, I compare different hyperparameter constellations' quantification success based on the fraction of variance their topic proportions explain when modeling the forecaster's point predictions. Therefore, I run the Word2Sense-LDA over the whole corpus without separating the institutes to ensure identical topics and enable comparability. Then, I model the most important numeric forecasts (pGDP, pINF, pUN) for each institute separately using the respective topic proportions as explanatory variables.

Further, I measure each institute's adjusted R^2 in a two-step procedure. Firstly, I identify the economic topics from the overall topic pool using a multivariate least absolute shrinkage and selection operator (Lasso) regression and excluding all topics without informational value for the forecasts.[3]. The remaining topics are employed in univariate linear regressions, resulting in three adjusted R^2 per institute, which are averaged to calculate each institute's overall coefficient of determination.

To provide insights regarding different hyperparameter constellations' effects on quantification success, I assess around 1500 hyperparameter constellations using 200 iterations of the Gibbs sampler to reduce computation time. Figure 1 presents the mean and standard deviation between the institutes adjusted R^2 for each fitted topic model. For a multitude of models, the mean adjusted R^2 is close to or below the standard deviation, indicating that the trained topics do successfully quantify the textual forecasts of one or two institutes while completely failing to represent the others. To avoid analyzing a topic model with very uneven quantification success for the different institutes, I calculate a score by subtracting the standard deviation of each institute's adjusted R^2 from the overall mean adjusted R^2.

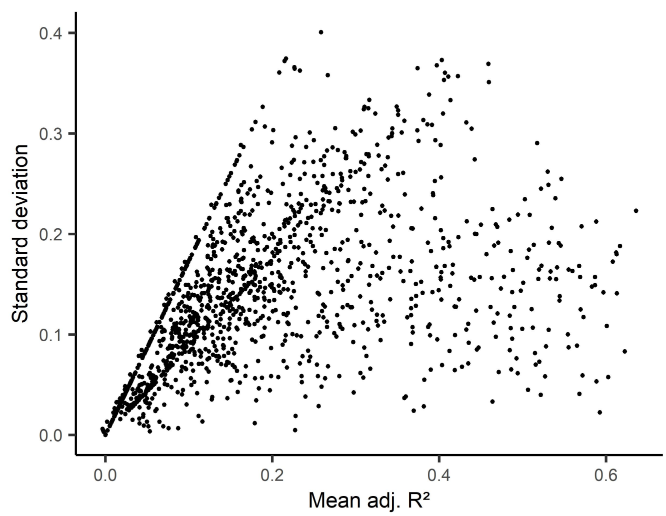


Figure 1. The institutes' adjusted R^2 for different hyperparameter constellations.
The x-axis presents the mean adjusted R^2 over the institutes, while the y-axis presents their standard deviation.

Figure 2 presents uni- and bivariate partial regression plots showing the estimated ceteris paribus effects of different hyperparameters. A higher number of topics K seemingly always increase performance potentially through topics being better tailored to the individual institutes and periods. However, too many topics complicate interpretability as economic subjects are fragmented over too many topics. Also, the number of topics has a significant impact on computation time. Therefore, I limit K to 125 for further analyses. The most impactful parameter is α, with the standard value $50/K$ (Griffiths and Steyvers 2004) achieving the highest scores. Variations of β and n result in minor performance differences.

It has to be noticed that the random initial distributions of the Gibbs sampler have a more substantial effect on topic model scores compared to small hyperparameter variations close to the optimal values, as the scattering shows. Thus, it is more promising to fit multiple models with

3. Optimally, the topics with economic content should be selected by qualitatively interpreting their word clouds' theme according to the selection in 4.3. However, this approach is unfeasible for the large pool of topic models.

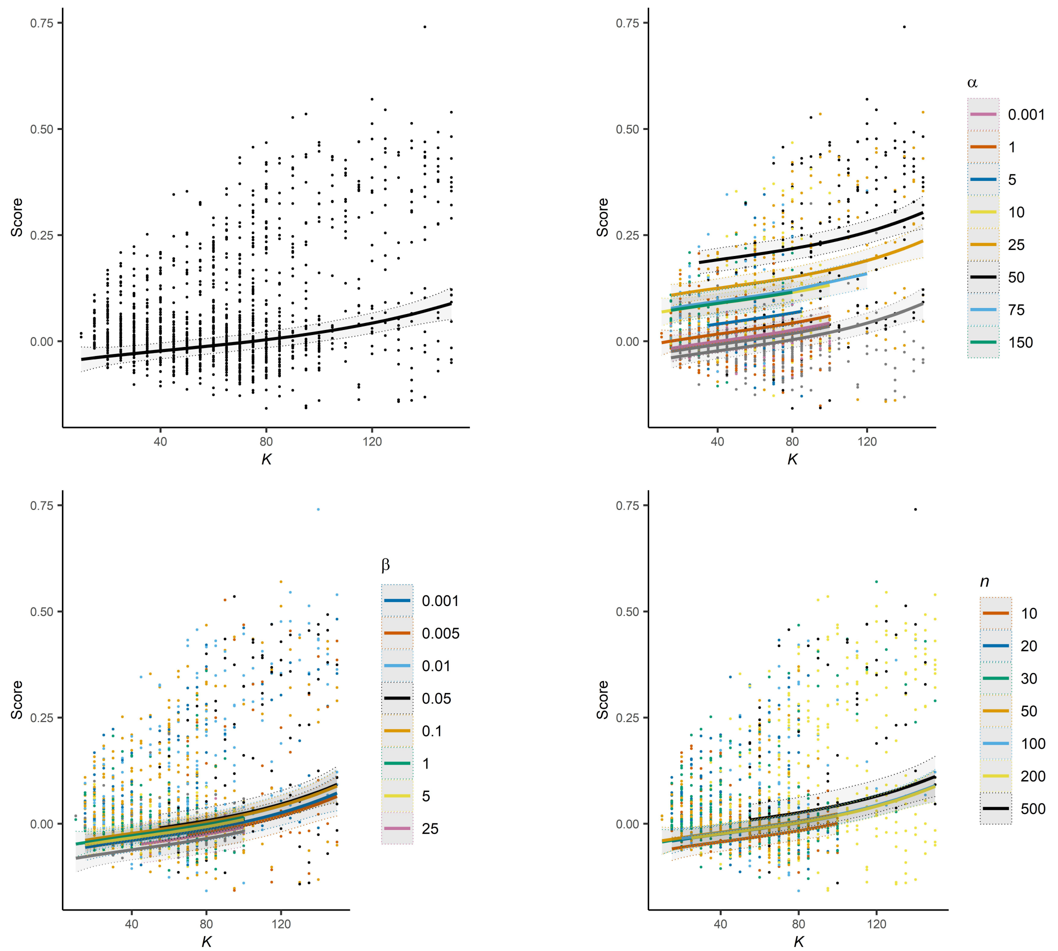


Figure 2. Partial regression plots.
Points show actual and lines show predicted adj. R^2 values for different hyperparameter constellations. Predictions are conducted with a cubic regression using K as a numeral, while other hyperparameters are included as categorical variables. Gray ribbons indicate 95% confidence intervals. The legend shows only the eight best parameter values; other parameters are plotted in grey. α input values are divided by the number of topics.

small hyperparameter variations close to the optimal hyperparameters to obtain a topic model with high quantification success.[4]

Following the studying of different hyperparameters' effects, I adjust the stop-word dictionary and reoptimize the hyperparameters for the new corpus fitting a smaller number of topic models.[5] I increase the Gibbs sampler's number of iterations to 400 and set $\alpha = 50/K$ and $n = 200$,[6] while varying K and β around 100 and 0.05, to fit multiple similar calibrated topic models with varying initial distributions while using the same seed. The second optimization provides a topic model with 100 topics and a sufficient score of 0.44 for the best fit, which will be evaluated in the analysis.[7]

4.3 Topic selection and aggregation

The 100 obtained topics require further preparation to allow an economically meaningful evaluation. First, I select topics with macroeconomic themes. Second, the large number of topics requires aggregating selected topics into broader macroeconomic subjects. While analyzing and comparing different topics with similar themes is informative, I rather focus on a broader picture.

I use word clouds as the main tool to identify the theme of a topic. Each word cloud shows up to 25 of its most probable terms according to ϕ values. The term with the highest ϕ probability is placed in the center. Other terms are sorted around the center term, with their size indicating their ϕ probabilities relative to the center term. Terms with lower ϕ probabilities than 20% of the center terms' ϕ probability are not depicted in the word cloud. Hence, a word cloud consisting only of a few terms indicates low dispersion of a topic's word distribution. Furthermore, a term's color saturation increases with its γ probabilities, with red shades showing probabilities of 50% or more.

A supervised interpretation of topics' themes and categorization into subjects is naturally subjective.[8] While most topics' themes are quite clearly interpretable and assignable to the subjects related to macroeconomic aggregates, there are some edge cases when the theme is unclear or could fit into multiple subjects. For example, a topic that thematizes labor unit costs might be relevant for a labor market subject or a trade subject. In such cases, I decide the classification by comparing the ϕ and γ probabilities of terms unambiguously connected to one subject.

For a topic to be categorized, terms with high ϕ and γ probabilities in its word cloud should mainly have an unambiguous economic meaning by themself. This definition excludes adjectives and adverbs that modify economic terms without being inherently economic, such as sentiment terms. While forecast reports' sentiment contains important information value, as several studies show (e.g. Hollrah, Sharpe, and Sinha 2020, Müller 2022), this study focuses on forecast quantification through revealing textual priorities. Examples of topics without an inherently economic theme are depicted in Figure 3.

Some topics have economic themes that are not of interest to this study as they are either microeconomic or unspecific. Topic 24 in Figure 4 is an example of a topic with a microeconomic theme concerning the industrial sector.[9] While industrial production is an important predictor of

4. The α parameter could be optimized further within the range of 25 to 75. However, due to the high impact of the initial distribution on the score, I refrain from further optimization of the α parameter.

5. Due to an oversight, only terms with four or more letters are included in the vocabulary of the first optimization leading to unacceptable information losses by excluding acronym as "GDP" or "ECB." The first optimization fulfilled its purpose of providing insights into different hyperparameters' effects, but its resulting topic models are unsuited for further analysis. The new corpus lowers the minimum threshold for including terms in the vocabulary to two letters.

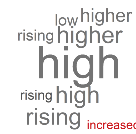
6. $n = 500$ achieved insignificantly higher scores while requiring considerably more computation time.

7. The model's exact hyperparameters are $K = 100$, $\alpha = 50$, $\beta = 0.15$, and $n = 200$.

8. While an objective unsupervised method might be preferable, I find the approach of Blair, Bi, and Mulvenna (2020) employing distance metrics for aggregating topics unsuitable for this study as topics with related themes according to economic theory might have very dissimilar ϕ distributions and vice-versa. Hence, a supervised approach is better for precisely analyzing qualitative forecasts.

9. While "commerce" is a correct translation of the term "Gewerbe," a better translation for "verarbeitenden Gewerbe" would be "manufacturing industry." The term "verarbeitenden" has the second highest ϕ probabilities and is translated as "processing" in the word cloud. The terms conjugation suggests a usage as "verarbeitenden Gewerbe."

(a) Topic 90



(b) Topic 81

Figure 3. Showcase: Non-economic topics.
All terms are automatically translated using DeepL. Different German terms might be translated into identical English terms causing duplicate appearances. The size of a term and its position indicate its ϕ probabilities. Only terms with ϕ values that exceed 20% of the topics' highest ϕ value are plotted to a limit of 25. The color saturation implies the term's γ value, with red shades indicating probabilities above 50 percent.

economic growth and employment, the highly specialized sector topic does not clearly fit into growth or employment subjects. Economic terms used in a broader macroeconomic context as "production," have a low chance of getting assigned to Topic 24, as depicted by their very low γ probabilities.[10] Hence, excluding the topic and other microeconomic topics from further analysis mostly removes terms of low relevance for macroeconomic assessments.

Topic 78 includes multiple economic terms such as "incoming orders," "production," "equipment investments," "consumer spending," or "exports." As each of these economic terms relates to different macroeconomic subjects, the topic cannot be assigned clearly to a macroeconomic subject. Instead, the topic's proportions might measure the size of introductions or overview sections in forecast reports. Excluding Topic 30 does lead only to minor information losses due to the low γ of the relevant economic terms.

Table 4 shows eight subjects representing 38 topics that I could identify within the remaining topics.[11]

Topics part of the same subject vary in the specificity or emphasis of their themes. For demonstration purposes, I present four topics with labor market themes in Figure 5. Topic 71 measures the discourse around "persons,"[12] potentially referring to sections discussing released or forecasted numbers of employed or unemployed people. Topic 40 shows a similar theme with a more dispersed ϕ distribution. Thus, both labor market topics have unemployment themes that vary in their specificity. The emergence of the two themes might be explained by different textual patterns in labor market sections for different institutes and periods. A sole interpretation of one of these topics might lead to misconclusions as each topic's proportions captures the discussion concerning unemployment only partly.

Topic 89 and Topic 25 depict different themes relating to the labor market subject. The for-

10. The word cloud shows high γ probabilities of the term "trade," which is the translation of "Handel." When forecasters refer to foreign trade, they typically use the term "Außenhandel."

11. It also would be interesting to analyze how monetary-policy topics vary with key macroeconomic variables. However, especially the earlier business cycle reports usually contain specific sections devoted to comments on macroeconomic policies, making it challenging to disentangle key aspects of monetary policy from general policy topics. While I could present such a general policy topic, the interplay of monetary policy and fiscal policy within such a topic makes its interpretation difficult. To ensure clarity and accuracy, I focus solely on presenting two specific policy topics that the topic model identifies separately from the general policy discussions.

12. In this context, the German term "Personen" is best translated as "persons," distinct from "people."

(a) Topic 24

(b) Topic 30

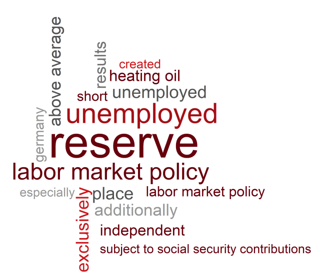
Figure 4. Examples of not utilized economic topics.
All terms are automatically translated using DeepL. Different German terms might be translated into identical English terms causing duplicate appearances. The size of a term and its position indicate its ϕ probabilities. Only terms with ϕ values that exceed 20% of the topics' highest ϕ value are plotted to a limit of 25. The color saturation implies the term's γ value, with red shades indicating probabilities above 50 percent.

Table 4. Subjects

Subject	Number of Topics
Recessions	5
Growth	7
Inflation	2
Investment	3
Labor Market	7
Government Revenue	3
Government Expenditure	4
Trade	7

(a) Topic 71

(b) Topic 40



(c) Topic 89

(d) Topic 25

Figure 5. Topics of the labor market subject.

All terms are automatically translated using DeepL. Different German terms might be translated into identical English terms causing duplicate appearances. The size of a term and its position indicate its ϕ probabilities. Only terms with ϕ values that exceed 20% of the topics' highest ϕ value are plotted to a limit of 25. The color saturation implies the term's γ value, with red shades indicating probabilities above 50 percent.

mer topic measures the discourses concerning labor market policy and the latter concerns labor compensation. As I investigate the priorities between broader subjects in this study, I aggregate topics all topics with labor market-related themes into a single aggregated topic, measuring the proportion of the entire subject.

5 Analysis of business cycle-dependent subjects

5.1 The approach

The upcoming section examines the association between forecasters' priorities and various business cycle stages. To achieve this objective, I begin by presenting the aggregated topics and estimating correlation coefficients between the topic proportions of the institutes and relevant macroeconomic variables. I calculate these correlations considering up to two leads and lags, allowing for a comprehensive analysis of forecasters' priorities.

To ensure the robustness of the coefficients, I split the sample and perform several subsample analyses. These subsample analyses aim to find potential variations and sensitivities in the results, providing valuable insights into the consistency and generalizability of my findings. Lastly, I conduct a multivariate regression analysis, which serves the purpose of isolating the impact of each aggregate topic on the recession indicator.

Given my primary research focus on analyzing forecasters' subject priorities in different business cycle stages, I will only present topics that consistently demonstrate consistent and significant correlations with the recession indicator in this section.

5.2 Peaks and contractions

The analysis starts with topics that positively correlate with the recession dummy, thus associating with peaks or contractions. The recession topic in Figure 6 combines five recession-themed topics with high γ or θ values for terms such as "recession," "insecurity," or "crisis." The latter terms exhibit gamma proportions of above 50%, as the color saturation of the word cloud in the top left shows. The conditional γ probability of the term "recession" getting assigned to the recession topic is below 50%, as forecasters use the term frequently when discussing other subjects. The top right subfigure displays recession periods, the topic's cyclical component for all institutes, and the cyclical component of an aggregated macroeconomic variable (economic growth for this topic).

The middle and bottom subfigures show the topic's correlation coefficients with various lags and leads of different key macroeconomic variables. The coefficients between the topics and the macroeconomic variables at 0 half-years show that all institutes prioritize the recession subject when the current economy is in a recession half-year, as the middle-left subfigure shows. The coefficients at -1 and -2 half-years show an equally or even slightly stronger correlation between forecasters' prioritizing recessions and recessions one or two half-years ago.

Only the ifo and the IfW institutes significantly prioritize recessions in their forecast reports when the economy shrinks in the upcoming one or two half-years, as their coefficients at 1 and 2 half-years show. Thus, all institutes' recession topics correlate more strongly with past recession half-years than with current or upcoming recession half-years suggesting that most of the time, the economy probably was already in a recession when institutes prioritize discussions of recessions. A visual analysis of the top right subfigure confirms that the recession topic usually peaks during or after recessions. The institutes prioritized recession-related matters exclusively before the recession following the second oil price shock. One would expect clear preceding recession-related content in forecast reports if forecasters accurately assess potential recession risks. Therefore, forecasters' quantified priorities corroborate the literature's finding that forecasters mostly fail to forecast recessions in advance (Fildes and Stekler 2002; Heilemann and Stekler 2013).

While qualitative forecasts miss the recession's beginning, the other correlation coefficients

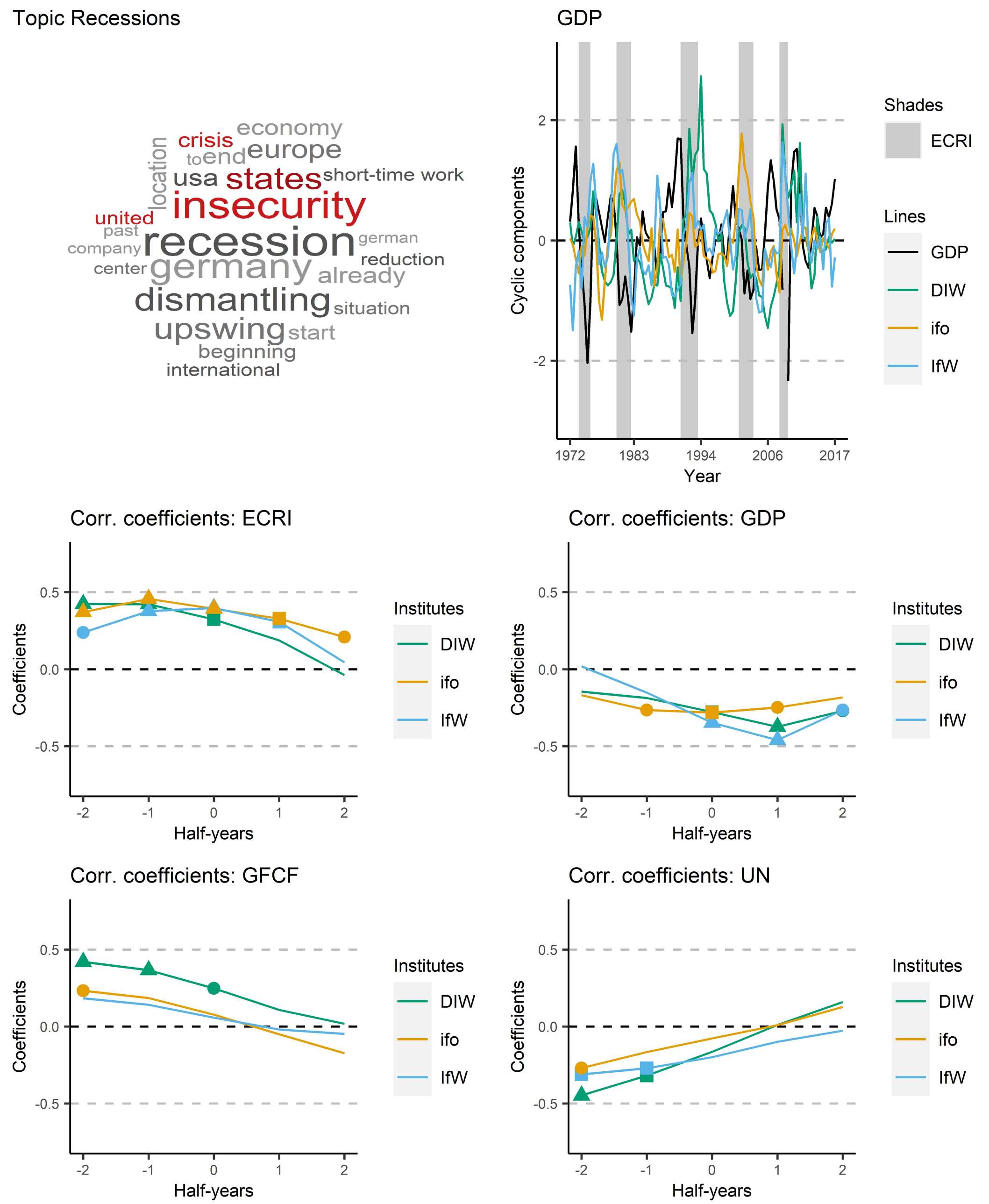


Figure 6. Topic Recessions

Word cloud (top left subfigure): See caption below Figure 3 - Visualization (top left subfigure): Cyclic components greater than three or smaller than minus three are not plotted. - Correlation plot (others): Correlation coefficients are calculated using Pearson for numeric variables and Kendall rank for the recession dummy. The x-axis depicts the macroeconomic variables' leads (positive values) or lags (negative values) in half-years. Symbols indicate p-values triangle: $p < 0.001$, rectangle: $p < 0.01$, circle: $p < 0.05$.

show that economic conditions worsen in the year after forecasters attribute high importance to the recession topic. The cyclical components of growth and the investment rate have their lowest coefficient in the subsequent half-year. The unemployment rate is significantly below-trend in the preceding periods while turning insignificantly for the current and future periods. Although the correlations with the current and future unemployment rates are insignificant, the rising coefficients following significant below-trend levels indicate rising unemployment.

Even though forecasters miss the turning point when discussing recessions, the institutes might anticipate further large decreases in economic activity. Thus forecasters' discussions of recession-related issues neither signal the beginning nor the end of a recession.[13]

The topic that captures the end of the growth phase most significantly is the inflation topic in Figure 7. The topic correlates strongly with upcoming recessions and shows declining correlation coefficients for the growth rate. Thereby, all institutes' correlation coefficients with recessions one year ahead exceed their respective coefficients of the recession topic. Furthermore, the topic significantly correlates with the absence of recessions or above-trend growth rates in the past. Therefore, forecasters discuss inflation-related matters primarily before peaks at the end of an expansion phase and deprioritize the subject once the turning point lies in the past. The cyclical components' visualization confirms that the topic consistently precedes recessions, with the only exception being the early 90s recession following German Reunification. It is reasonable to assume that forecasters valued the importance of this unique historical event and its overall impacts on the German economy higher than inflation-related matters, relatively displacing the inflation topic in their forecasting reports.

Notably, the topic indicates upcoming above-trend inflation rates less strongly, as the lower and less consistent correlation coefficients show. Despite the relatively weak correlation, forecasters allocate a substantial proportion of their reports to inflation when they anticipate a high inflation rate. However, it is important to note that such anticipated inflation triggers a response from monetary policymakers, as evidenced by the correlation between the topic of inflation and both current and future above-trend interest rates. During the sample period, each increase in interest rates was followed by an economic recession. This reaction by monetary policymakers to rising inflation likely served to moderate price increases, thereby explaining the relatively weak correlation between inflation topic and actual inflation rates.

The pattern's consistency raises questions about why forecasters neither predicted the recessions nor dedicated more significant parts of their reports to discussing recession risks. Imminent before the recession, their reports focus primarily on inflation, implying that forecasters might systematically overestimate inflation risks and underestimate growth risks during phases of contractive monetary policies.

This hypothesis finds support in the investigations of the FOMC protocols around the Great Recession by Stekler and Symington (2016), which assert that the central bankers were primarily worried about inflation instead of GDP in advance of the recession. The authors state that the committee explicitly recognized the downside risks to growth throughout 2007 but considered inflation of more concern until the beginning of the recession in December 2007, when they saw increasing risk for growth and economic uncertainty. In the second quarter of 2008, they shifted back to finding the downside risks for GDP decreased and became increasingly more worried about inflation again. The focus on inflation worries reached its high in August 2008. Their primary concern switched back to GDP only with the collapse of Lehman Brothers in September 2008. In October 2008, their inflation worries finally diminished.

While Stekler and Symington (2016) confirm the overestimation of inflation risks, the authors do

13. To reduce information overload, I do not discuss correlations that do not change the overall interpretation of a topic. The inflation and the real interest rates show declining correlation coefficients, as one would expect considering the presented coefficients.

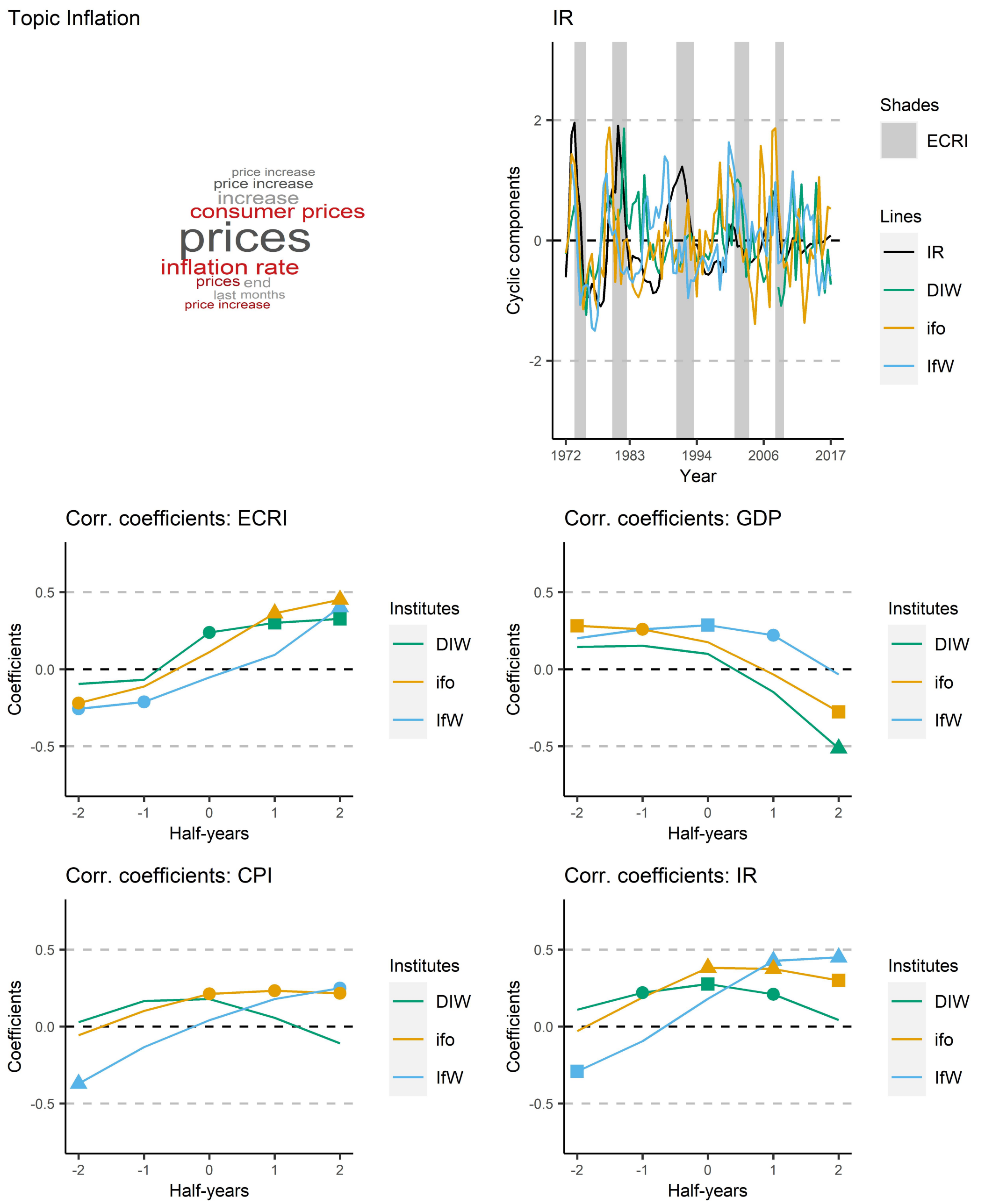


Figure 7. Topic Inflation

Word cloud (top left subfigure): See caption below Figure 3 - Visualization (top left subfigure): Cyclic components greater than three or smaller than minus three are not plotted. - Correlation plot (others): Correlation coefficients are calculated using Pearson for numeric variables and Kendall rank for the recession dummy. The x-axis depicts the macroeconomic variables' leads (positive values) or lags (negative values) in half-years. Symbols indicate p-values triangle: $p < 0.001$, rectangle: $p < 0.01$, circle: $p < 0.05$.

not connect these observations to a failure to recognize the recession. Instead, they state that the committee, for the most part, assessed the risks upfront of the recession correctly and note: "With hindsight, we cannot fault the forecasting process, but the FOMC did make one serious forecasting error: their failure to predict the Great Recession." They provide multiple potential explanations for this miscast: inaccurate real-time GDP data, general reluctance to forecast recessions to avoid self-fulfilling forecasts, overestimating effects from already implemented fiscal- and monetary policies to aid the economy, and political constraints.

While these points are valid explanations, the recurring pattern of forecasters concentrating on inflation while failing to predict recessions indicates a connection between both phenomena that might further explain the miscasts. A potential explanation for a systematic failure to predict recessions when expecting inflation lies in the interaction of imperfect expectations and uncertainty about monetary policy transmission delays. Angeletos, Huo, and A. Sastry (2021) show that forecasters' inflation expectations sluggishly respond to shocks initially, apparently due to noisy information and overextrapolation. After the first underreaction, forecasters' beliefs about inflation tend to overreact and overshoot the actual outcomes.

Additionally, Havranek and Rusnak 2012 find that economists' estimations for monetary policy's transmission lags are highly heterogenous, reaching from 25 to 50 months with an average of 29 months. Furthermore, the authors find that transmission lags increase with greater financial development, which might give forecasters and policymakers the impression that policy effects are already in place. Together imperfect expectations and uncertainty of transmission lags unfold a hazardous potential for monetary policy overreactions, inducing or amplifying economic downturns. Due to the sluggish expectation adaptation, they appear suddenly for forecasters still concerned with inflation's upside risks.

5.3 Troughs and expansions

The topic associated with economic expansions is the growth topic in Figure 8, which encompasses broad terms such as "economy," "gross domestic product," and "production." Word cloud terms like "stronger" and "increase," together with the correlation coefficients, indicate strong growth. The topic's coefficients for GDP and the ECRI index are most significant in the current period, while the DIW's topic exhibits a stronger correlation with growth rates in future periods. The current unemployment rate remains above trend, reflecting lingering signs of past economic weakness. However, the significant negative coefficients for future unemployment suggest that the rate is likely to decline as economic growth unfolds. Despite high growth and falling unemployment, the coefficients for the inflation rate remain below zero; thus, the economy shows no apparent signs of overheating.

Forecasters place greater emphasis on qualitative evaluations of growth in the midst of an ongoing expansion, rather than at the early stages of a recovery. As a result, the growth topic largely resembles the recession topic, since it captures information that remains relevant beyond the turning point rather than anticipating it.

Instead, forecasters place greater emphasis on investment-related matters in their textual analysis around economic troughs, as illustrated in Figure 9. The cyclical components of investment topic tend to peak toward the end of recessions, as evidenced by a shift from positive to negative correlations with the recession dummy. This turning point is further supported by rising growth coefficients and the visual pattern of the topic's cyclical component, which typically surges just before the end of most recessions. Coefficients for the unemployment rate remain positive, consistent with the observation that unemployment usually peaks shortly after a recession ends.

Forecasters' increased emphasis on investment around economic troughs may suggest that they are detecting early signs of a recovery in investment. However, the investment topic is strongly negatively correlated with the investment rate, indicating that investment typically declines more

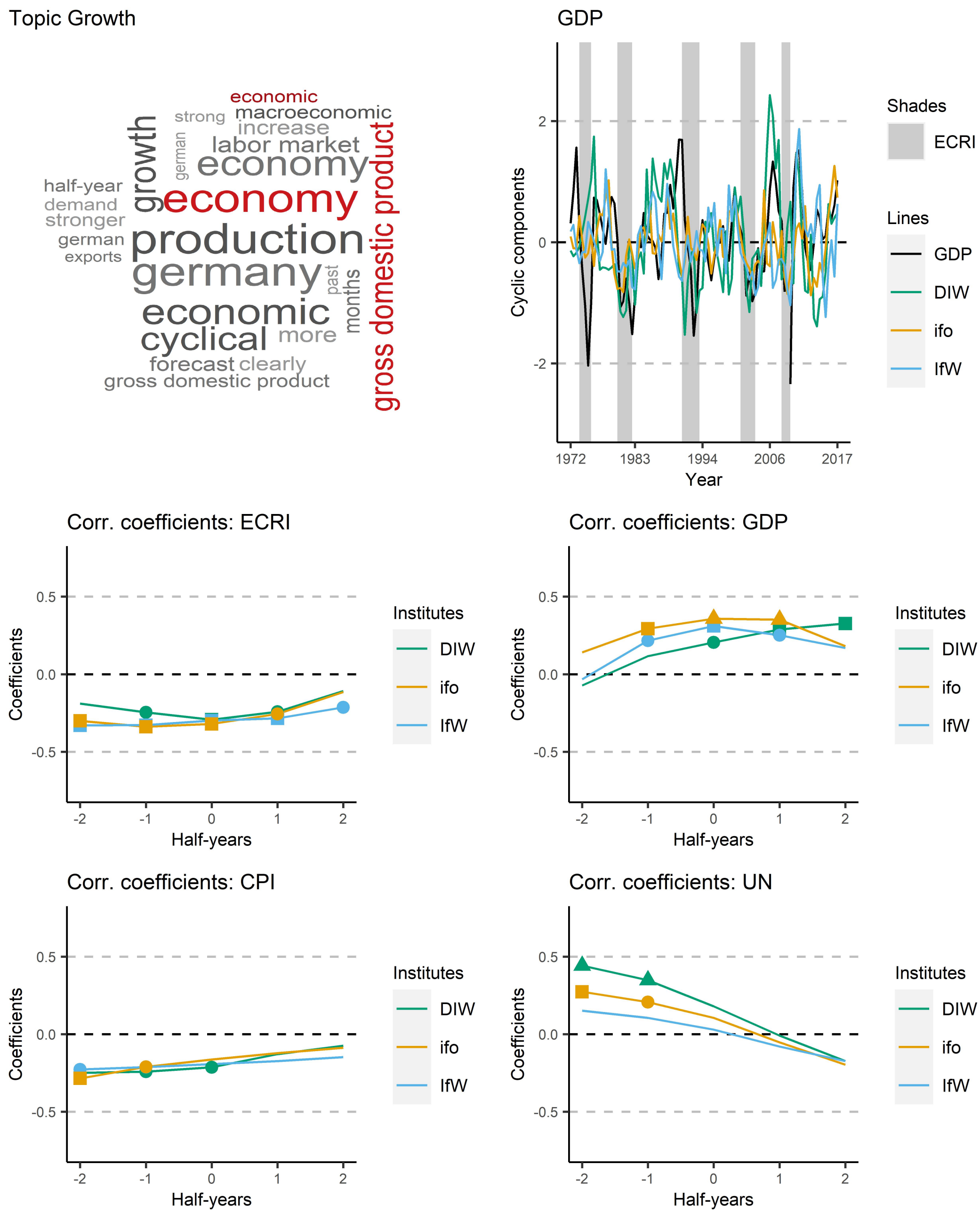


Figure 8. Topic Growth

Word cloud (top left subfigure): See caption below Figure 3 - Visualization (top left subfigure): Cyclic components greater than three or smaller than minus three are not plotted. - Correlation plot (others): Correlation coefficients are calculated using Pearson for numeric variables and Kendall rank for the recession dummy. The x-axis depicts the macroeconomic variables' leads (positive values) or lags (negative values) in half-years. Symbols indicate p-values triangle: $p < 0.001$, rectangle: $p < 0.01$, circle: $p < 0.05$.

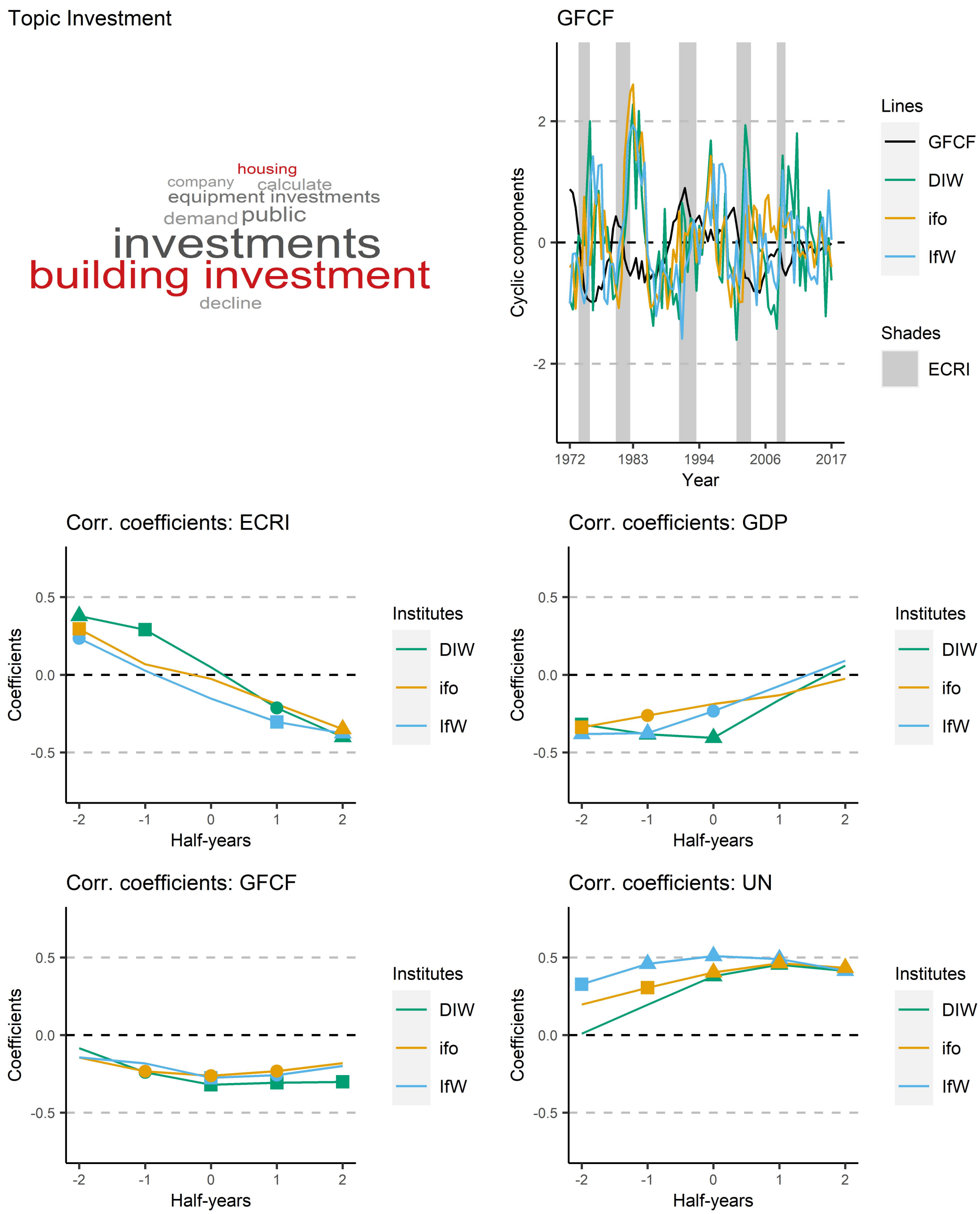


Figure 9. Topic Investment

Word cloud (top left subfigure): See caption below Figure 3 - Visualization (top left subfigure): Cyclic components greater than three or smaller than minus three are not plotted. - Correlation plot (others): Correlation coefficients are calculated using Pearson for numeric variables and Kendall rank for the recession dummy. The x-axis depicts the macroeconomic variables' leads (positive values) or lags (negative values) in half-years. Symbols indicate p-values triangle: $p < 0.001$, rectangle: $p < 0.01$, circle: $p < 0.05$.

sharply than other GDP components and shows no clear signs of recovery at that stage. Additionally, the topic's word cloud includes terms such as "decline," reflecting a negative sentiment toward investment. These findings imply that recoveries during the sample period were not driven by investment, reducing the likelihood that forecasters observed early signs of a rebound in aggregate investment.

An alternative explanation is that forecasters were pessimistic about investment precisely because it underperformed relative to other GDP components. In this case, the relative importance of the investment topic would rise not because investment was recovering, but because forecasters placed less emphasis on other, improving areas of the economy. This would suggest that forecasters recognized early signs of recovery while still highlighting weak spots, thereby demonstrating an awareness of the specific economic conditions at the time and indirectly anticipating the end of the recession.

It is not possible to definitively explain forecasters' emphasis on investment during economic troughs, nor to determine whether they identified early signs of recovery, based on this quantitative finding alone. Future research could offer more nuanced insights into this recurring pattern by qualitatively examining forecasters' investment- and growth-related expectations around economic turning points.

5.4 Subsample analyses

The 45-year sample period encompasses significant developments in economic theory, forecasting methods, and changes in institutional leadership. However, testing the robustness of correlations with respect to theoretical or methodological advances is not feasible, as this would require precise information on when such changes were implemented by forecasters. As noted by Döpke, Fritsche, and Waldhof (2019), German forecasting institutes have predominantly continued to rely on traditional approaches, with limited adoption of more recent methodological innovations such as machine learning and DSGE models. Similarly, assessing the impact of changes in institutional leadership is not viable, as the resulting subsample periods would be too short to yield meaningful insights. Instead, I use two major structural events - German reunification and the introduction of the Euro - as breakpoints for subsample analysis. The goal is to examine whether the correlation between the recession indicator and key topical priorities changed across these periods.

Figure 10 depicts a sample split at the German Reunification, approximately dividing the sample in half. Although the levels and significances of the coefficients exhibit slight variations, they mostly demonstrate a similar qualitative trend, indicating that forecasters discussed the topics during comparable stages of the business cycle in both periods. However, the growth topic stands as a significant exception, with insignificant coefficients for future recessions following the Reunification. As the topic seems to focus on the current period primarily, this difference might arise from smaller time intervals between recessions.

Figure 11 illustrates a sample split with the introduction of the Euro. The second subsample analysis further confirms that the textual patterns did not undergo substantial changes over the sample period, except for the growth topic that further loses predictive power. However, the significance of investment and inflation topics increases in the later subsample. These enhanced capabilities in identifying turning points might be attributed to the significant increase in the length of forecasting reports since the 2000s. More lengthy and thus more detailed reports favor specific rather specific and repress more general topics, as shown by Foltas (2022). Accordingly, the more general recession and growth topics reveal weaker coefficients and more variance between the different institutes.

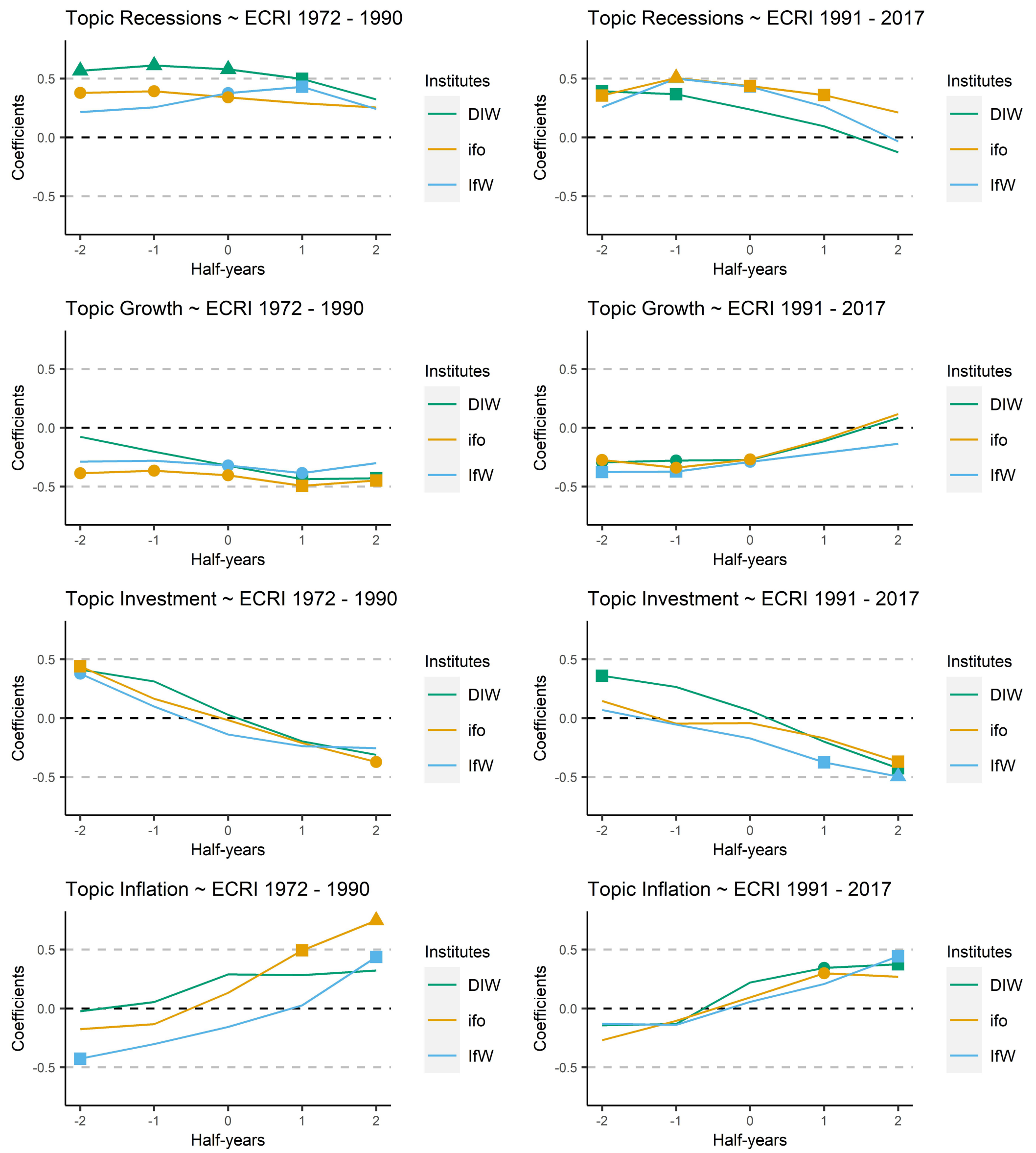


Figure 10. Robustness Check: German Reunification

Correlation coefficients are calculated using Pearson for numeric variables and Kendall rank for the recession dummy. The x-axis depicts the macroeconomic variables' leads (positive values) or lags (negative values) in half-years. Symbols indicate p-values triangle: $p < 0.001$, rectangle: $p < 0.01$, circle: $p < 0.05$.

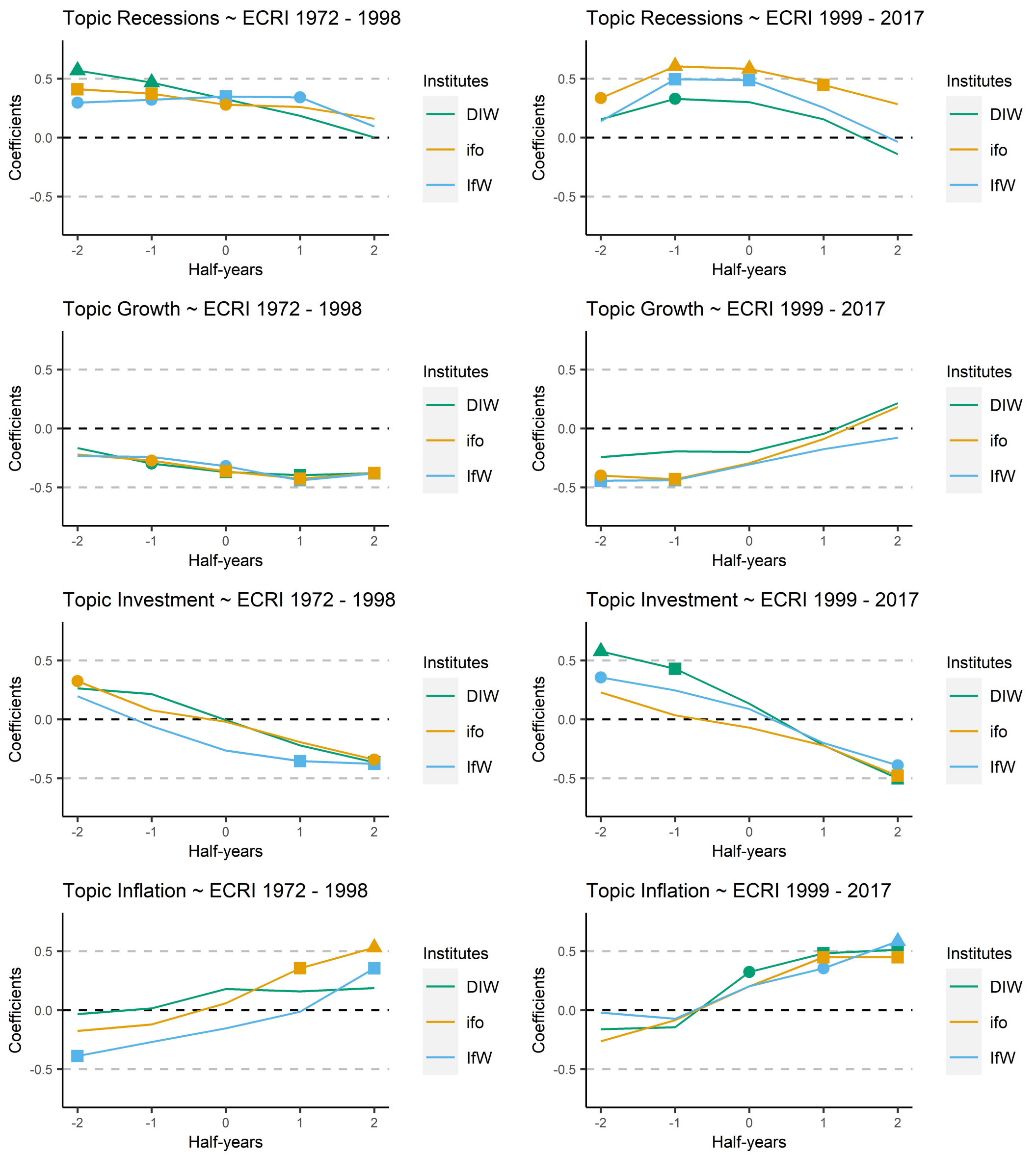


Figure 11. Robustness Check: Euro Introduction

Correlation coefficients are calculated using Pearson for numeric variables and Kendall rank for the recession dummy. The x-axis depicts the macroeconomic variables' leads (positive values) or lags (negative values) in half-years. Symbols indicate p-values triangle: $p < 0.001$, rectangle: $p < 0.01$, circle: $p < 0.05$.

Table 5. Probit analyses

	ECRI lags (negative values) and leads (positive values) in half-years				
	-2	-1	0	1	2
Topic Recessions	1.435^{***}	1.678^{***}	1.329^{***}	1.085^{***}	0.176
	(0.421)	(0.430)	(0.382)	(0.407)	(0.443)
Topic Growth	-1.394^{**}	-1.553^{***}	-1.550^{***}	-1.763^{***}	-2.017^{***}
	(0.575)	(0.557)	(0.515)	(0.566)	(0.669)
Topic Inflation	-1.157^{**}	-0.820^{*}	0.409	0.968^{**}	1.888^{***}
	(0.539)	(0.477)	(0.326)	(0.404)	(0.566)
Topic Investment	0.754^{***}	0.141	−0.268	-1.108^{**}	-2.235^{***}
	(0.289)	(0.271)	(0.284)	(0.432)	(0.700)
Constant	-1.356^{***}	-1.184^{***}	-0.973^{***}	-1.071^{***}	-1.507^{***}
	(0.300)	(0.258)	(0.201)	(0.225)	(0.362)
Observations	90	91	92	91	90
Log Likelihood	−30.021	−30.523	−33.732	−29.782	−22.875
Akaike Inf. Crit.	70.043	71.045	77.463	69.564	55.751
Note:	*p<0.05; **p<0.01; ***p<0.001				

5.5 Multivariate analysis

After conducting the individual analysis of forecasters' cycle-dependent priorities, I incorporate the topics into a multivariate probit model to isolate each topic's relationship with the recession indicator. For this purpose, the topics' cyclical components are averaged across the institutes and used as independent variables in five models, each employing different lags or leads of the recession dummy. I present the results in Table 5.

The averaged recession topic most strongly relates to recessions in the past and becomes insignificantly for recessions two half-years ahead, confirming the previously found lagging prioritization of recessions in business cycle reports. The coefficients of the growth topic are significantly negative across all periods. The declining coefficients suggest that forecasters prioritize growth forward-looking, which may initially seem contradictory to previous results. However, the coefficients validate my main finding that forecasters prioritize growth in the middle of expansion phases, not in proximity to turning points.

Forecasters' prioritizing inflation significantly relates to expansion in the past and recession phases in the future, with the topics' significance increasing with the recession indicators of leads and lags. Accordingly, forecasters' prioritizing investment topics mark the recession's end and an upcoming upswing. Thus, both combined topics confirm their previously found association with turning points in a multivariate analysis.

6 Insights into growth-independent subjects

This section presents topics with no or only subordinated correlations with growth or the recession indicator, thus not providing substantial insights into forecasters' priorities around economic turning points. As these topics are not of my primary research interest, I discuss them only briefly to limit the scope of this paper and highlight potential use cases for quantifying forecasters' priorities.

6.1 Unemployment

The DIW and the IfW prioritize the labor market when unemployment rates are above trend levels, as Figure A1 depicts.[14] Although growth and unemployment are connected through Okun's law, only the IfW's topic demonstrates a relatively weak negative correlation with growth. Research investigating the linkage between growth and unemployment shows that Okun's law is not a strict relationship but rather a "rule of thumb" that changes over time and has short-term exceptions (Knotek 2007). For Germany, Nebot, Beyaert, and García-Solanes (2019) estimates the relationship between growth and unemployment to be weak, potentially attributable to firms' reluctance to lay off trained workers ("labor hoarding hypothesis") and institutional restrictions on worker layoffs ("institutional rigidity hypothesis"). Thus, the weak correlation of the labor market topic with economic growth can be explained by the weakness of Okun's law in Germany and its changes over time.

While the IfW strongly prioritizes the labor market in anticipation of rising unemployment, the DIW's topic appears to be more backward-looking, and the ifo's topic shows no correlation with unemployment rates. When faced with institutional divergences, it provides further insights to investigate the aggregated topics' components. Figure A2 illustrates the wage-themed Topic 25, which is aggregated into the labor market topic. The figure reveals a negative correlation between ifo's topic and unemployment rates. While the DIW and IfW tend to deprioritize labor market-related issues during periods of lower unemployment, the ifo could instead shift its emphasis from unemployment to wages, preserving the relative length of their labor market subject

In conclusion, researchers may discover interesting patterns not only by comparing forecasters' subject priorities but also by conducting more detailed analyses that investigate the emphases within specific subjects. By undertaking fine-grained future analysis, future researchers could enhance the understanding of how forecasters process information and identify potential areas of miscasts.

6.2 Fiscal policy

Figure A3 and A4 measure forecasters' prioritization of government revenues and expenditures, which could be seen as different aspects of a fiscal policy subject. The institutes do not vary their fiscal policy priorities according to government spending except for the DIW, whose government spending topic negatively correlates with actual upcoming expenditure decreases.

The prioritization of either income or spending aspects of fiscal policy appears to be influenced by the prevailing unemployment rate. Specifically, the IfW and ifo institutes tend to emphasize taxes and social contributions when unemployment rates are high. This emphasis implies that these institutes may be inclined to recommend reforms in those areas to address and mitigate the challenges associated with elevated unemployment. On the other hand, the DIW does not exhibit the same behavior as IfW and ifo.

Forecasters demonstrate a greater tolerance for government spending when unemployment rates are elevated, as evidenced by their reduced prioritization of this matter. This change in focus

14. All three institutes substantially prioritized the labor market in the early 2000s. This coincided with a series of major labor market reforms known as the "Agenda 2010", which aimed to reduce unemployment. These reforms generated intense discussions among German economists and experts, thus providing an explanation for the notable emphasis placed on the labor market topic during that time period.

potentially signifies the institutes' recognition of government expenditures as a tool for economic stabilization and countercyclical policies.

Overall, investigating forecasters' policy priorities during specific economic situations can reveal the different fundamental beliefs held by the institutes, thus enabling researchers to measure and map the economic policy landscape more precisely.

6.3 Trade

Figure A5 depicts the prioritization of trade within forecast reports. All institutes prioritize trade before more favorable German trade balances, demonstrating their ability to correctly anticipate increases in global demand for German goods and services. However, there is no correlation between the topic and economic growth, even though trade has been a crucial factor for German growth (Dustmann *et al.* 2014).

Further, the reasons for the positive correlation between prioritizing trade and actual trade surpluses are unclear, as it is not evident why forecasters would deprioritize foreign trade when expecting the German account balance to shrink. Hence, a negative correlation or no correlation at all between the trade topic and the balance of trade is also plausible.

There are several potential explanations for the positive correlation between prioritizing trade and its topic:

- Forecasters may be less successful in predicting decreasing trade balances, leading to a lack of prioritization of trade before downward shifts in the balance.
- Decreasing global demand could cause forecasters to allocate higher proportions to assessments of dangers and uncertainties for the broader German economy, which are covered by the recession topic.
- Forecasters predicting a decline in demand for German products and services could trigger policy reactions aimed at enhancing German competitiveness, thereby preventing unfavorable trade balances.

The trade topic shows that further investigations are needed to fully understand the complex relationship between forecasters prioritizing trade, the balance of trade, and economic growth.

7 Conclusion

I introduced a novel approach to uncover embedded information within forecast reports by measuring forecasters' subjects' priorities. These priorities reflect forecasters' perceptions of the importance of various macroeconomic subjects within the current business cycle phase. As such, quantified forecasters' priorities offer an additional layer of valuable information in forecast reports, complementing forecasters' quantitative predictions and sentiment.

I expanded upon previous studies using the Word2Sense-LDA topic model to identify latent topics in forecast reports (e.g. Foltas 2022) to extract forecasters' priorities of three German forecasting institutes from 1970 to 2017. Firstly, I introduced the concept of subjects, which allows classifying topics with related themes under a principal theme, enabling the aggregation of topics, which enhances coherence while achieving a desired level of specificity. Secondly, I decomposed proportions into trend and cyclic components, with the latter serving as a measure of forecasters' priorities. Thirdly, I proposed a novel approach that optimizes the topic model's hyperparameters to capture the embedded information in the reports effectively by leveraging business cycle forecasts.

Through analyzing the priorities of forecasters in different business cycle phases, I validate the literature's findings that forecasters typically fail to predict recessions (Fildes and Stekler 2002; Heilemann and Stekler 2013) as the institutes do not prioritize recession-related matters in advance of economic downturns. Instead, I reveal that forecasters primarily focus on inflation around the peak, affirming the observations made by Stekler and Symington (2016) in their investigation of FOMC protocols surrounding the Great Recession. The pattern's robustness suggests a systemic underestimation of the impact of contractive monetary policies' on GDP. This forecasting error could be explained by increasing transmission lags associated with greater financial development (Havranek and Rusnak 2012) and sluggish expectations (Angeletos, Huo, and A. Sastry 2021).

Around troughs, forecasters allocate substantial portions of their reports to aggregated investment. Simultaneously, investment rates typically reach their lowest levels during the early stage of a recovery. This consistency between textual prioritization and macroeconomic data indicates that forecasters have a better understanding of the specific economic circumstances around troughs. However, it remains unclear whether forecasters correctly recognize the recession's end without further investigations.

Additionally, I identified the labor market, fiscal policy, and foreign trade as promising areas for priority-based investigations. Analyzing how forecasters shift their focus within these subjects during macroeconomic fluctuations, or how institutions differ in their approaches, could yield valuable insights into the forecasting process and ultimately help improve forecast accuracy.

Data availability statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work, I used DeepL and Grammarly to improve the clarity and readability of the text. After using these tools, I reviewed and edited the content as needed and take full responsibility for the content of the publication.

References

Angeletos, G.-M., Z. Huo, and K. A. Sastry. 2021. "Imperfect Macroeconomic Expectations: Evidence and Theory." *NBER Macroeconomics Annual* 35 (1): 1–86.

Blair, S. J., Y. Bi, and M. D. Mulvenna. 2020. "Aggregated topic models for increasing social media topic coherence." *Applied Intelligence* 50:138–156. https://doi.org/10.1007/s10489-019-01438-z.

Blei, D. M., and J. D. Lafferty. 2006. "Dynamic topic models." In *ICML '06: Proceedings of the 23rd international conference on Machine learning,* edited by Association for Computing Machinery, 113–120. https://doi.org/10.1145/1143844.1143859.

Blei, D. M., A. Y. Ng, and M. I. Jordan. 2003. "Latent Dirichlet allocation." *Journal of Machine Learning Research* 3:993–1022.

Churchill, R., and L. Singh. 2022. "The Evolution of Topic Modeling." *ACM Computing Surveys* 54 (10s): 1–35. https://doi.org/10.1145/3507900.

Clements, M. P., and J. J. Reade. 2020. "Forecasting and forecast narratives: The Bank of England Inflation Reports." *International Journal of Forecasting* 36 (4): 1488–1500.

Dieng, A. B., F. J. R. Ruiz, and D. M. Blei. 2019. "The Dynamic Embedded Topic Model," arXiv:1907.05545.

———. 2020. "Topic Modeling in Embedding Spaces." *Transactions of the Association for Computational Linguistics* 8:439–453.

Döhrn, R., and C. M. Schmidt. 2011. "Information or Institution?" *Jahrbücher für Nationalökonomie und Statistik* 231 (1): 9–27. https://doi.org/10.1515/jbnst-2011-0103.

Döpke, J., and U. Fritsche. 2006. "Growth and inflation forecasts for Germany a panel-based assessment of accuracy and efficiency." *Empirical Economics* 31 (3): 777–798.

Döpke, J., U. Fritsche, and C. Pierdzioch. 2017. "Predicting recessions with boosted regression trees." *International Journal of Forecasting* 33:745–759.

Döpke, J., U. Fritsche, and G. Waldhof. 2019. "Theories, techniques and the formation of German business cycle forecasts." *Jahrbücher für Nationalökonomie und Statistik* 239 (2): 203–241.

Dustmann, C., B. Fitzenberger, U. Schönberg, and A. Spitz-Oener. 2014. "From Sick Man of Europe to Economic Superstar: Germany's Resurgent Economy." *Journal of Economic Perspectives* 28 (1): 167–188. https://doi.org/10.1257/jep.28.1.167.

Fildes, R., and H. Stekler. 2002. "The state of macroeconomic forecasting." *Journal of Macroeconomics* 24 (4): 435–468.

Foltas, A. 2022. "Testing Investment Forecast Efficiency with Forecasting Narratives." *Journal of Economics and Statistics* 242 (2): 191–222.

Foltas, A., and C. Pierdzioch. 2022a. "Business-cycle reports and the efficiency of macroeconomic forecasts for Germany." *Applied Economics Letters* 29 (10): 867–872. https://doi.org/10.1080/13504851.2021.1896668.

———. 2022b. "On the efficiency of German growth forecasts: an empirical analysis using quantile random forests and density forecasts." *Applied Economics Letters* 29 (17): 1644–1653. https://doi.org/10.1080/13504851.2021.1954594.

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

Havranek, T., and M. Rusnak. 2012. "Transmission lags of monetary policy: A meta-analysis," William Davidson Institute Working Paper No. 1038.

Heilemann, U., and H. O. Stekler. 2013. "Has The Accuracy of Macroeconomic Forecasts for Germany Improved?" *German Economic Review* 14 (2): 235–253. https://doi.org/10.1111/j.1468-0475.2012.00569.x.

Heinisch, K., C. Behrens, J. Döpke, A. Foltas, U. Fritsche, T. Köhler, K. Müller, J. Puckelwald, and H. Reichmayr. 2023. “The IWH Forecasting Dashboard: From Forecasts to Evaluation and Comparison.” *Jahrbücher für Nationalökonomie und Statistik* 0 (0). https://doi.org/10.1515/jbnst-2023-0011.

Hollrah, C. A., S. A. Sharpe, and N. R. Sinha. 2020. “The Power of Narratives in Economic Forecasts.” *Finance and Economics Discussion Series,* no. 001, https://doi.org/10.17016/FEDS.2020.001.

Jones, J. T., T. M. Sinclair, and H. O. Stekler. 2020. “A textual analysis of Bank of England growth forecasts.” *International Journal of Forecasting* 36 (4): 1478–1487.

Jones, T., and W. Doane. 2023. “Package ‘textmineR’,” Functions for Text Mining and Topic Modeling, no. 3.0.5.

Ke, S., J. Montiel Olea, and J. Nesbit. 2022. “Robust Machine Learning Algorithms for Text Analysis.” (Yale Sch. Manag., Yale Univ., New Haven, CT), Unpublished manuscript, http://www.joseluismontielolea.com/lda_2022.pdf.

Kholodilin, K. A. 2005. “Forecasting the German Cyclical Turning Points: Dynamic Bi-Factor Model with Markov Switching.” *Jahrbücher für Nationalökonomie und Statistik* 225 (6): 653–674. https://doi.org/10.1515/jbnst-2005-0606.

Knotek, E. S. 2007. “How useful is Okun’s law?” *Economic Review-Federal Reserve Bank of Kansas City* 92 (4): 73–103.

Köhler, T., and J. Döpke. 2023. “Will the last be the first? Ranking German macroeconomic forecasters based on different criteria.” *Empirical Economics* 64 (2): 797–832. https://doi.org/10.1007/s00181-022-02267-9.

Müller, K. 2022. “German forecasters’ narratives: How informative are German business cycle forecast reports?” *Empirical Economics* 62 (5): 2373–2415. https://doi.org/10.1007/s00181-021-02100-9.

Nebot, C., A. Beyaert, and J. García-Solanes. 2019. “New insights into the nonlinearity of Okun’s law.” *Economic Modelling* 82:202–210. https://doi.org/10.1016/j.econmod.2019.01.005.

Rybinski, K. 2021. “Ranking professional forecasters by the predictive power of their narratives.” *International Journal of Forecasting* 37 (1): 186–204. https://doi.org/10.1016/j.ijforecast.2020.04.003.

Stekler, H. O. 2007. “The future of macroeconomic forecasting: Understanding the forecasting process.” *International Journal of Forecasting* 23 (2): 237–248.

Stekler, H., and H. Symington. 2016. “Evaluating qualitative forecasts: The FOMC minutes, 2006–2010.” *International Journal of Forecasting* 32 (2): 559–570. https://doi.org/10.1016/j.ijforecast.2015.02.003.

Thompson, L., and D. Mimno. 2020. “Topic Modeling with Contextualized Word Representation Clusters,” arXiv preprint arXiv:2010.12626.

Appendix

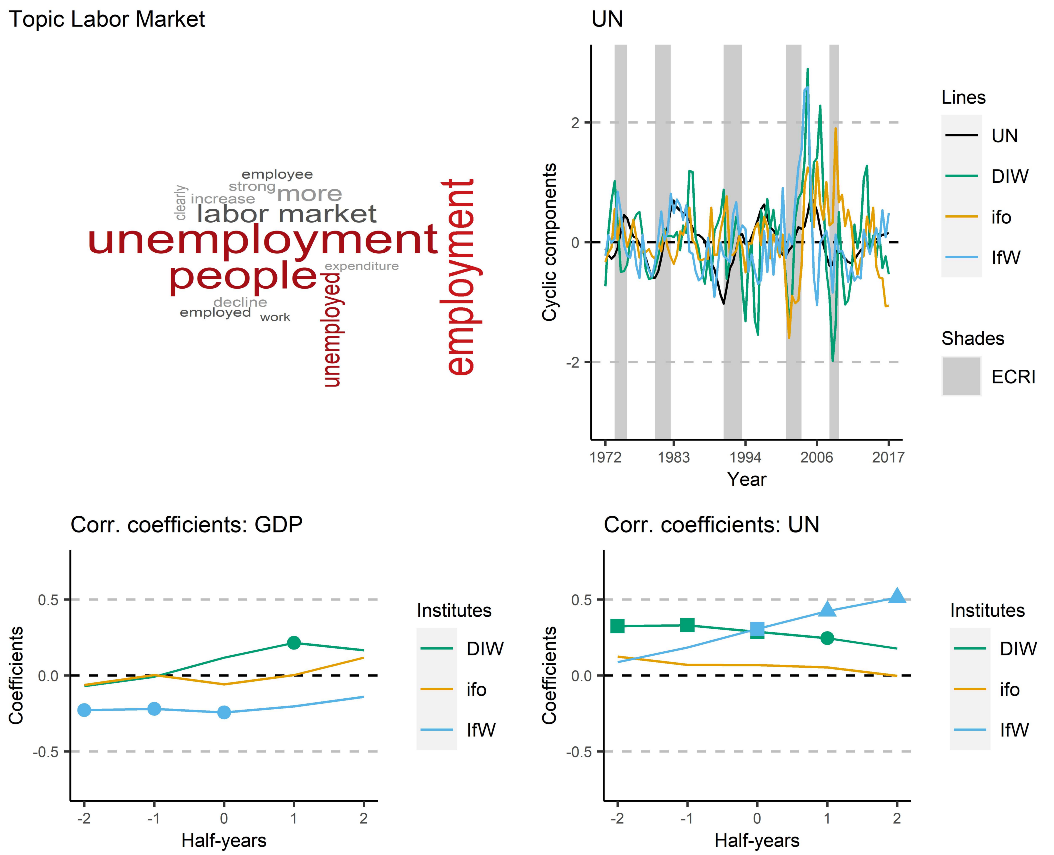


Figure A1. Topic Labor Market
Word cloud (top left subfigure): See caption below Figure 3 - Visualization (top left subfigure): Cyclic components greater than three or smaller than minus three are not plotted. - Correlation plot (others): Correlation coefficients are calculated using Pearson for numeric variables and Kendall rank for the recession dummy. The x-axis depicts the macroeconomic variables' leads (positive values) or lags (negative values) in half-years. Symbols indicate p-values triangle: $p < 0.001$, rectangle: $p < 0.01$, circle: $p < 0.05$.

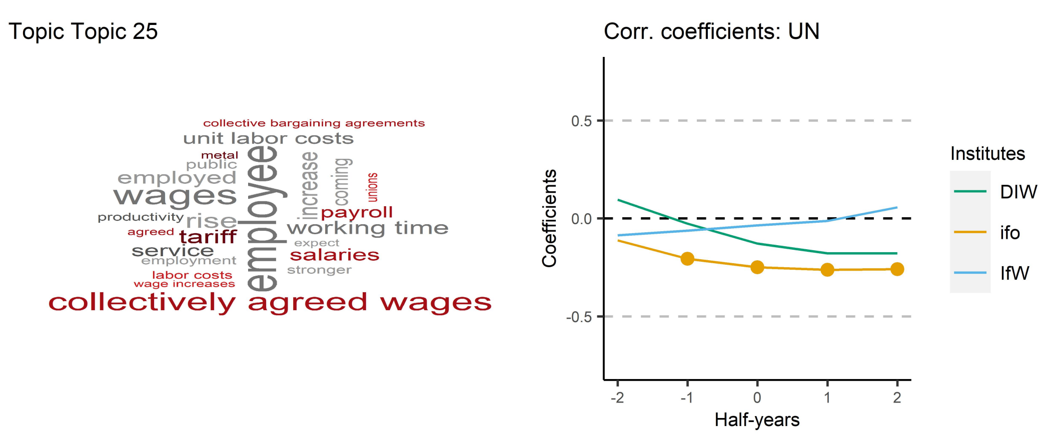


Figure A2. Topic 25
Word cloud (top left subfigure): See caption below Figure 3 - Correlation plot (others): Correlation coefficients are calculated using Pearson for numeric variables and Kendall rank for the recession dummy. The x-axis depicts the macroeconomic variables' leads (positive values) or lags (negative values) in half-years. Symbols indicate p-values triangle: $p < 0.001$, rectangle: $p < 0.01$, circle: $p < 0.05$.

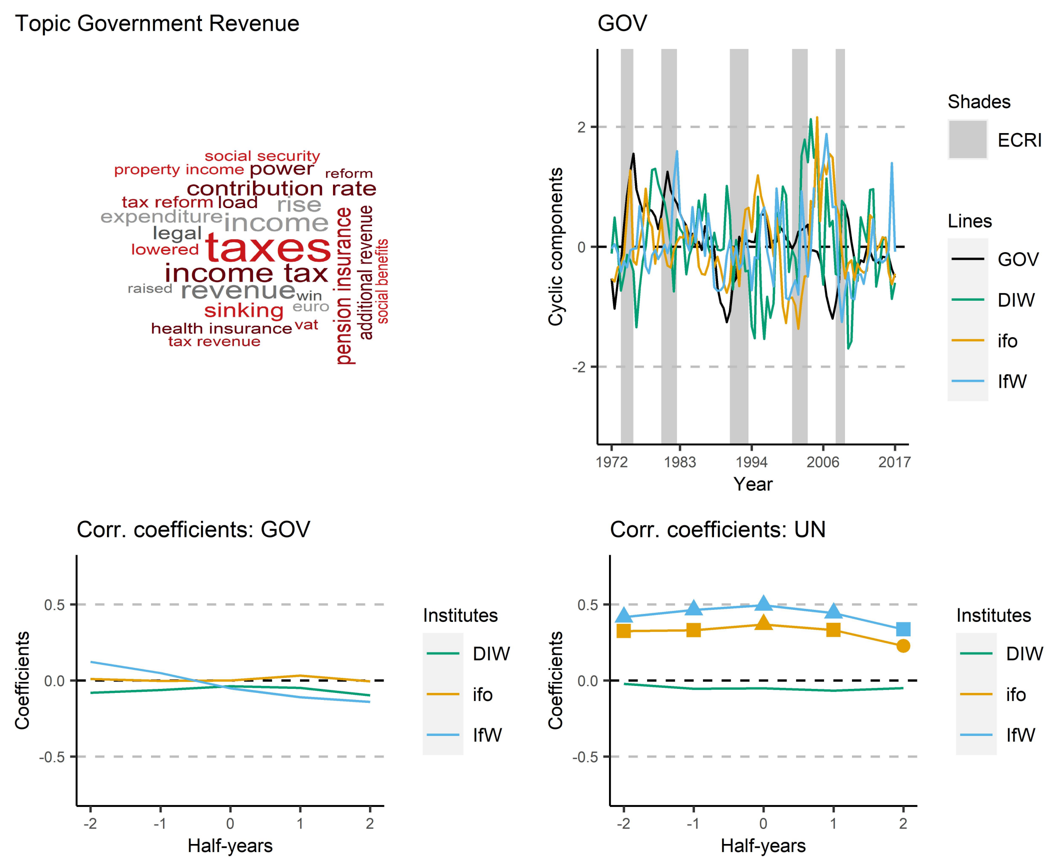


Figure A3. Topic Government Revenue

Word cloud (top left subfigure): See caption below Figure 3 - Visualization (top left subfigure): Cyclic components greater than three or smaller than minus three are not plotted. - Correlation plot (others): Correlation coefficients are calculated using Pearson for numeric variables and Kendall rank for the recession dummy. The x-axis depicts the macroeconomic variables' leads (positive values) or lags (negative values) in half-years. Symbols indicate p-values triangle: $p < 0.001$, rectangle: $p < 0.01$, circle: $p < 0.05$.

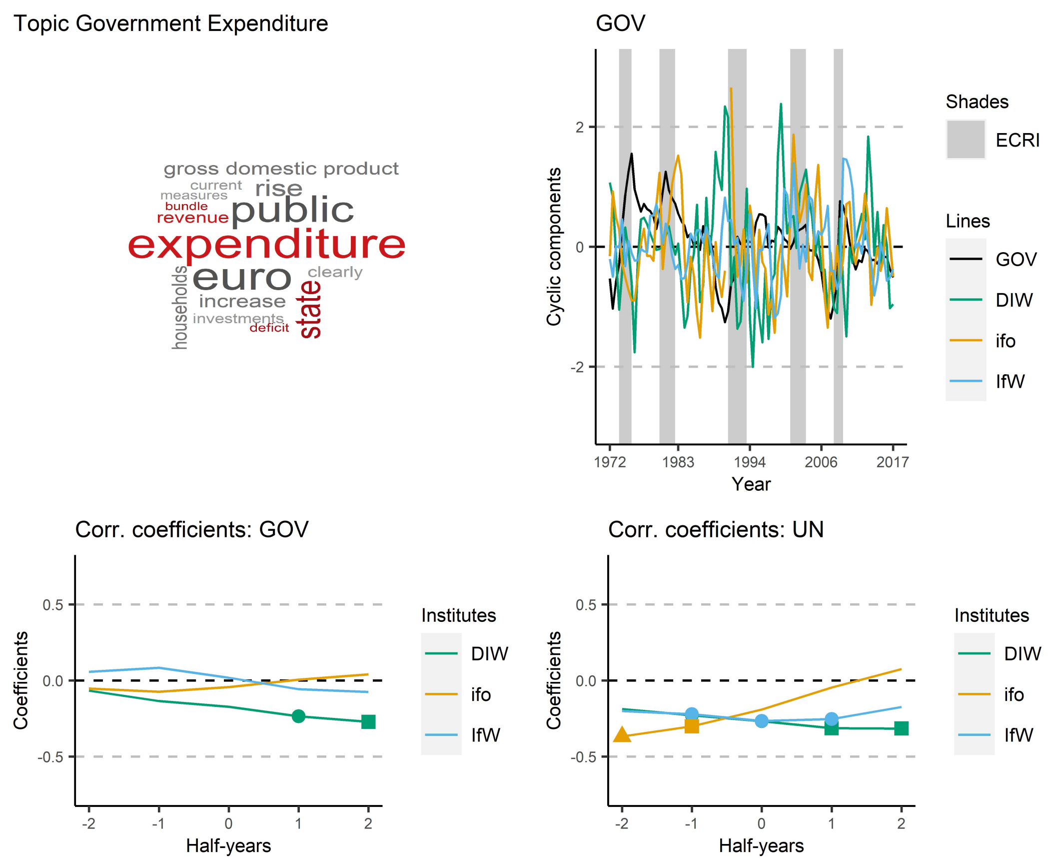


Figure A4. Topic Government expenditure

Word cloud (top left subfigure): See caption below Figure 3 - Visualization (top left subfigure): Cyclic components greater than three or smaller than minus three are not plotted. - Correlation plot (others): Correlation coefficients are calculated using Pearson for numeric variables and Kendall rank for the recession dummy. The x-axis depicts the macroeconomic variables' leads (positive values) or lags (negative values) in half-years. Symbols indicate p-values triangle: $p < 0.001$, rectangle: $p < 0.01$, circle: $p < 0.05$.

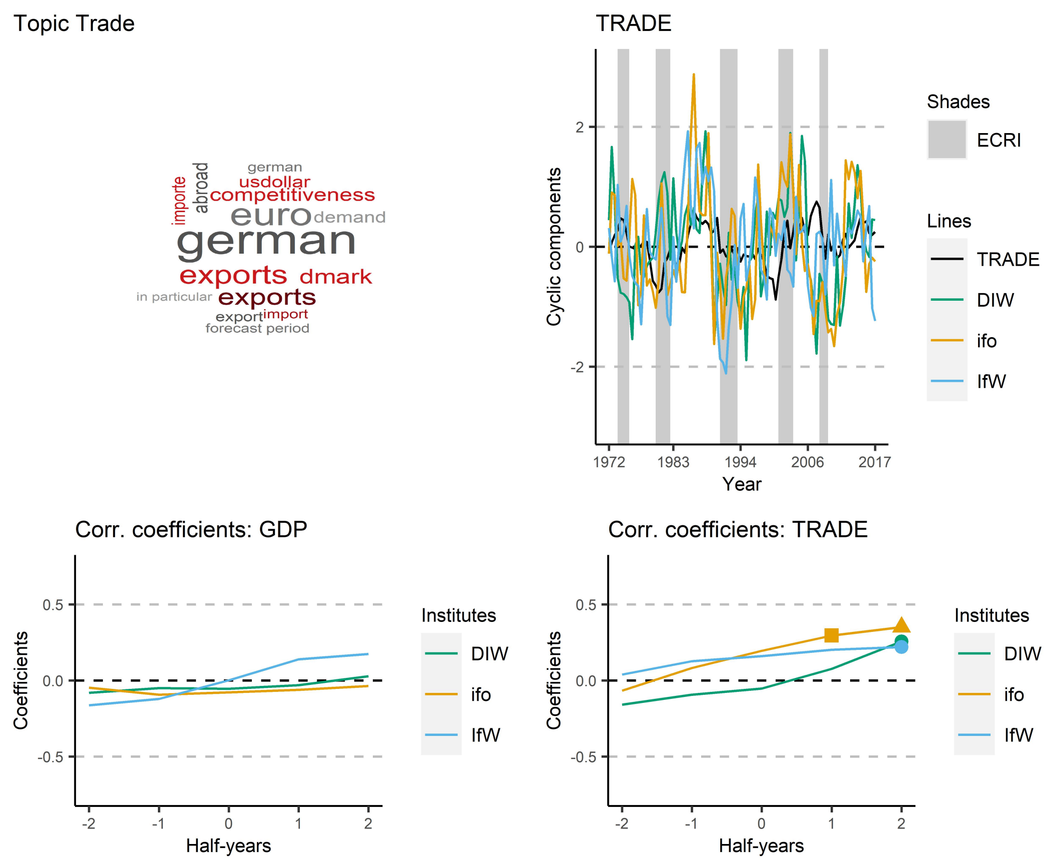


Figure A5. Topic Trade
Word cloud (top left subfigure): See caption below Figure 3 - Visualization (top left subfigure): Cyclic components greater than three or smaller than minus three are not plotted. - Correlation plot (others): Correlation coefficients are calculated using Pearson for numeric variables and Kendall rank for the recession dummy. The x-axis depicts the macroeconomic variables' leads (positive values) or lags (negative values) in half-years. Symbols indicate p-values triangle: $p < 0.001$, rectangle: $p < 0.01$, circle: $p < 0.05$.