Check for
updates

German forecasters' narratives: How informative are German business cycle forecast reports?

Karsten Müller[1] (iD)

Abstract

Based on German business cycle forecast reports covering 10 German institutions for the period 1993–2017, the paper analyses the information content of German forecasters' narratives for German business cycle forecasts. The paper applies textual analysis to convert qualitative text data into quantitative sentiment indices. First, a sentiment analysis utilizes dictionary methods and text regression methods, using recursive estimation. Next, the paper analyses the different characteristics of sentiments. In a third step, sentiment indices are used to test the efficiency of numerical forecasts. Using 12-month-ahead fixed horizon forecasts, fixed-effects panel regression results suggest some informational content of sentiment indices for growth and inflation forecasts. Finally, a forecasting exercise analyses the predictive power of sentiment indices for GDP growth and inflation. The results suggest weak evidence, at best, for in-sample and out-of-sample predictive power of the sentiment indices.

Keywords Textual analysis · Sentiment · Macroeconomic forecasting · Forecast evaluation · Germany

1 Introduction

German business cycle forecast reports offer quantitative point forecasts and qualitative text data for growth and inflation, among other variables. The qualitative texts describe forecasters' views on the macroeconomic situation and development. And, the narratives also express the forecasters' expectations about the future economic development. Using the narratives, the forecasters' expectations can be objectified by

✉ Karsten Müller
    karsten.mueller@hs-merseburg.de

[1] University of Applied Sciences Merseburg, Eberhard-Leibnitz-Straße 2, 06217 Merseburg, Germany

applying textual analysis methods to generate sentiment indices. The key issue is to analyse whether the forecasters' narratives contain additional information beyond the quantified forecasts.

The evaluation of German and international business cycle forecasts has traditionally focused on the analysis of quantitative point forecasts. A large number of existing studies have examined the accuracy and efficiency of German macroeconomic forecasts (see e.g. Heilemann and Stekler 2013; Fritsche and Tarassow 2017; Döpke et al. 2019, and the literature cited therein). Prior research suggests three key insights. First, macroeconomic forecasts for Germany are (mostly) unbiased, but inefficient (see e.g. Döpke et al. 2010; Krüger and Hoss 2012). Second, forecast errors seem to be stable on average over decades which are neither increasing nor decreasing in tendency (Heilemann and Stekler 2013). Third, no forecaster's performance is uniformly superior (Döpke and Fritsche 2006), and there are not significant institutional differences in accuracy across a long time horizon (Döhrn and Schmidt 2011).[1]

Recently, another forecast evaluation approach, which uses qualitative text as data, has become increasingly popular. In this context, textual analysis methods are applied to convert qualitative text data into quantitative scores. The generated indices are used for forecast evaluation tests. Two major strands of the literature can be identified.

One strand will be subsumed here under the term 'elicited forecasts', which was used by Jones et al. (2020). This concept applies a manual scoring procedure to quantify qualitative assessments about the future stance of the economy. Goldfarb et al. (2005) mapped newspaper articles published during the *Great Depression* into an index series using a scoring system to compare the quantified qualitative assessments with numerical forecasts and realized values. A series of forecast evaluation studies applied the developed scoring procedure of Goldfarb et al. (2005) in several contexts to generate *elicited forecasts* to evaluate them (see e.g. Lundquist and Stekler 2012; Stekler and Symington 2016; Mathy and Stekler 2018). The recent analysis of Jones et al. (2020) investigates the Bank of England's growth forecasts using *elicited forecasts* over the period 2005–2015. The more general research question as to whether the text contains additional information for the numerical forecasts is similar to this work. Jones et al. (2020) find that the economic development in the UK is accurately represented by the *elicited forecasts*. Moreover, regression results suggest informational content of the text index in the sense that they can improve the Bank of England's numerical growth nowcasts and one-quarter-ahead forecasts.

A second strand of the literature uses computational text analysis methods to generate text-based sentiment indices. Clements and Reade (2020) and Sharpe et al. (2020) are two seminal related studies. The latter study applies computational text analysis to quantify the 'tonality' (the degree of optimism versus pessimism) of the Federal Reserve Board's Greenbooks and examines whether this measure has predictive power for the economic development over the period 1972–2009. The investigation shows some predictive power of the Greenbook tonality on Greenbook numerical GDP growth and unemployment forecasts, as well as on private GDP forecasts. The latter point implies that the sentiment index also covered policy-relevant information

---

[1] Several studies suggest similar results for example for the US (Batchelor 1990), Japan (Ashiya 2006) or Austria (Fortin et al. 2020).

(Sharpe et al. 2020). Clements and Reade (2020) analyse whether the narratives in the Bank of England's Inflation Reports contain useful information about the future course of GDP growth and inflation between 1997 and 2018. Encompassing tests show some informational content for predicting GDP forecast errors for one and two quarters ahead, but no evidence that sentiment indices are useful to predict forecast revisions. Both studies use the dictionary-based approach to generate sentiment indices, and both studies show that 'an important element of economic forecasting is in the accompanying narrative' (Sharpe et al. 2020, p. 31).

Considering German forecasters' narratives, Fritsche and Puckelwald (2018) analyse the topics of German business cycle forecast reports using generative models. The authors find that textual expressions vary with the business cycle, which is in line with the hypothesis of adaptive expectations. But, a number of questions regarding German forecasters' narratives remain to be addressed.

There is a broader and growing literature in (computational) textual analysis in economics, finance, and accounting (see e.g. Loughran and McDonald 2016; Gentzkow et al. 2019, and the literature cited therein). The following examples give a selective overview of literature that is related to this paper. For example, Shapiro et al. (2020) for the US and Lamla et al. (2020) for Germany use textual analysis tools to create news media sentiment indicators. Both studies has provided evidence for a correlation between news media sentiment indicators and the business cycle and show that sentiment indicators can serve as predictors of the future stance of the economy. Another strand of the literature concerns the predictability of stock market activity. Tetlock (2007), Tetlock et al. (2008) and Garcia (2013) use a dictionary-based approach to generate sentiment indices via news coverage. Loughran and McDonald (2011, 2016) developed a finance-specific dictionary to improve the forecasting performance relative to existing linguistic dictionaries. Jegadeesh and Wu (2013) and Manela and Moreira (2017) apply text regression methods to predict stock market outcomes, while Jegadeesh and Wu (2013) show that text regression-based sentiment indices are superior to sentiment indices based on Loughran and McDonald (2011) dictionary in an out-of-sample forecast environment. The analysis of central bank communication is another topic in text mining. Jegadeesh and Wu (2017) find incremental information value in the Federal Open Market Committee meeting minutes. The authors use a generative model to quantify the tone and the topics of texts. Tillmann and Walter (2018) apply dictionary-based sentiment indices to analyse the tone of Bundesbank and ECB speeches. The authors find significant divergences between the tone of the two institutions. An additional topic is about measuring policy uncertainty. Baker et al. (2016) developed the prominent economic policy uncertainty index (EPU) by analysing news coverage with a dictionary method. Using a (nonlinear) text regression method to construct an EPU for Belgium, Tobback et al. (2018) show that they have improved the predictive power of the EPU.

This paper makes several contributions to the literature on forecast evaluation and textual analysis. First, German forecasters' narratives were considered using textual analysis methods. Second, previous studies have almost exclusively focused on dictionary methods to generate sentiment indices. To the best of the author's knowledge, this paper is the first in forecast evaluation to apply (linear) text regression approaches, and additionally, it uses a recursive estimation technique. Third, the paper tests why fore-

casters' narratives have predictive power. Although recent studies discussed several explanatory hypotheses, the answer is still insufficiently explored.

The purpose of the paper is to analyse German forecasters' narratives and the question as to whether the forecasters' stories and expectations contain additional information relative to numerical forecasts. Based on 534 business cycle forecast reports covering 10 German institutions from 1993 to 2017, the paper creates sentiment indices using text mining techniques. Regression results suggest that some sentiment indices can reduce the absolute magnitude of the quantitative forecast errors for GDP growth and inflation forecasts. German forecasters' narratives are informative for the accuracy of German business cycle forecasts. One explanation might be that forecasters' narratives contain useful information about the future stance of the German economy. An in-sample and out-of-sample forecasting exercise tests whether the sentiment indices can predict the evolution of German economic activity. Forecasting results indicate weak in-sample predictive power and out-of-sample predictive power of the sentiment indices.

The following section explains the methodology used to convert qualitative text data into quantitative sentiment scores. Section 3 describes the employed text corpus and numerical data. Section 4 analyses the empirical results, and Sect. 5 concludes and discusses these results.

2 Methodology: sentiment analysis

There are various computational analysis methods to connect word counts to attributes to generate sentiment indices, e.g. dictionary-based methods, text regression methods, generative models, and word embeddings (Gentzkow et al. 2019). This paper uses dictionary-based methods and text regression methods to convert qualitative text data into quantitative indices.

Furthermore, qualitative measures can only be directly related to macro-variables, provided that they are appropriately scaled (Clements and Reade 2020, p. 1491). Hence, all weighted sentiment indices are standardized to have a mean equal to zero and a standard deviation equals to one. In order to avoid bias in the measure, all weighted sentiments are normalized by the total number of words per report to account for varying text lengths and numbers of documents per year (Fritsche and Puckelwald 2018).

2.1 Dictionary-based method

Following Clements and Reade (2020) and Sharpe et al. (2020), the dictionary-based method is applied to develop sentiment indices. In fact, three well-established linguistic dictionaries are used to generate five different indices.

– First, the word list is prepared by Bannier et al. (2018). This is the German equivalent of the English original dictionary provided by Loughran and McDonald (2016). The last-mentioned word list is well established for textual analysis in finance- and accounting-specific contexts. The word list prepared by Bannier et al.

(2018) includes over 2200 positive and 10,000 negative word forms. The dictionary is binary coded for polarity in positive and negative terms.

– Second, there is a forecast-specific German dictionary-based on Sharpe et al. (2020). According to Di Fatta et al. (2015), words have different connotations and meanings in different contexts, and sentiment indices have to be adapted to the content to which they have been applied. To this end, Sharpe et al. (2020) developed a forecast-specific word list which excludes words that have special meanings in an economic forecasting context. The word list contains 205 positive and 103 negative words (see Tables 8, 9) and is binary coded like the previous one.

– Finally, there is the SentimentWortschatz (SentiWS) dictionary (Remus et al. 2010). The SentiWS dictionary contains a German-specific word list for sentiment analysis. The current version (v2.0) contains about 16,000 positive and 18,000 negative word forms, and unlike the other two dictionaries, it includes weights for polarity within the interval of $[-1; 1]$.

Two different score systems will be applied for the two binary dictionary-based sentiments (hereinafter called 'Bannier' and 'Sharpe'). Sentiment score number one consists of the difference between positive word count, P, and negative word count, N, normalized by the total number of words, T, per report:

$$\text{Sentiment score}_1 = (P - N)/T \tag{1}$$

The second sentiment score (polarity score) is defined as the quotient of the difference between positive and negative word counts and the sum of positive and negative words:

$$\text{Sentiment score}_2 = (P - N)/(P + N) \tag{2}$$

In contrast, the SentiWS index is a continuous score. The score of each word sums up over all words and is normalized by the total number of words per report.

2.2 Automatic variable selection approach

The automatic variable selection approach is a promising text regression method to generate regression-based sentiment indices (e.g. Pröllochs et al. 2018). In contrast to the dictionary-based method, here the required dictionary is not given and will be recursively estimated. In fact, the estimated parameters will be updated by expanding the estimation windows by one observation in chronological order (see Sect. 2.3). Generally, text regression methods introduce a regularization penalty that reduces the complexity, number, and size of the predictors included in the model. Penalized linear models use each word in the text corpus as explanatory variables, shrink non-informative noise variables to zero, and select decisive variables (Pröllochs et al. 2015).

Regularization methods can serve as mathematical mechanisms to extract important terms, which is why it is a common tool for variable selection in data science (Pröllochs et al. 2018; Varian 2014). Given a standard multivariate regression with y (dependent

variable) as a linear function of β_0 (constant) and x_j (explanatory variable), the penalty term of the form:

$$\lambda \sum_{j=1}^{P} \left[(1-\alpha)|\beta_j| + \alpha|\beta_j^2| \right] \tag{3}$$

can be added (Varian 2014). Setting $\alpha = 0$, the term Eq. 3 reduces to the linear l_1-norm penalty $\lambda \sum_{j=1}^{P} |\beta_j|$, which represents the *least absolute shrinkage and selection operator* (LASSO) introduced by Tibshirani (1996). Formally, the LASSO estimator is given by (Pröllochs et al. 2015):

$$\hat{\beta}_{\text{LASSO}} = \arg\min_{\beta} \sum_{i=1}^{N} \left[y_i - \beta_0 + \sum_{j=1}^{P} \beta_j x_{ij} \right]^2 + \lambda \sum_{j=1}^{P} |\beta_j| \tag{4}$$

where x_{ij} are the document terms (words $j = 1, \ldots, P$) for forecast report $i = 1, \ldots, N$, and y_i represents the 12-month-ahead fixed horizon growth and inflation forecasts as response variables. If $\lambda = 0$, the penalty reaches zero, and we get the classical OLS estimator by simply minimizing the residual sum of squares. The higher λ, the larger the penalty shrinkage gets, with the result that more coefficients end up being zero. The optimal λ^* is estimated by minimizing the mean squared error (MSE) (Dimpfl and Kleiman 2019):

$$\text{MSE}_{\text{CV}}(\lambda) = \frac{1}{K} \sum_{i=1}^{K} \frac{1}{n_i} ||y_i - X_i \hat{\beta}_{\text{LASSO}}^{-i}||_2^2 \tag{5}$$

using an established 10-fold cross-validation, where n_i is the size of ith subsample. Therefore, the data are split into K subsets, one part i is removed, the coefficients $\hat{\beta}_{\text{LASSO}}^{-i}$ are estimated, and the cross-validated $\text{MSE}_{\text{CV}}(\lambda)$ is calculated for any given value of λ.

In contrast, setting $\alpha = 1$ shortens the term Eq. 3 to the quadratic l_2-norm penalty $\lambda \sum_{j=1}^{P} \beta_j^2$, and the ridge estimator is implemented (Pröllochs et al. 2015):

$$\hat{\beta}_{\text{Ridge}} = \arg\min_{\beta} \sum_{i=1}^{N} \left[y_i - \beta_0 + \sum_{j=1}^{P} \beta_j x_{ij} \right]^2 + \lambda \sum_{j=1}^{P} \beta_j^2 \tag{6}$$

Again, the tuning parameter λ is the regularization penalty. The quadratic penalty l_2-norm follows similar characteristics to the LASSO penalty: if λ reaches zero, we get OLS coefficients; if λ moves towards infinity, the coefficients come down to zero. However, in contrast to the LASSO regularization, the ridge estimator does not set

explicitly some coefficients equal to zero (Pröllochs et al. 2015).[2] Again, the optimal λ^* is estimated by minimizing the MSE using 10-fold cross-validation.

Equations 4 and 6 are used to estimate the LASSO and ridge regression coefficients $\hat{\beta}_{LASSO}$ and $\hat{\beta}_{Ridge}$. The magnitude of $\hat{\beta}_{LASSO}$ and $\hat{\beta}_{Ridge}$ serve as the weight and a measure of variable importance, specifying which variables (words) are included in the final dictionary (Pröllochs et al. 2015). A linear rule is then applied to calculate document ith sentiment score. Again, the document's score is defined as the continuous score normalized by the total number of words.

2.3 Recursive estimation

In order to guarantee that no information is produced and used for tests for forecast efficiency and predictive power that are (hypothetically) not known for forecaster in time t, a recursive estimation technique will be applied for sentiment indices based on the automated variable selection approach. First, a sufficiently large text corpus is generated as a basis (pre-estimation corpus) using business cycle forecast reports from the period 1993–1998, including 74 observations. Second, based on the pre-estimation corpus, a recursive estimation approach is applied, expanding the estimation windows by one observation per estimation in chronological order. In fact, the following procedure is executed in each recursive estimation step: First, the extended text corpus is established and weighted; second, the optimal λ^* is estimated by minimizing the MSE using 10-fold cross-validation; third, LASSO and ridge estimator (Eqs. 4, 6) are used to estimate the respective dictionaries and weights ($\hat{\beta}_{LASSO}$ and $\hat{\beta}_{Ridge}$); finally, the respective sentiment (document) score is calculated and stored in a common series.

3 Corpus and data

3.1 The text corpus

The plain corpus includes business cycle forecast reports for Germany issued by 10 institutions with different institutional backgrounds. First, the corpus covers the six largest economic research institutes in Germany that are formally politically and economically independent. These comprise the five publicly founded institutes, the Ifo Institute Munich (Ifo), the Berlin Institute (DIW), the Essen Institute (RWI), the Halle Institute (IWH), the Kiel Institute (IfW), and the privately funded Hamburg Institute (HWWI).[3] Second, the corpus contains institutes that are funded by interest groups: the employer's institute of the German economy located in Cologne (IW Köln), and the trade union's macroeconomic policy institute (IMK). Third, the corpus includes the 'joint diagnosis' (GD), the economic projection of the leading research institutes as an institution within the process of economic policy advice. Fourth, the corpus covers

---

[2] Ridge regularization is introduced as an opposite of LASSO because the ridge estimator cannot benefit from a parsimonious model (Pröllochs et al. 2018). Therefore, the elastic net, a mixture of both regularization methods, is not absolutely necessary for this investigation.

[3] Until 2005, the HWWI was known as HWWA and mainly funded by public money. It became a privately funded institute in 2006.

a financial institution, the Bundesbank. The German central bank is another formally politically and economically independent public institution.

The entire corpus contains 534 documents.[4] There is a wider range of potential business cycle forecast reports for Germany than the selected institutes that did not meet the defined criteria. For the selection, a range of criteria was checked:

– *Business cycle forecast (sub-)section* Business cycle forecast reports are heterogeneous in size and content. Some reports are structured into different subsections like recent national or international economic development, business cycle forecasts, economic policy advices, or methodological explanations. Other reports are miscellaneous texts of various themes and cannot be split in a meaningful way. Therefore, business cycle reports should contain a clearly defined forecast (sub-)section.
– *Time range* The corpus covers business cycle forecast reports for Germany from 1993 to 2017 to circumvent the German reunification and possible misspecification for East and West Germany.
– *Forecasters' experiences* Continuity and regularity of publication within the examined period ensure forecasters' experiences in the field of economic forecasting, ensuring a sufficient level of homogeneity in language across institutes.
– *Language homogeneity* The (relatively short) period of 25 years as well as forecasters' experiences assures a sufficient degree of homogeneity in language over time.
– *Quantitative forecast availability* To use a comparative sample for growth and inflation forecast analysis, only business cycle forecast reports with a calculable fixed horizon forecast for growth and inflation will be used. The availability of numerical point forecasts of growth and inflation for the current and next year restricts the number of incorporated forecast reports (see Sect. 3.2).
– *Forecasting date* The forecasting date is distributed over the whole year, depending on respective institutional practice and the frequency of publication. In most cases, the frequency of publication is bi-annual or higher.
– *Text availability* Another criterion was the public availability of business cycle forecast reports, which is why private institutes like banks are not included.

As a result, 534 business cycle forecast reports for Germany issued by 10 institutions are used for the creation of the corpus. In the first step of textual analysis, data cleaning and linguistic pre-processing are applied to all texts. In fact, line breaks, numbers and words with fewer than four characters are eliminated, lower cases were introduced, stopwords (e.g. from German linguistic stopword lists or names) and sparse terms where a word that occurs in less than 10% of documents are removed. With reference to Zipf's law (Zipf 1949), the texts are weighted with their term frequency—inverse document frequency (tf-idf).[5] Zipf's law for empirical language implies that a word's frequency is inversely proportional to its rank. Consequently, the corpus is adjusted for

---

[4] See Table 7 for an overview.

[5] The principle behind the tf-idf weighting scheme is that the more often a word appears in a document, the more important it is (term frequency). But, the more the word appears in all documents, the less important it is (inverse document frequency). The tf-idf weighting scheme is a commonly used metric in text analysis literature (see e.g. Loughran and McDonald 2011; Sharpe et al. 2020).

Fig. 1 Wordcloud of German business cycle forecast reports. *Notes* Own illustration

that symptom. Figure 1 shows the wordcloud of the weighted corpus. The wordcloud sort terms frequency in descending order. The larger the word, the more often the term occurs. The wordcloud shows that the weighted corpus includes a lot of important forecast-specific vocabulary, for example 'Anstieg' (growth), 'Prognose' (forecast), and 'Exporte' (exports).[6]

Finally, Porter's stemming algorithm (Porter et al. 1980) is used to truncate the different word forms to its base forms.[7]

3.2 The sample

The incorporated business cycle forecast reports for Germany typically contain numerical fixed event forecasts of growth and inflation for the current and next year. Depending on the forecast date, the forecast horizon of fixed event forecasts varies from one up to 11 months. Heilemann and Müller (2018) show in a forecast evaluation study for Germany that forecast accuracy decreases with increasing forecast horizon, and that differences in forecast accuracy are mainly determined by the different timings of the production of the forecasts.[8]

Furthermore, uncertainty and cross-sectional dispersion of fixed event forecasts show a pronounced seasonal pattern (Dovern et al. 2012). Consequently, fixed horizon forecasts are used to reduce different forecast horizons within one quarter. The method of Dovern and Fritsche (2008), Heppke-Falk and Hüfner (2004) and Smant (2002)

---

[6] Nevertheless, the pre-processed corpus contains some meaningless terms as 'gegenüber' (in relation to) or 'deutlich' (obvious). To avoid a selection bias, the linguistic stopword lists were not manually expanded.

[7] German is a morphologically rich language and the text corpora is a specific economic text corpora, and therefore, the meaning of a word is crucial. Stemming reduces different word forms to its base forms and to retain the meaning and semantic interpretation of the word (Jivani 2011). Porter's stemming algorithm is one of the best stemming algorithms; it has a lower error rate and it is a light stemmer (Jivani 2011). Thus, the stemming procedure reduces complexity without losing the meaning of the word form. In contrast, lemmatization reduces the word forms to its root forms and the semantic interpretation can be lost (Jivani 2011).

[8] An analysis of forecast revision patterns shows an inverse L-curve relationship between accuracy and shortening forecast horizon (Heilemann and Müller 2018).
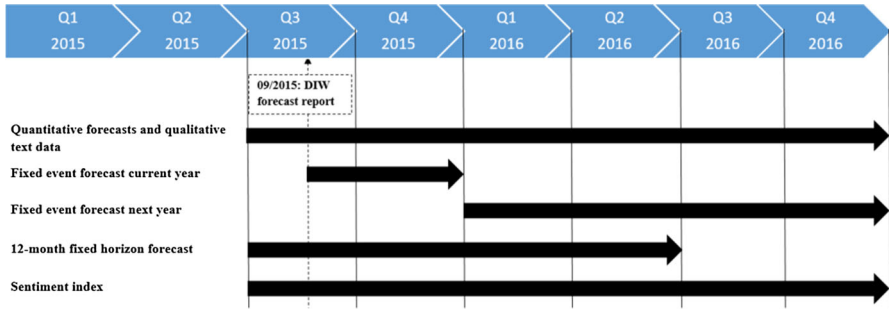
Fig. 2 Time line of forecast horizons and construction for DIW forecast report. *Notes* Own illustration

$$\hat{y}_{i,t}^{12} = \frac{4 - q + 1}{4} \tilde{y}_{i,t}^{0} + \frac{q - 1}{4} \tilde{y}_{i,t}^{1} \tag{7}$$

is applied to construct 12-month-ahead fixed horizon forecasts for growth and inflation. Given current $(\tilde{y}_{i,t}^{0})$ and next $(\tilde{y}_{i,t}^{1})$ year fixed event forecast, q is equal the quarter where the forecast is done. Subsequent, the fixed horizon forecast is approximated as a quarterly weighted average of their share in both years. For example, considering the forecasts of the Berlin institute from September 2015, $\tilde{y}_{i,t}^{0} = 1.8$ and $\tilde{y}_{i,t}^{1} = 1.9$, q is equal to three and the 12-month-ahead fixed horizon forecast $\hat{y}_{i,t}^{12} = 1.85$.

Moreover, forecast narratives cannot distinguish between different forecast horizons within a quantitative textual analysis. All in all, nine different sentiment indices will be calculated for each forecasting report at time t.

Figure 2 depicts the different forecast horizons and the construction of 12-month-ahead fixed horizon forecast and sentiment indices using an forecast report of the German institute of economic research (DIW Berlin).

Besides, seasonally adjusted and finally revised real GDP is used for realized GDP growth (quarterly data, source Federal Statistical Office 2019b). Finally, the revised consumer price index is used for actual inflation outcome (monthly data, source Federal Statistical Office 2019a).[9] (Dovern et al. 2012) point out that the approximation error in the fixed horizon series in Eq. 7 could result in a correlation if dependent variable and regressors are constructed in the same way. To avoid this, the annualized cumulative percentage change from past quarter $t - h$ to current quarter t is used for the realized values. Thus, $h = 4$ denotes the forecasting horizon in quarters based on the 12-month-ahead fixed horizon forecasts.

The forecast error is defined as $e_t = A_t - P_t$—the realized value in period t minus the forecast made in period $t - j$. Hence, a positive forecast error represents an underestimation of the growth (inflation) rate, and vice versa, whereas a negative forecast error corresponds to an overestimation.

Table 1 provides an overview of some standard error measures of forecast evaluation (see for example, Fildes and Stekler 2002) for the pooled data of the introduced sample. On the whole, the error measures correspond to previous forecast evaluation studies

---

[9] In forecast evaluation contexts, it is appropriate to use first published (real-time) data or the last available revised data (Döpke et al. 2019). Here, the revised data are used because of data availability.

Table 1 Descriptive statistics on forecast accuracy in Germany, 1993–2017 . *Source*: Authors' own calculations

	Growth forecasts	Inflation forecasts
Number of observations	534	534
Mean error	− 0.051	− 0.135
Mean absolute error	1.715	0.685
Root mean squared error	2.578	0.862
Theil's inequality coefficient	1.000	0.546
Number of overestimations	274	292
Number of underestimations	260	242
Information content	1.398	1.217
χ^2-test	0.000	0.000
AUROC	0.746	0.763

See text for details. Values are rounded to three decimal places. For the χ^2-test p values are reported. The Mean Error: $ME = \frac{1}{T}\sum_{t=1}^{T} e_t$, where e_t is the forecast error in each period, defined as actual A_t (in t) minus predicted P_t (in $t-1$ for period t). $t = 1, \ldots, T$ is the time index. The Mean Absolute Error: $MAE = \frac{1}{T}\sum_{t=1}^{T} |e_t|$. The Root Mean Squared Error: $RMSE = \sqrt{\frac{1}{T}\sum_{t=1}^{T} e_t^2}$. The Theil's Inequality Coefficient: $\text{Theil U} = \dfrac{\sqrt{\frac{1}{T}\sum_{t=1}^{T} |e_t^2|}}{\sqrt{\frac{1}{T}\sum_{t=1}^{T} |A_t^2|}}$

for Germany (Heilemann and Stekler 2013; Döpke et al. 2019). The ME is nearly zero, indicating unbiased forecasts. Growth forecasts MAE and RMSE are on average large compared to Heilemann and Stekler (2013) and Döpke et al. (2019) due to the forecasting error in the *Great Recession* 2008/2009.[10]

Considering the ability to forecast turning points, three directional analysis measures are included. Referring to Diebold and Lopez (1996, p. 28) and Merton (1981), the information content of a forecast series is calculated.[11] The forecasts beat a pure coin-flip if the informational content has a value above one. Second, a χ^2-test validate whether the forecasts are significant better than chance, testing the null hypothesis of no information content of the forecasts under investigation. The results indicate that both, growth and inflation forecasts, have an significant informational content at conventional significance levels. In addition, the area under a receiver operating curve (AUROC), a frequently used measure of the quality of directional forecasts (see, e.g. Berge and Jordà 2011; Pierdzioch and Rülke 2015; Liu and Moench 2016) is calculated. An AUROC < 0.5 indicate that forecasts are even worse than pure coin-flip and an AUROC $= 0.5$ that forecasts are indistinguishable from a pure coin-flip because the ROC curve coincides with the 45° line. An AUROC > 0.5 and < 1 beat the coin-flip, whereas an AUROC $= 1$ represents perfect forecasts. Considering the AUROC for

---

[10] Calculations without the period of the *Great Recession* in 2008/2009 results in similar error measures.

[11] The information content addressing the relation between the number of forecasts with correctly predicted direction by the number of all forecasts.

growth and inflation forecasts, both forecasts beat again pure coin-flip and indicate to some directional accuracy.

4 Empirical results

4.1 Sentiments' characteristics

Table 2 gives an overview of sentiment characteristics.

Considering dictionary metrics as positive and negative entries and standard statistical measures, Table 2 shows how different the individual sentiment approaches work. The ridge estimation results show that the ridge estimator does not explicitly set some coefficients equal to zero. In contrast to the LASSO estimator, the ridge approach selects much more words as its LASSO counterpart.

Tables 10, 11 and 12 list in a full sample example the (stemmed) dictionaries and weights generated by the automated variable selection approach. Table 10 shows the estimated 71 words and their coefficients according to LASSO regression with real GDP growth forecasts as the response variable (hereinafter 'LASSO_GDP_P'). The term with the most positive weight is 'upswing' ('Aufschwung'), which in German is also a synonym for 'boom' or 'recovery', whereas 'drastic' ('drastisch') is the word with the most negative coefficient. The list of plausible words and weight with respect to GDP development is long, i.e. 'export dynamic' ('Exportdynamik'), 'continuation' ('Fortsetzung'), 'lively' ('schwungvoll') with positive coefficients, or 'deep' ('tief'), 'layoffs' ('Entlassungen'), and 'shrink' ('schrumpfen') with negative coefficients. Nevertheless, the list contains few outliers whose economic sense is not immediately clear, e.g. 'a third' ('drittel'), or where the words have a non-intuitive weight, such as 'recover' ('erholen').[12]

Similar patterns can be observed in other text regression-based dictionaries. Table 11 lists the estimated 69 words and weights according to LASSO regressions, with inflation forecasts as the response variable (hereinafter 'LASSO_INF_P'). Table 12 list ridge regression results for real GDP growth forecasts (hereinafter 'Ridge_GDP_P') and inflation forecasts (hereinafter 'Ridge_INF_P'). Both tables list the top 30 estimated words with the largest positive and negative coefficients.

Figures 3 and 4 give a visual impression of the generated sentiment indices. The figures illustrate the sentiment values per business cycle forecast report aggregated over years and across institutes, in combination with the realized real GDP growth, or inflation rate, respectively. Panels (a) to (i) present for each sentiment specification the aggregate sentiment value per year on the left axis (solid line), and the realized value of GDP growth, respective inflation, on the right axis (dashed line).

Considering each of the panels from (a) to (i) separately, we can conclude that each sentiment specification varies in its pattern. Concerning, for instance, the *Great Recession* in 2008–09, it can be seen that some sentiment indices are closer to the real development, i.e. LASSO_GDP forecast in Fig. 3, whereas some sentiment indices

---

[12] An extended pursuit of stopwords could reduce some 'outliers' to a minimum. But first, the objective of this paper is not to find the best stopword list, and, second, the few outliers should not matter from a purely statistical point of view.

Table 2 Overview dictionaries metrics

Dictionary	Bannier (1,2)	Sharpe (1,2)	SentiWS	LASSO (GDP)	LASSO (inflation)	Ridge (GDP)	Ridge (inflation)
Dictionary type	Binary	Binary	Weighted	Weighted	Weighted	Weighted	Weighted
Total entries	7619	292	22972	71	69	2359	2359
Positive entries in %	1363	196	10863	42	38	1257	1161
	17.9	67.1	47.3	59.2	55.1	53.3	49.2
Negative entries in %	6256	96	12109	29	31	1102	1198
	82.1	32.9	52.7	40.8	44.9	46.7	50.8
Average score	–	–	− 0.0515	− 0.0032	0.0002	0.0000	0.0000
Stand. deviation	–	–	0.2153	0.0302	0.0159	0.0021	0.0017

Own representation. Full sample example

(a) Bannier1

(b) Bannier2

(c) Sharpe1

(d) Sharpe2

(e) SentiWS

(f) LASSO (GDP forecast)

(g) LASSO (inflation forecast)

(h) Ridge (GDP forecast)

(i) Ridge (inflation forecast)
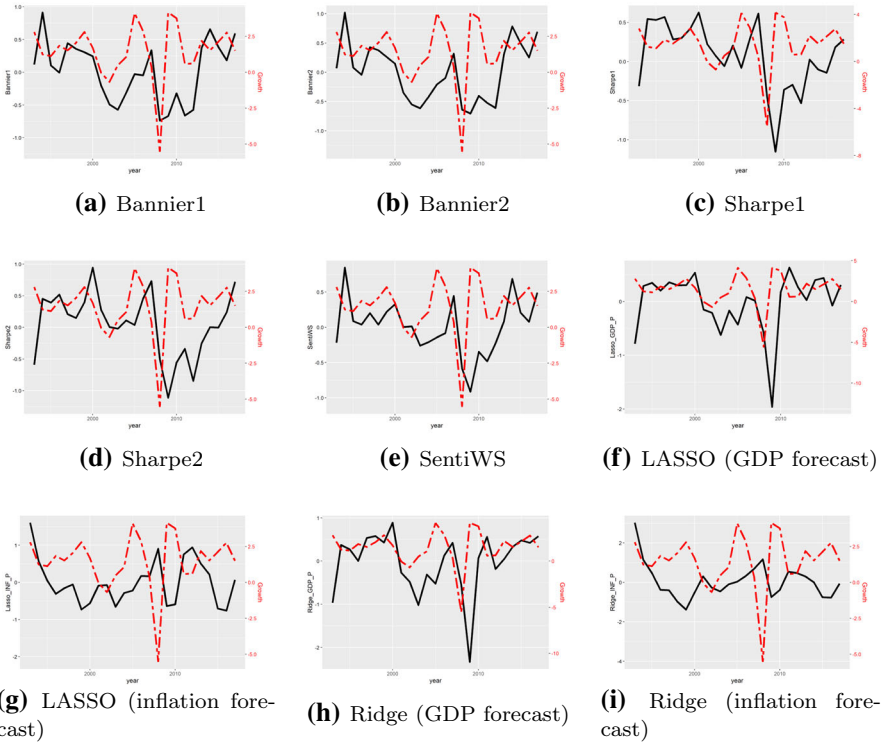
Fig. 3 Sentiment indices (solid line) and realized growth (dashed line), aggregate over years and institutes

have a longer time lag, i.e. Sharpe 1 in Fig. 3. Other sentiment indices are even ahead of the real development, i.e. Sharpe 2 in Fig. 4. Another picture illustrates a (partly) countercyclical behaviour. For example, Bannier1 and Bannier2 in Fig. 4 show this countercyclical behaviour, which could be explained by a huge time lag or an opposite polarity of terms.

In summary, the generated sentiment indices differ across patterns and in amplitude, as well as in terms of time lag and lead.

4.2 Forecast efficiency

Forecast efficiency analysis is used to test whether the narratives of German business cycle reports contain useful information for the numerical forecasts of German forecasters. More precisely, we test whether the sentiment indices can be used to improve the accuracy of the quantitative point forecasts. In particular, we test for weak and strong efficiency of forecasts by using the specification of Holden and Peel (1990):

$$e_{i,t} = \beta_{0,i} + \beta_1 e_{i,t-1} + \beta_2 \text{Sentiment}_{i,t} + u_{i,t}, \tag{8}$$

and test the joint null hypothesis $H_0 : \beta_{0,i} = \beta_1 = \beta_2 = 0$.

(a) Bannier1

(b) Bannier2

(c) Sharpe1

(d) Sharpe2

(e) SentiWS

(f) LASSO (GDP forecast)

(g) LASSO (inflation forecast)
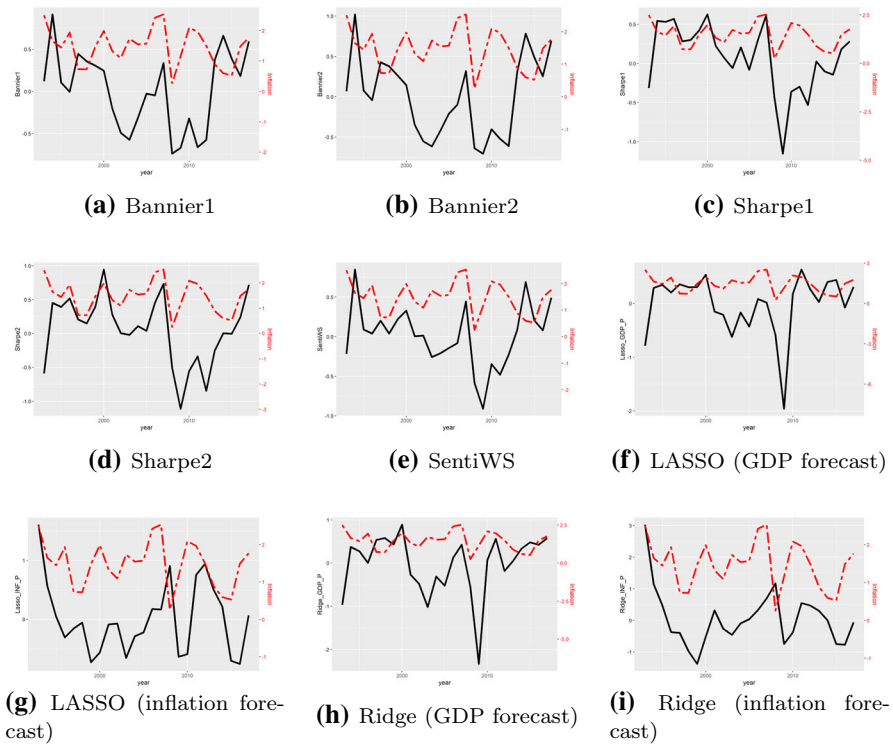
(h) Ridge (GDP forecast)

(i) Ridge (inflation forecast)

Fig. 4 Sentiment indices (solid line) and realized growth (dashed line), aggregate over years and institutes

In Eq. 8, $e_{i,t}$ is the forecast error of forecaster i in time t, $\beta_{0,i}$ is institution's i individual effect, $e_{i,t-1}$ is the institution's forecast error made in $t-1$, Sentiment$_{i,t}$ is the forecaster's sentiment index at time t as exogenous variable which is known by the forecasters on the forecasting date, and $u_{i,t}$ is the error term. Forecasts are weakly efficient if the forecast errors are not autocorrelated, and forecasts are strongly efficient if there is no variable that helps to predict the forecast errors, including the lagged forecast error. Optimal forecasts should consider all available information at the date of the forecast. A fixed effects estimation approach is used to account for individual institutional effects, such as different forecast horizons during the quarter. According to Gaibulloev et al. (2014), panel-corrected standard errors (PCSE) suggested by Beck and Katz (1995) are reliable for panel type T>N to deal with unit heterogeneity and panel heteroscedasticity. The standard test statistics are reliable and the Nickell bias (Nickell 1981) is negligible (see Gaibulloev et al. 2014, and the literature cited therein). [13] Estimates are corrected for serial and cross-sectional correlation. Comparable forecast evaluation studies have used this kind of robust standard errors (see, among others, Keane and Runkle 1990; Kauder et al. 2017; Döpke et al. 2019).

---

[13] Therefore, it is not necessary to employ the dynamic panel estimator proposed by Arellano and Bond (1991)

Table 3 Tests for efficiency of forecasts—1999-2017

	Dependent variable: growth forecast error[a]									
Constant	−0.203*** (0.057)	0.079 (0.132)	0.078 (0.132)	0.052 (0.131)	0.052 (0.130)	0.077 (0.132)	0.086 (0.128)	0.056 (0.127)	0.083 (0.125)	−0.057 (0.124)
IGDP_FE	−[b]	−0.212*** (0.058)	−0.206*** (0.058)	−0.182*** (0.057)	−0.167*** (0.057)	−0.196*** (0.057)	−0.099* (0.057)	−0.221*** (0.054)	0.002 (0.058)	−0.188*** (0.052)
Bannier1		0.118 (0.135)								
Bannier2			0.032 (0.126)							
Sharpe1				−0.324** (0.151)						
Sharpe2					−0.402*** (0.136)					
SentiWS						−0.152 (0.155)				
Lasso_GDP_P							−0.736*** (0.145)			
Lasso_INF_P								−0.761*** (0.124)		
Ridge_GDP_P									−1.093*** (0.166)	
Ridge_INF_P										−1.341*** (0.159)

Table 3 continued

Dependent variable: growth forecast error[a] (continued summary statistics)

Observations	387	387	387	387	387	387	387	387
R^2	0.043	0.045	0.057	0.063	0.097	0.122	0.142	0.198
Efficiency test [p value]	[< 0.001]	[0.002]	[< 0.001]	[< 0.001]	[< 0.001]	[< 0.001]	[< 0.001]	[< 0.001]

Dependent variable: inflation forecast error[a]

Constant	− 0.062	− 0.062	− 0.058	− 0.058	− 0.063	− 0.063	− 0.067	− 0.065	− 0.106
	0.042	0.042	0.042	0.041	0.042	0.042	0.039	0.042	0.037
$-$[b]	–								
IINF_FE	− 0.108**	− 0.108**	− 0.121**	− 0.132***	− 0.109**	− 0.109**	− 0.045	− 0.128**	− 0.067
	(0.050)	(0.050)	(0.050)	(0.051)	(0.050)	(0.052)	(0.047)	(0.054)	(0.047)
Bannier1	0.023								
	(0.045)								
Bannier2		0.019							
		(0.040)							
Sharpe1			0.073						
			(0.049)						
Sharpe2				0.113***					
				(0.043)					
SentiWS					− 0.011				
					(0.047)				

Table 3 continued

	Dependent variable: inflation forecast error[a]								
Lasso_GDP_P						−0.0005 (0.049)			
Lasso_INF_P							−0.323*** (0.043)		
Ridge_GDP_P								0.046 (0.049)	
Ridge_INF_P									−0.568*** (0.054)
Observations	387	387	387	387	387	387	387	387	387
R^2	0.013	0.013	0.020	0.030	0.013	0.013	0.157	0.015	0.269
Efficiency test [p value]	[0.028]	[0.085]	[0.033]	[0.004]	[0.088]	[0.091]	[<0.001]	[0.062]	[<0.001]

Standard errors are in parentheses; p values are in brackets

[a] Cross-section SUR (PCSE) standard errors and covariances (d.f. corrected) following the method of Beck and Katz (1995)

[b] The function in *R* to calculate the overall intercept does not work with one-dimensional objects, it requires at least two explanatory variables. ***, **, and * denote rejection of the null hypothesis at the 1, 5, and 10 % significance level, respectively

Table 3 presents the estimated parameters and the standard errors (in parentheses) of the individual coefficients and the p-value [in brackets] for the joint efficiency test. In almost all cases, the weak efficiency condition of no serial correlation of the forecast errors has to be rejected for GDP growth forecasts. Moreover, test results with sentiment indices indicate several significant influences of forecasters' narratives for forecast accuracy. For both Sharpe sentiment indices, as well as for all text regression-based sentiment indices, the null of no correlation has to be rejected at a conventional significance level. The negative coefficients indicate that a higher sentiment value correlates with a higher GDP prediction in that smaller (or negative) forecast errors imply higher forecast values. In addition, all specifications reject the joint test on efficiency. But it is not clear whether the autocorrelated forecast error or the sentiment indices are the reason for the rejection of the joint tests.

Considering inflation forecasts, again, the lagged forecast error has generally a significant influence on the forecast error of the following period, at a conventional significance level. Moreover, we find some hints for explanatory power of the narratives on the numerical point forecast errors. Sharpe2 and the LASSO, as well as the ridge sentiment with inflation forecast as response variable, are significantly correlated with the forecast error. Both text regression-based sentiment indices are the only two out of nine specifications that also reject the joint efficiency hypothesis without having autocorrelated errors. The varying signs of sentiment indices' coefficients indicate sentiment indices with different polarity. Thus, rising inflation, e.g. the word 'inflation', could have both positive and negative weights, depending on the given dictionary (dictionary-based methods) and the used response variable (text regression methods).

The efficiency test results suggest that forecasters' narratives have informational power for the forecast errors at the time when the forecasts were made, implying that the numerical forecasts do not make efficient use of all available information. Previous studies (e.g. Döpke et al. 2010, 2019) confirm that forecasts for Germany are not strongly (in part weakly) efficient by not incorporating all available information. But they never test the narratives of the forecaster itself. Sentiment indices, based on business cycle forecast reports, seem informative for the accuracy of German business cycle forecasts.[14] Thus, forecasters' narratives contain information which is not exhausted by numerical forecasts. One explanation might be that the forecasters' narratives contain useful information about the future stance of the German economy.

4.3 Predictive power

To test whether the narratives of German business cycle forecast reports contain useful information for the future stance of the German economy, the paper applies an in-sample and an out-of-sample forecast exercise.

---

[14] Robustness checks with the last known forecast error instead of the lagged forecast error support this finding. The results are available on request.

4.3.1 In-sample forecasting regressions

Following Estrella and Hardouvelis (1991), Stock and Watson (2003) and Ferreira (2018), single forecasting equations are used to predict actual GDP growth and the inflation rate of changes. The in-sample and (pseudo) out-of-sample forecasting exercise tests whether text-based sentiment indices have predictive power for actual GDP growth and inflation. Similar methods were used to find predictors of economic activity (Estrella and Hardouvelis 1991) or predictors of business cycle fluctuations (Ferreira 2018). In order to do that, the sentiment indices are transformed by averaging all observations per quarter to build quarterly time series as explanatory variables. Hence, we get a quarterly time series with 100 observations from 1993Q1 to 2017Q4. The dependent variable in the basic forecasting regression is the annualized cumulative percentage change in real GDP respectively inflation. Following (Estrella and Hardouvelis 1991; Stock and Watson 2003):

$$\hat{Y}_{t|t+h} = (400/h)[\ln(Y_{t+h}/Y_t)] \tag{9}$$

where Y_t and Y_{t+h} denote the level of real GDP (consumer price index) in period t and $t + h$, $\hat{Y}_{t|t+h}$ is the annualized cumulative percentage change from current quarter t to future quarter $t + h$, and $h = 4$ denotes the forecasting horizon in quarters. The single forecasting equation is provided by (Ferreira 2018):

$$\hat{Y}_{t|t+h} = \alpha + \underbrace{\sum_{i=1}^{p} \rho_i \hat{Y}_{t-i}}_{\text{Lag. endog. var.}} + \underbrace{\sum_{j=0}^{q} \beta_j \text{SI}_{t-j}}_{\text{Sentiment indices}} + \underbrace{\sum_{m=1}^{3} \sum_{j=0}^{q} \gamma_j^m \text{IN(m)}_{t-j}}_{\text{Control variables}} + \epsilon_{t+h} \tag{10}$$

where SI_t denotes the respective sentiment index, and IN(m) represents German leading indicators as control variables. The control variables are also standardized by subtracting the mean from each variable and dividing it by its standard deviation. The forecast horizon h is set to four quarters to capture the annualized cumulative percentage change of GDP growth ($\hat{Y}_{t|t+h}$), respectively inflation, from current quarter t to future quarter $t + h$. To hold the model parsimonious, the lag length p of the endogenous variable is set to one, and q is set equal to 0.

The single forecast regression given in Eq. 10 reduces under the simplifying assumption to a simple forecast equation, as suggested by Estrella and Hardouvelis (1991). According to Estrella and Hardouvelis (1991), the overlapping forecasting horizons provoke a moving average error term of order $h - 1$, resulting in consistent but inefficient estimates. Therefore, Newey and West (1987)-corrected standard errors for heteroscedasticity and autocorrelation are applied with a lag length set equal to three ($h = 4$) in line with Estrella and Hardouvelis (1991).[15]

---

[15] An automatic selection method for the number of lags is given by Andrews (1991) approximation rule. Another widely used method is to determine the lag length simply to the integer part of $T^{\frac{1}{4}}$, where T is the sample size (Greene 2012).

As control variables for the forecasting regressions, several admitted economic predictors for the German business cycle are introduced:[16]

- First, the term 'spread' (long-term interest rate minus the short-term interest rate) serves as a monetary control variable. The long-term interest rate serves the yield on debt securities outstanding issued by residents with mean residual maturity of more than nine and up to 10 years (monthly average, source Deutsche Bundesbank 2020). As the short-term interest rate, the EURIBOR 3-month funds money market rate is used (monthly average, source Deutsche Bundesbank 2020).
- Second, total orders received by the German industry serves as the industry control variable. We take the change over the previous month at constant prices, calendar and seasonally adjusted orders (source: Deutsche Bundesbank 2020)
- Third, the Ifo business climate index as leading business cycle indicator (monthly data, source Ifo institute 2020)

Table 4 presents the in-sample forecasting regression results, including selected business cycle indicators as control variables given by Eq. 10. While neither the lagged endogenous variable nor the Ifo business climate index is significantly different from zero, the order inflow and the spread interest rate have a significant impact on the average GDP growth rate. All control variables have the expected sign and a notable magnitude, indicating to a robust specification. Considering the generated sentiment indices, it can be seen that the coefficients are statistically significant only in three out of nine cases. The bag-of-words approach of Bannier1 and both text regression-based sentiments with inflation prediction as response variable (LASSO_INF_P, Ridge_INF_P) are statistically different from zero at conventional significance levels.

Noteworthy is the performance of text regression-based sentiment indices with inflation forecasts as response variables, instead of GDP growth prediction. It seems that this 'wrong' macroeconomic target variable captures the real GDP development as well.[17] This results can be a hint that GDP sub-aggregates, such as investments and consumption, could be promising response variables for text analysis tools to predict GDP growth.

Table 5 presents results regarding inflation in-sample forecasting regressions. Both dictionary-based Bannier sentiment indices have a significant influence on the average growth rate of inflation over the next four quarters. Both sentiment indices are negatively correlated with the target variable.[18] However, most of the generated sentiment indices do not show a significant impact on the average growth rate of inflation over the next four quarters at a conventional significance level.

In brief, changes in the narratives have weak in-sample predictive power on the average growth rate of GDP and inflation over the next four quarters.

---

[16] For a detailed discussion about German business cycle leading indicators, see Heinisch and Scheufele (2018) and the literature cited therein

[17] The reason for the correlations are the generated dictionaries. For example, consider the full sample dictionary and weights for LASSO_INF_P in Table 11 again. Words such as 'recovery' ('erholung'), 'stable' ('stabil'), and 'expansive' ('expansiv') have negative weights, whereas words such as 'slow down' ('abkühlung') and 'deficit' ('verlust') have positive weights. All these words are related to GDP growth but have a reversed sign in relation to GDP growth, which explains the correlation and the negative coefficient.

[18] The negative polarity of inflation is not surprising, given the finance-specific context of the dictionary. There is no 'right' sign of coefficient; it depends only on the given polarity (or weight).

Table 4 Forecasting equations including sentiment indices and control variables for Germany, GDP, 1999Q1 to 2017Q4

	Dependent variable: average growth rate of GDP over the next four quarters									
Lagged endog. var.	0.098	0.092	0.100	0.149	0.113	0.092	0.101	0.120	0.054	0.040
	(0.211)	(0.206)	(0.207)	(0.199)	(0.201)	(0.193)	(0.192)	(0.196)	(0.205)	(0.178)
Order inflow	0.838***	0.733***	0.757***	0.904***	0.860***	0.822***	0.837***	0.745***	0.831***	0.646***
	(0.172)	(0.158)	(0.162)	(0.173)	(0.172)	(0.170)	(0.183)	(0.163)	(0.167)	(0.167)
Interest rate spread	0.962**	1.037**	1.044**	0.913*	0.937*	0.977*	0.962**	0.785*	0.986**	0.629*
	(0.463)	(0.467)	(0.477)	(0.476)	(0.485)	(0.526)	(0.454)	(0.401)	(0.495)	(0.356)
Ifo business climate	0.077	−0.075	−0.070	0.105	0.107	0.068	0.079	0.023	0.044	0.271
	(0.435)	(0.472)	(0.478)	(0.434)	(0.458)	(0.473)	(0.473)	(0.397)	(0.479)	(0.321)
Bannier1		0.628*								
		(0.356)								
Bannier2			0.515							
			(0.342)							
Sharpe1				−0.498						
				(0.363)						
Sharpe2					−0.185					
					(0.303)					
SentiWS						0.127				
						(0.766)				
Lasso_GDP_P							−0.015			
							(0.556)			
Lasso_INF_P								−1.018***		
								(0.275)		

Table 4 continued

	Dependent variable: average growth rate of GDP over the next four quarters								
Ridge_GDP_P								0.193	
								(0.577)	
Ridge_INF_P									−1.200***
									(0.287)
Constant	1.292**	1.332***	1.211**	1.272**	1.305***	1.286***	1.184**	1.370***	1.271***
	(0.515)	(0.489)	(0.498)	(0.506)	(0.476)	(0.441)	(0.481)	(0.439)	(0.432)
Observations	76	76	76	76	76	76	76	76	76
R^2	0.409	0.424	0.418	0.411	0.410	0.409	0.475	0.411	0.499

Robust (Newey and West 1987) standard errors in parentheses. Maximum lag length is set to 3 in accordance to Estrella and Hardouvelis (1991). ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$

Table 5 Forecasting equations including sentiment indices and control variables for Germany, Inflation, 1993Q1 to 2017Q4

	Dependent variable: average growth rate of inflation over the next four quarters									
Lagged endog. var.	0.116	0.034	−0.002	0.118	0.136	0.094	0.116	0.240	0.117	0.280
	(0.168)	(0.149)	(0.142)	(0.157)	(0.162)	(0.159)	(0.166)	(0.192)	(0.163)	(0.220)
Order inflow	0.110	0.149**	0.154**	0.078	0.089	0.124	0.108	0.105	0.111	0.096
	(0.072)	(0.075)	(0.073)	(0.072)	(0.070)	(0.077)	(0.078)	(0.069)	(0.072)	(0.070)
Interest rate spread	0.052	−0.006	−0.037	0.095	0.096	0.024	0.051	0.042	0.055	0.029
	(0.135)	(0.133)	(0.128)	(0.127)	(0.129)	(0.148)	(0.136)	(0.135)	(0.151)	(0.130)
Ifo business climate	0.255***	0.342***	0.373***	0.182**	0.180*	0.282***	0.262**	0.238***	0.246**	0.255***
	(0.077)	(0.066)	(0.069)	(0.090)	(0.098)	(0.066)	(0.112)	(0.082)	(0.113)	(0.079)
Bannier1		−0.304*								
		(0.169)								
Bannier2			−0.380**							
			(0.162)							
Sharpe1				0.297						
				(0.189)						
Sharpe2					0.234					
					(0.160)					

Table 5 continued

	Dependent variable: average growth rate of inflation over the next four quarters									
SentiWS						−0.155				
						(0.230)				
Lasso_GDP_P							−0.015			
							(0.164)			
Lasso_INF_P								−0.221		
								(0.185)		
Ridge_GDP_P									0.015	
									(0.133)	
Ridge_INF_P										−0.247
										(0.202)
Constant	1.299***	1.399***	1.437***	1.301***	1.270***	1.326***	1.299***	1.109***	1.299***	1.048***
	(0.294)	(0.254)	(0.239)	(0.272)	(0.281)	(0.278)	(0.298)	(0.344)	(0.293)	(0.398)
Observations	76	76	76	76	76	76	76	76	76	76
R^2	0.201	0.252	0.282	0.241	0.235	0.209	0.201	0.224	0.201	0.223

Robust (Newey and West 1987) standard errors in parentheses. Maximum lag length is set to 3 in accordance to Estrella and Hardouvelis (1991). *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

4.3.2 Out-of-sample forecasting performance

To evaluate the pseudo out-of-sample predictive power of the narratives, a reduced forecasting model of Eq. 10 is used to predict the 12-month-ahead average growth rate of real GDP respectively inflation:

$$\hat{Y}_{t|t+h} = \alpha + \sum_{i=1}^{p} \rho_i \hat{Y}_{t-i} + \sum_{j=0}^{q} \beta_j SI_{t-j} + \epsilon_{t+h} \tag{11}$$

Following Ferreira (2018), we include only the lagged endogenous variable to the forecasting model as an additional regressor. The training sample covers 56 observations for the period from 1999Q1 to 2013Q4. The test sample includes 20 observations for the period from 2014Q1 to 2017Q4, which meets the recommended value of 20 per cent of the full sample (Hyndman and Athanasopoulos 2018). The model will be re-estimated at each iteration of the pseudo out-of-sample exercise before each one-step-ahead forecast is computed. The number of lags of the endogenous variable (p) and the predictor variable SI_t (q) will be obtained by minimizing the Akaike information criterion (AIC) at each forecasting period. An autoregressive model is used as a comparative benchmark model. The order of the autoregressive model is also determined by minimizing the Akaike information criterion (AIC) at each forecasting period. In order to evaluate the predictive ability of the narratives, two common forecast evaluation metrics are calculated in a first step. The relative MAE:

$$\text{Relative MAE} = \frac{\frac{1}{T} \sum_{t=1}^{T} \left| e_t^{\text{SI}(k)} \right|}{\frac{1}{T} \sum_{t=1}^{T} \left| e_t^{\text{AR}} \right|} \tag{12}$$

with a linear loss function, and the relative MSE with quadratic loss:

$$\text{Relative MSE} = \frac{\frac{1}{T} \sum_{t=1}^{T} \left(e_t^{\text{SI}(k)} \right)^2}{\frac{1}{T} \sum_{t=1}^{T} \left(e_t^{\text{AR}} \right)^2} \tag{13}$$

is calculated by using the respective forecast error e_t of model 11 in relation to the benchmark autoregressive model. If the value of the relative measure is smaller than 1, the current model outperforms the benchmark model.

In a second step, a Diebold–Mariano test (Diebold and Mariano 1995; Harvey et al. 1997) is employed to test the out-of-sample forecasting performance. To this end, the null hypothesis of equal predictive accuracy (i.e. equal expected loss) between the forecasts with sentiment index and without (benchmark model). The one-sided

Table 6 Out of sample forecasting performance

	Relative MAE	Relative MSE	DM-statistic (linear)	p value (linear)	DM-statistic (quadratic)	p value (quadratic)
Dependent Variable: GDP growth						
Bannier1	1.064	1.063	− 0.486	0.686	− 0.329	0.629
Bannier2	1.106	1.098	− 1.024	0.847	− 0.540	0.705
Sharpe1	1.195	1.482	− 0.760	0.776	− 1.170	0.879
Sharpe2	1.222	1.404	− 0.985	0.838	− 1.397	0.919
SentiWS	1.580	2.611	− 3.441	1.000	− 2.411	0.992
Lasso_GDP_P	0.949	0.945	1.303	0.096	0.888	0.187
Lasso_INF_P	0.938	1.090	0.299	0.383	− 0.267	0.605
Ridge_GDP_P	1.282	1.408	− 1.963	0.975	− 2.063	0.980
Ridge_INF_P	0.858	0.766	0.548	0.292	0.563	0.287
Dependent Variable: Inflation rate						
Bannier1	1.080	1.100	− 0.830	0.797	− 0.553	0.710
Bannier2	1.049	1.082	− 0.470	0.681	− 0.415	0.661
Sharpe1	1.040	1.153	− 0.246	0.597	− 0.616	0.731
Sharpe2	0.950	0.991	0.655	0.256	0.107	0.457
SentiWS	1.017	1.011	− 0.087	0.535	− 0.030	0.512
Lasso_GDP_P	1.158	1.223	− 1.892	0.971	− 1.862	0.969
Lasso_INF_P	1.007	0.931	− 0.137	0.554	1.046	0.148
Ridge_GDP_P	1.071	1.217	− 0.529	0.702	− 0.993	0.840
Ridge_INF_P	0.934	0.889	2.524	0.006	1.975	0.024

Dependent variable: average growth rate of GDP over the next four quarters. DM-test statistic and p values refer to Diebold–Mariano test for predictive accuracy as compared to a simple AR(1) model (Diebold and Mariano 1995; Harvey et al. 1997). Newey and West (1987)-corrected for autocorrelation ($h = 3$)

alternative hypothesis that the forecasts without sentiment index is less accurate:[19]

$$H_0 : L\left(e_t^{\text{AR}}\right) = L\left(e_t^{\text{SI}(k)}\right) \text{ versus } H_1 : L\left(e_t^{\text{AR}}\right) > L\left(e_t^{\text{SI}(k)}\right) \tag{14}$$

where $L(e_t)$ represents the respective linear loss $L(e_t) = e_t$ or quadratic loss $L(e_t) = e_t^2$. Again, the Newey and West (1987) procedure is applied to correct for autocorrelation and the lag length is set equal to 3 ($h − 1$) following Estrella and Hardouvelis (1991).

Table 6 shows the pseudo out-of-sample forecasting performance results for real GDP growth and inflation. The first two columns present the relative forecast performance based on relative MAE and MSE measures. Considering GDP growth,

---

[19] The autoregressive benchmark model can be seen as nested in the sentiment model. Clark and McCracken (2001) show that the asymptotics of the Diebold–Mariano test can fail when comparing nested models. Diebold (2015) demonstrates that the Diebold–Mariano test is still useful and valid for comparing forecasts. Here, we simply ask whether the forecasts of one series is statistically more accurate than another, not whether one forecasting model is better than the other.

two forecasting series with regression-based sentiment indices (LASSO_GDP_P, Ridge_INF_P) beat the benchmark series in both relative measures, MAE and MSE, whereas one forecasting series (LASSO_INF_P) outperforms the benchmark series at least in relative MAE. In contrast, no dictionary-based sentiment index outperforms the benchmark forecasts in relative forecast performance metrics. Statistical tests to check whether the forecasting series with sentiment indices are more accurate as the benchmark forecasts without sentiment indices are given in lines three to six. The Diebold–Mariano tests for linear and quadratic losses do not reject the null hypothesis of equal predictive accuracy for all except one (linear: LASSO_GDP_P) forecasting series with sentiment indices at a conventional significance level. Thus, the generated sentiment indices do not seem to be a statistically powerful out-of-sample predictor for the average growth rate of GDP over the next four quarters. Forecasting performance results for inflation are also given in Table 6. On average, the relative forecast performance of the sentiment series are also weak, measured by the relative MAE and MSE. Again, two forecasting series (Sharpe2, Ridge_INF_P) outperform the benchmark series in relative MAE and relative MSE, whereas one forecast series with LASSO_INF_P index beat the benchmark series in relative MSE. Considering Diebold–Mariano tests, the null hypothesis of equal forecast accuracy can only be rejected for the forecasting series with Ridge_INF_P index in linear and quadrat forecast error environment. To summarize, forecasters' narratives in the form of sentiment indices have weak, at best, predictive ability regarding future GDP growth and inflation in a (pseudo) out-of-sample environment.

5 Discussion and conclusion

Based on 534 business cycle forecast reports covering 10 German institutions for the period 1993–2017, the paper analysed the information content of German forecasters' narratives for German business cycle forecasts and macroeconomic development. In order to do that, textual analysis is used to convert qualitative text data into quantitative sentiment indices.

In a first step, computational textual analysis methods are used to transform forecasters' expectations about the future macroeconomic development into nine sentiment indices.

Second, sentiment analysis shows that the generated sentiment indices vary in their behaviour, pattern, and amplitude. In addition, the sentiment indices differ in their timely relationship to the realized macroeconomic development. Some sentiment indices show nearly a parallel development to the realized value, while other sentiment indices lag behind the real development and a small number of exceptions (partly) lead, compared to the realized value.

Third, sentiment indices are used to test forecast efficiency for GDP growth and inflation forecasts. Using 12-month-ahead fixed horizon forecasts, fixed-effects panel regression results suggest several sentiment indices with informational content for GDP growth and inflation forecasts. German forecasters' narratives can enhance the accuracy of German business cycle forecasts. Overall, the results are in line with the findings of Jones et al. (2020), Sharpe et al. (2020) and

Clements and Reade (2020). The four-quarter forecast horizon is comparable with the results of Sharpe et al. (2020) for the Fed's Greenbook, whereas findings for the UK show shorter forecast horizons (Jones et al. 2020; Clements and Reade 2020).

Fourth, a forecasting exercise analysed the predictive power of sentiment indices for realized growth and inflation. This might explain why forecasters' narratives have predictive power for forecast errors. But the forecasting exercise finds modest evidence, at best, for this hypothesis. The results indicate weak in-sample and out-of-sample predictive power of the sentiment indices for the future stance of the economy. However, more sophisticated forecasting models, e.g. mixed-data sampling (MIDAS) regression models, could improve the results.

There are several explanatory hypotheses as regards why the narratives contain information that is not exhausted by numerical forecasts. One of these is information rigidity. Based on the hypothesis that forecast revisions have predictive power for forecast errors (Nordhaus 1987), Coibion and Gorodnichenko (2015) and Dovern et al. (2015) find some hints supporting this hypothesis using tests for numerical forecasts in an international setting. Kirchgässner and Müller (2006) also find some evidence that German forecasters are reluctant to revise numerical forecasts. In a similar vein, forecasters' narratives could be faster adjusted than their numerical counterparts. Sharpe et al. (2020) analysis for sticky point forecasts could only find weak evidence, at best, for this hypothesis. Another explanatory approach for the predictive power of forecasters' narratives is the 'modal-forecast explanation' (Sharpe et al. 2020, p. 5). This hypothesis is based on the concept that the sentiment indices are particularly informative about tail risks, whereas numerical forecasts unbalance the risks because they are modal rather than mean forecasts. Sharpe et al. (2020) findings suggest such an interpretation. An additional explanation could be that the forecast narrative offers a wider scope for individuality than the quantitative forecast. The numerical forecast is limited to a number. And the production of the forecasts also depends on the institutes' hierarchy and other influencing factors (see e.g. Fritsche and Heilemann 2010, for the Joint Diagnosis). Thus, the forecast report may allow the forecaster a higher degree of freedom. An study of the general issue—why forecasters' narratives have predictive power for forecast errors—could form part of further research.

Last but not least, there is not a single sentiment index or sentiment analysis approach which is generally superior to other methods. The forecast-specific dictionary (Sharpe et al. 2020) and text regression methods perform well in tests for forecast efficiency. Considering the predictive power for GDP growth and inflation, dictionary-based approaches and text regression methods perform relatively weakly. However, the sentiment analysis could be improved in further research using more sophisticated text analysis and machine learning tools.

Appendix

See Tables 7, 8, 9, 10, 11 and 12.

Table 7 List of included institutions and publications

Institution	Dates	Publication schedule	Source
German central bank (Deutsche Bundesbank)	2007–2017	Bi-annually	Monatsberichte der Deutschen Bundesbank
German Institute of Economic Research (DIW Berlin)	1993–2017	Bi-annually (1993–2004), quarterly (2005–2017)	Wochenbericht des DIW
Joint forecast of the German economic research institutes (Gemeinschaftsdiagnose)	1993–2017	Bi-annually	Various publications
Hamburg Institute of International Economics (HWWI) (HWWA until 2006)	1993–2017	Bi-annually (1993–2006), quarterly (2005–2017)	Konjunktur von morgen (1990–1996), Wirtschaftsdienst (1997–2017), occasionally press releases from 2015
Ifo Institute Munich	1993–2017	Bi-annually	Monatsberichte des ifo Instituts für Wirtschaftsforschung (1993–2000), ifo Schnelldienst (2001–2017)
Kiel Institute for World Economics (IfW)	1993–2017	Quarterly	Die Weltwirtschaft (1993–2005), Kieler Diskussionsbeiträge (2006–2014), Kieler Konjunkturberichte (2015–2017)
Macroeconomic Policy Institute (IMK)	2005–2017	Quarterly	IMK report
German Economic Institute (IW Köln)	1995–2017	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

Table 8 Forecasting specific word list: positive words (205 words)

English	German	English	German
assurance	Zusicherung	endorse	billigen
assure	versichern	energetic	energetisch
attain	erreichen	engage	engagieren
attractive	attraktiv	enhance	verbessern
auspicious	vielversprechend	enhancement	Verbesserung
backing	Unterstützung	enjoy	genießen
befitting	angemessen	enrichment	Anreicherung
beneficial	vorteilhaft	enthusiasm	Begeisterung
beneficiary	Begünstigter	enthusiastic	enthusiastisch
benefit	Vorteil	envision	vorstellen
benign	gutartig	excellent	exzellent
better	besser	exuberance	Überschwang
bloom	Blütezeit	exuberant	überschwänglich
bolster	Nackenrolle	facilitate	erleichtern
boom	Boom	faith	Glaube
boost	Verstärkung	favor	Gefälligkeit
bountiful	freigiebig	favorable	günstig
bright	hell	feasible	durchführbar
buoyant	schwungvoll	fervor	Inbrunst
calm	ruhig	filial	kindlich
celebrate	feiern	flatter	flacher
coherent	kohärent	flourish	blühen
comeback	wiederbelebung	fond	zärtlich
comfort	Komfort	foster	fördern
comfortable	komfortabel	friendly	freundlich
commend	empfehlen	gain	Gewinn
compensate	kompensieren	generous	großzügig
composure	Gelassenheit	genuine	echt
concession	Konzession	good	gut
concur	übereinstimmen	happy	glücklich
conducive	förderlich	heal	heilen
confide	anvertrauen	healthy	gesund
confident	selbstbewusst	helpful	hilfreich
constancy	Beständigkeit	hope	Hoffnung
constructive	konstruktiv	hopeful	hoffnungsvoll
cooperate	kooperieren	hospitable	gastfreundlich
coordinate	Koordinate	imperative	unerlässlich

Table 8 continued

English	German	English	German
credible	glaubwürdig	impetus	Impulsgeber
decent	anständig	impress	beeindrucken
definitive	definitiv	impressive	beeindruckend
deserve	verdienen	improve	verbessern
desirable	wünschenswert	improvement	Verbesserung
discern	erkennen	inspire	inspirieren
distinction	Unterscheidung	irresistible	unwiderstehlich
distinguish	unterscheiden	joy	Freude
durability	Haltbarkeit	liberal	liberal
eager	begierig	lucrative	lukrativ
earnest	ernst	manageable	überschaubar
ease	Leichtigkeit	mediate	vermitteln
easy	einfach	mend	ausbessern
encourage	ermutigen	mindful	achtsam
encouragement	Ermutigung		
moderation	Moderation	revival	Aufschwung
onward	vorwärts	revive	wieder aufleben
opportunity	Gelegenheit	ripe	reif
optimism	Optimismus	rosy	rosig
optimistic	optimistisch	salutary	heilkräftig
outrun	überschreiten	sanguine	blutrot
outstanding	ausstehend	Satisfactory	Zufriedenstellend
overcome	überwinden	Satisfy	Befriedigen
paramount	hervorragend	Sound	Sound
particular	speziell	Soundness	Solidität
patience	Geduld	Spectacular	Spektakulär
patient	Patient	Stabilize	Stabilisieren
peaceful	friedlich	Stable	Stabil
persuasive	überzeugend	Stable	Stabil
pleasant	angenehm	Steadiness	Stetigkeit
please	bitte	Steady	Langsam
pleased	zufrieden	Stimulate	Stimulieren
plentiful	reichlich	Stimulation	Stimulation
plenty	Fülle	Subscribe	Abonnieren

Table 8 continued

English	German	English	German
positive	positiv	Succeed	Erfolgreich
potent	stark	Success	Erfolg
precious	kostbar	Successful	Erfolgreich
pretty	hübsch	Suffice	Es genügt
progress	Fortschritt	Suit	Anzug
progressive	progressiv	Support	Unterstützung
prominent	bekannt	Supportive	Unterstützende
promise	Versprechen	Surge	Surge
prompt	Eingabeaufforderung	Surpass	Übertrifft
proper	ordentlich	Sweeten	Süßstoff
prosperity	Wohlstand	Sympathetic	Sympathisch
rally	Kundgebung	Sympathy	Mitgefühl
readily	bereit	Synthesis	Synthese
reassure	beruhigen	Temperate	Gemüßigt
receptive	empfänglich	Thorough	Gründlich
reconcile	versöhnen	Tolerant	Tolerant
refine	verfeinern	tranquil	ruhig
reinstate	wiederherstellen	tremendous	riesig
relaxation	Entspannung	undoubtedly	zweifellos
reliable	zuverlässig	unlimited	unbegrenzt
relief	Erleichterung	upbeat	optimistisch
relieve	entlasten	upgrade	Upgrade
remarkable	bemerkenswert	uplift	Auftrieb
remarkably	bemerkenswert	upside	aufwärts
repair	Reparatur	upward	nach oben
rescue	Rettung	valid	gültig
resolve	auflösen	viable	tragfähig
resolved	gelöst	victorious	siegreich
respectable	respektabel	virtuous	tugendhaft
respite	Aufschub	vitality	Vitalität
restoration	Wiederherstellung	warm	warm
restore	wiederherstellen	welcome	willkommen

Own translation based on DeepL Pro. Based on the English version (Sharpe et al. 2020)

Table 9 Forecasting specific word list: negative words (103 words)

English	German	English	German
adverse	nachteilig	hurt	verletzt
afflict	belasten	illegal	illegal
alarming	beunruhigend	insecurity	Unsicherheit
apprehension	Besorgnis	insidious	heimtückisch
apprehensive	ängstlich	instability	Instabilität
awkward	unangenehm	interfere	eingreifen
bad	schlecht	jeopardize	gefährden
badly	schlecht	jeopardy	Gefahr
bitter	bitter	lack	Mangel
bleak	trostlos	languish	schmachten
bug	Fehler	loss	Verlust
burdensome	beschwerlich	mishap	Missgeschick
corrosive	korrosiv	negative	negativ
danger	Gefahr	nervousness	Nervosität
daunting	beängstigend	offensive	beleidigend
deadlock	Sackgasse	painful	schmerzhaft
deficient	unzulänglich	paltry	armselig
depress	niederdrücken	pessimistic	pessimistisch
depression	Krise	plague	Plage
destruction	Vernichtung	plight	Notlage
devastation	Abbau	poor	schlecht
dim	schwach	recession	Rezession
disappoint	enttäuschen	sank	gesunken
disappointment	Enttäuschung	scandal	Skandal
disaster	Katastrophe	scare	schreck
discomfort	Unbehagen	sequester	absondern
discouragement	Entmutigung	sluggish	träge
dismal	trostlos	slump	Einbruch
disrupt	unterbrechen	sour	sauer
disruption	Störung	sputter	spritzen
dissatisfied	unzufrieden	stagnant	stagnierend
distort	verzerren	standstill	Stillstand
distortion	Verzerrung	struggle	kämpfen
distress	Notlage	suffer	ertragen
doldrums	Flaute	terrorism	Terrorismus
downbeat	deprimierend	threat	Bedrohung
emergency	Notfall	tragedy	Tragödie
erode	erodieren	tragic	tragisch
fail	scheitern	trouble	Ärger
failure	Versagen	turmoil	Aufruhr

Table 9 continued

English	German	English	German
fake	Fälschung	unattractive	unattraktiv
falter	zögern	undermine	untergraben
feeble	schwach	undesirable	unerwünscht
feverish	fieberhaft	uneasiness	Unbehagen
fragile	zerbrechlich	uneasy	unbehaglich
gloom	Tristesse	unfavorable	ungünstig
gloomy	düster	unforeseen	unvorhergesehen
grim	grimmig	unprofitable	unrentabel
harsh	rau	unrest	Unruhe
havoc	Verwüstung	violent	gewalttätig
hit	treffen	war	Krieg
horrible	schrecklich		

Own translation based on DeepL Pro. Based on the English version (Sharpe et al. 2020)

Table 10 Dictionary and weights—Lasso GDP (71 words)

Words	Weight	Words	Weight
aufschwung	0.0939	steuersenkungen	0.0000
fortsetzung	0.0458	schwellenländ	0.0000
fortgesetzt	0.0358	fiskalischen	0.0000
erreicht	0.0294	inlandsnachfrag	0.0000
bewirkt	0.0272	beachten	0.0000
schwungvol	0.0260	fort	0.0000
erweiterungsinvestitionen	0.0246	einstellen	0.0000
steigend	0.0217	historisch	0.0000
dynamik	0.0207	absatzperspektiven	− 0.0009
verbrauchskonjunktur	0.0192	abschwung	− 0.0015
einsparmaßnahmen	0.0188	erholen	− 0.0017
exportdynamik	0.0180	geld	− 0.0026
japan	0.0159	konjunkturschwäch	− 0.0030
schwellenländern	0.0134	bezugsdau	− 0.0047
guten	0.0100	historischen	− 0.0050
südostasien	0.0094	minus	− 0.0053
westdeutschland	0.0092	anpassungen	− 0.0061
belaufen	0.0087	extrem	− 0.0065
abflachen	0.0081	massiv	− 0.0083
ostdeutschland	0.0080	konjunkturpaket	− 0.0118
hohen	0.0059	getroffen	− 0.0174
lohnabschlüss	0.0052	talfahrt	− 0.0187
betragen	0.0048	verschlechterten	− 0.0220

Table 10 continued

Words	Weight	Words	Weight
drittel	0.0047	schwachen	− 0.0247
industrieländern	0.0039	verschlechtert	− 0.0253
turbulenzen	0.0038	schlechten	− 0.0332
vorkrisenniveau	0.0018	schrumpfen	− 0.0337
erreichten	0.0014	unterauslastung	− 0.0342
läßt	0.0011	einbruch	− 0.0342
außenwert	0.0008	unterhang	− 0.0389
reichlich	0.0006	entlassungen	− 0.0392
unterschied	0.0003	tief	− 0.0479
wachstum	0.0002	reduzieren	− 0.0480
längerfristigen	0.0001	stabilisierung	− 0.1052
arbeitslosenversicherung	0.0000	drastisch	− 0.1427
wachstumspakt	0.0000		

Full sample example. The table presents the LASSO text regression-based dictionary with weights over the full corpus from 1993–2017. The weights are rounded to the 4th decimal. Response variable: GDP forecast (12-month-ahead fixed horizon forecast)

Table 11 Dictionary and weights—Lasso inflation (69 words)

Words	Weight	Words	Weight
investoren	0.0369	ältere	0.0005
lockerung	0.0232	mehrwertsteu	0.0001
tarifabschlüss	0.0229	mehrwertsteuererhöhung	0.0000
nachlassenden	0.0201	stützt	0.0000
abkühlung	0.0200	flüchtling	− 0.0005
nahrungsmittelpreis	0.0198	jahresverlaufsr	− 0.0006
arbeitslosenversicherung	0.0181	unsich	− 0.0009
halb	0.0171	abwertung	− 0.0010
durchsetzen	0.0168	esvg	− 0.0012
treuhandanstalt	0.0161	leistungsausweitungen	− 0.0013
zurückbilden	0.0127	abnehmenden	− 0.0019
westeuropa	0.0125	stufe	− 0.0034
westdeutsch	0.0119	einbruch	− 0.0035
kurzfristigen	0.0117	investitionsausgaben	− 0.0035
westdeutschen	0.0112	gang	− 0.0038
eingestellt	0.0109	stützen	− 0.0039
lohnabschlüss	0.0098	gesamt	− 0.0039
verlust	0.0096	drastisch	− 0.0052
gute	0.0084	land	− 0.0055
bundesrepublik	0.0079	kranken	− 0.0055
steuererhöhungen	0.0079	entscheidungen	− 0.0079

Table 11 continued

Words	Weight	Words	Weight
zahlungen	0.0079	zugang	− 0.0083
insolvenzgeldumlag	0.0075	krankenkassen	− 0.0096
preisauftrieb	0.0072	festigung	− 0.0108
staatsausgaben	0.0066	kindergeld	− 0.0142
lohnpolitik	0.0052	bundesverfassungsgericht	− 0.0149
sparneigung	0.0049	niedrigen	− 0.0155
tätigen	0.0039	wirkung	− 0.0175
konjunkturindikatoren	0.0026	expansiv	− 0.0204
abschreibungsbedingungen	0.0025	stabil	− 0.0255
inlandskonzept	0.0019	erholung	− 0.0296
gelegen	0.0019	unterauslastung	− 0.0335
beitragssatz	0.0018	arbeitslosigkeit	− 0.0547
schwächeren	0.0015	gesunkenen	− 0.0582
mäßig	0.0013		

Full sample example. The table presents the LASSO text regression-based dictionary with weights over the full corpus from 1993–2017. The weights are rounded to the 4th decimal. Response variable: inflation forecast (12-month-ahead fixed horizon forecast)

Table 12 Dictionary and weights—Ridge GDP and inflation—Top 30 positive and negative words in descending order

Positive words	Weight	Negative words	Weight
Ridge GDP			
fortsetzung	0.0095	wirken	− 0.0061
fortsetzen	0.0081	geld	− 0.0062
zunehmen	0.0081	verhindert	− 0.0063
aufschwung	0.0079	druck	− 0.0063
dynamik	0.0075	einbruch	− 0.0064
schwungvol	0.0074	instrument	− 0.0066
steigend	0.0071	massiv	− 0.0067
bleiben	0.0071	zurückgehen	− 0.0069
privaten	0.0067	einmalig	− 0.0070
steigenden	0.0066	rückgang	− 0.0071
erweiterungsinvestitionen	0.0065	talfahrt	− 0.0074
bleibt	0.0063	einstellen	− 0.0075
verbessert	0.0063	schlechten	− 0.0077
günstig	0.0060	reduzieren	− 0.0078
fortgesetzt	0.0059	unterauslastung	− 0.0079
hohen	0.0058	rückläufigen	− 0.0079
bewirkt	0.0057	schwachen	− 0.0080
verbrauchskonjunktur	0.0056	schrumpfen	− 0.0085

Table 12 continued

Positive words	Weight	Negative words	Weight
investitionsdynamik	0.0056	verschlechtert	− 0.0085
betragen	0.0055	unterhang	− 0.0087
erreicht	0.0055	erholen	− 0.0087
höheren	0.0053	tief	− 0.0087
erstmal	0.0051	unternehmen	− 0.0088
lohnabschlüss	0.0050	stabilisieren	− 0.0088
kräftige	0.0050	entwicklung	− 0.0088
investitionsklima	0.0050	sinken	− 0.0094
kräftigen	0.0049	verschlechterten	− 0.0096
konjunkturaufschwung	0.0049	entlassungen	− 0.0107
stütze	0.0049	stabilisierung	− 0.0117
guten	0.0048	drastisch	− 0.0118
Ridge inflation			
unternehmen	0.0085	ursächlich	− 0.0037
tarifabschlüss	0.0073	preiserhöhungsspielräum	− 0.0038
durchsetzen	0.0072	zehnjährig	− 0.0038
real	0.0071	durchgeführt	− 0.0038
investoren	0.0067	wirken	− 0.0039
trotz	0.0062	eingeschränkt	− 0.0039
lohnabschlüss	0.0061	zentralbank	− 0.0039
konjunkturindikatoren	0.0060	krankenkassen	− 0.0040
nachlassenden	0.0059	festigung	− 0.0040
nahrungsmittelpreis	0.0059	gang	− 0.0040
mäßig	0.0052	kranken	− 0.0041
abkühlung	0.0052	drastisch	− 0.0042
lockerung	0.0051	quartalsdurchschnitt	− 0.0043
halb	0.0050	jahresveränderungsr	− 0.0045
aufwärtsentwicklung	0.0049	druck	− 0.0045
marktanteil	0.0049	früherer	− 0.0046
transferzahlungen	0.0048	profitieren	− 0.0046
eckdaten	0.0047	bundesverfassungsgericht	− 0.0051
gesamtwirtschaftlich	0.0047	erholung	− 0.0051
schwächeren	0.0045	abnehmenden	− 0.0051
tendenziel	0.0045	stabil	− 0.0054
steuererhöhungen	0.0045	erwerbspersonen	− 0.0054
beleben	0.0044	stützen	− 0.0056
bestimmt	0.0044	unterauslastung	− 0.0058
hohen	0.0043	niedrigen	− 0.0059
sparneigung	0.0043	bleibt	− 0.0061

Table 12 continued

Positive words	Weight	Negative words	Weight
vordergrund	0.0041	gesunkenen	− 0.0068
inlandskonzept	0.0041	arbeitsmarkt	− 0.0084
verbesserung	0.0040	bleiben	− 0.0095
lohnpolitik	0.0040	arbeitslosigkeit	− 0.0113

The table presents the Ridge text regression-based dictionary with weights over the full corpus from 1993 to 2017. The weights are rounded to the 4th decimal. Response variable: GDP forecast (12-month-ahead fixed horizon forecast)

References

Andrews DW (1991) Heteroskedasticity and autocorrelation consistent covariance matrix estimation. Econometrica 59(3):817–858

Arellano M, Bond S (1991) Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. Rev Econ Stud 58(2):277–297

Ashiya M (2006) Forecast accuracy and product differentiation of Japanese institutional forecasters. Int J Forecast 22(2):395–401

Baker S, Bloom N, Davis S (2016) Measuring economic policy uncertainty. Q J Econ 131(4):1593–1636

Bannier CE, Pauls TT, Walter A (2018) Content analysis of business communication: introducing a German dictionary. J Bus 89(1):79–123

Batchelor RA (1990) All forecasters are equal. J Bus Econ Stat 8(1):143–144

Beck N, Katz JN (1995) What to do (and not to do) with time-series cross-section data. Am Political Sci Rev 89(3):634–647

Berge TJ, Jordà Ò (2011) Evaluating the classification of economic activity into recessions and expansions. Am Econ J Macroecon 3(2):246–277

Clark TE, McCracken MW (2001) Tests of equal forecast accuracy and encompassing for nested models. J Econom 105(1):85–110

Clements MP, Reade JJ (2020) Forecasting and forecast narratives: the bank of England inflation reports. Int J Forecast 36(4):1488–1500

Coibion O, Gorodnichenko Y (2015) Information rigidity and the expectations formation process: a simple framework and new facts. Am Econ Rev 105(8):2644–78

Deutsche Bundesbank (2020) Time series data base. https://www.bundesbank.de/Navigation/EN/Statistics/Time_series_databases/time_series_databases.html. Accessed 5 April 2020

Di Fatta G, Reade JJ, Jaworska S, Nanda A (2015) Big social data and political sentiment: the tweet stream during the UK general election 2015 campaign. In: 2015 IEEE international conference on smart city/socialcom/sustaincom (smartcity). IEEE, pp 293–298

Diebold FX (2015) Comparing predictive accuracy, twenty years later: a personal perspective on the use and abuse of Diebold–Mariano tests. J Bus Econ Stat 33(1):1–9

Diebold FX, Lopez JA (1996) Forecast evaluation and combination. Handb Stat 14:241–268

Diebold FX, Mariano RS (1995) Comparing predictive accuracy. J Bus Econ Stat 13(3):253–263

Dimpfl T, Kleiman V (2019) Investor pessimism and the German stock market: exploring google search queries. German Econ Rev 20(1):1–28

Döhrn R, Schmidt C (2011) Information or institution? On the determinants of forecast accuracy. J Econ Stat (Jahrbuecher fuer Nationalökonomie und Statistik) 231(1):9–27

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

Döpke J, Fritsche U, Siliverstovs B (2010) Evaluating German business cycle forecasts under an asymmetric loss function. OECD J J Bus Cycle Meas Anal 1:1–18

Döpke J, Fritsche U, Müller K (2019) Has macroeconomic forecasting changed after the great recession? Panel-based evidence on forecast accuracy and forecaster behavior from Germany. J Macroecon 62:103–135

Dovern J, Fritsche U (2008) Estimating fundamental cross-section dispersion from fixed event forecasts (787). DIW Berlin Discussion Paper

Dovern J, Fritsche U, Slacalek J (2012) Disagreement among forecasters in G7 countries. Rev Econ Stat 94(4):1081–1096

Dovern J, Fritsche U, Loungani P, Tamirisa N (2015) Information rigidities: comparing average and individual forecasts for a large international panel. Int J Forecast 31(1):144–154

Estrella A, Hardouvelis GA (1991) The term structure as a predictor of real economic activity. J Finance 46(2):555–576

Federal Statistical Office (2019a) Preise, Verbraucherpreisindizes für Deutschland, Lange Reihen ab 1948. https://www.destatis.de. Accessed 14 Mar 2019

Federal Statistical Office (2019b) Volkswirtschaftliche gesamtrechnungen, Bruttoinlandsprodukt ab 1970, Vierteljahres- und Jahresergebnisse. https://www.destatis.de. Accessed 14 Sept 2019

Ferreira T (2018) Stock market cross-sectional skewness and business cycle fluctuations. International Finance Discussion Papers 1223, Board of Governors of the Federal Reserve System (U.S.). https://www.fedinprint.org/items/fedgif/1223.html. Accessed 5 May 2020

Fildes R, Stekler H (2002) The state of macroeconomic forecasting. J Macroecon 24(4):435–468

Fortin I, Koch SP, Weyerstrass K (2020) Evaluation of economic forecasts for Austria. Empir Econ 58(1):107–137

Fritsche U, Heilemann U (2010) Too many cooks? The German joint diagnosis and its production. Technical Report 1/2010, DEP (Socioeconomics) Discussion Papers, Macroeconomics and Finance Series

Fritsche U, Puckelwald J (2018) Deciphering professional forecasters' stories: analyzing a corpus of textual predictions for the German economy. DEP (Socioeconomics) discussion papers—macroeconomics and finance series 4/2018, Hamburg. http://hdl.handle.net/10419/194021

Fritsche U, Tarassow A (2017) Vergleichende Evaluation der Konjunkturprognosen des Instituts für Makroökonomie und Konjunkturforschung an der Hans-Böckler-Stiftung für den Zeitraum 2005-2014. IMK Study 54, Düsseldorf. http://hdl.handle.net/10419/156388

Gaibulloev K, Sandler T, Sul D (2014) Dynamic panel analysis under cross-sectional dependence. Political Anal 22(2):258–273

Garcia D (2013) Sentiment during recessions. J Finance 68(3):1267–1300

Gentzkow M, Kelly B, Taddy M (2019) Text as data. J Econ Lit 57(3):535–74

Goldfarb RS, Stekler HO, David J (2005) Methodological issues in forecasting: insights from the egregious business forecast errors of late 1930. J Econ Methodol 12(4):517–542

Greene WH (2012) Econometric analysis, 7th edn. Pearson, New York

Harvey D, Leybourne S, Newbold P (1997) Testing the equality of prediction mean squared errors. Int J Forecast 13(2):281–291

Heilemann U, Müller K (2018) Wenig Unterschiede-Zur Treffsicherheit internationaler Prognosen und Prognostiker. AStA Wirtschafts-und Sozialstatistisches Archiv 12(3–4):195–233

Heilemann U, Stekler HO (2013) Has the accuracy of macroeconomic forecasts for Germany improved? German Econ Rev 14(2):235–253

Heinisch K, Scheufele R (2018) Bottom-up or direct? Forecasting German GDP in a data-rich environment. Empir Econ 54(2):705–745

Heppke-Falk K, Hüfner FP (2004) Expected budget deficits and interest rate swap spreads-evidence for France, Germany and Italy. Deutsche Bundesbank Discussion Paper (40/2004)

Holden K, Peel DA (1990) On testing for unbiasedness and efficiency of forecasts. Manch Sch 58(2):120–127

Hyndman R, Athanasopoulos G (2018) Forecasting: principles and practice, 2nd edn. OTexts. https://otexts.com/fpp2/. Accessed 03 Sept 2020

Ifo institute (2020) Business climate index. http://www.cesifo-group.de/ifoHome/facts/Survey-Results/Business-Climate.html. Accessed 04 May 2020

Jegadeesh N, Wu D (2013) Word power: a new approach for content analysis. J Financial Econ 110(3):712–729

Jegadeesh N, Wu DA (2017) Deciphering fedspeak: the information content of fomc meetings. Technical report, SSRN. https://ssrn.com/abstract=2939937. Accessed 19 Oct 2019

Jivani AG (2011) A comparative study of stemming algorithms. Int J Comput Technol Appl 2(6):1930–1938

Jones JT, Sinclair TM, Stekler HO (2020) A textual analysis of bank of England growth forecasts. Int J Forecast 36(4):1478–1487

Kauder B, Potrafke N, Schinke C (2017) Manipulating fiscal forecasts: evidence from the German states. FinanzArchiv Public Finance Anal 73(2):213–236

Keane MP, Runkle DE (1990) Testing the rationality of price forecasts: new evidence from panel data. Am Econ Rev 80:714–735

Kirchgässner G, Müller UK (2006) Are forecasters reluctant to revise their predictions? Some German evidence. J Forecast 25(6):401–413

Krüger JJ, Hoss J (2012) German business cycle forecasts, asymmetric loss and financial variables. Econ Lett 114(3):284–287

Lamla MJ, Lein SM, Sturm JE (2020) Media reporting and business cycles: empirical evidence based on news data. Empir Econ 59(3):1085–1105

Liu W, Moench E (2016) What predicts US recessions? Int J Forecast 32(4):1138–1150

Loughran T, McDonald B (2011) When is a liability not a liability? Textual analysis, dictionaries, and 10-ks. J Finance 66(1):35–65

Loughran T, McDonald B (2016) Textual analysis in accounting and finance: a survey. J Account Res 54(4):1187–1230

Lundquist K, Stekler HO (2012) Interpreting the performance of business economists during the great recession. Bus Econ 47(2):148–154

Manela A, Moreira A (2017) News implied volatility and disaster concerns. J Financ Econ 123(1):137–162

Mathy G, Stekler H (2018) Was the deflation of the depression anticipated? An inference using real-time data. J Econ Methodol 25(2):117–125

Merton RC (1981) On market timing and investment performance. I. An equilibrium theory of value for market forecasts. J Bus 54:363–406

Newey W, West K (1987) A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. Econometrica 55(3):703–08

Nickell S (1981) Biases in dynamic models with fixed effects. Econometrica 1417–1426

Nordhaus WD (1987) Forecasting efficiency: concepts and applications. Rev Econ Stat 69(4):667–674

Pierdzioch C, Rülke JC (2015) On the directional accuracy of forecasts of emerging market exchange rates. Int Rev Econ Finance 38:369–376

Porter MF et al (1980) An algorithm for suffix stripping. Program 14(3):130–137

Pröllochs N, Feuerriegel S, Neumann D (2015) Generating domain-specific dictionaries using Bayesian learning. ECIS 2015 Completed Research Papers (Paper 144)

Pröllochs N, Feuerriegel S, Neumann D (2018) Statistical inferences for polarity identification in natural language. PLoS ONE 13(12):1–21

Remus R, Quasthoff U, Heyer G (2010) Sentiws—a publicly available german-language resource for sentiment analysis. In: Calzolari N, Choukri K, Maegaard B, Mariani J, Odijk J, Piperidis S, Rosner M, Tapias D (eds) Proceedings of the 7th international language resources and evaluation (LREC'10), European language resources association (ELRA), Valletta, Malta

Shapiro AH, Sudhof M, Wilson D (2020) Measuring news sentiment. Federal Reserve Bank of San Francisco. Working Paper 2017-01

Sharpe SA, Sinha NR, Hollrah CA (2020) The power of narratives in economic forecasts. FEDS Working Paper (2020-001)

Smant DJ (2002) Has the European Central Bank followed a Bundesbank policy? Evidence from the early years. Kredit und Kapital 35(3):327–343

Stekler H, Symington H (2016) Evaluating qualitative forecasts: the fomc minutes, 2006–2010. Int J Forecast 32(2):559–570

Stock JH, Watson MW (2003) Forecasting output and inflation: the role of asset prices. J Econ Lit 41(3):788–829

Tetlock PC (2007) Giving content to investor sentiment: the role of media in the stock market. J Finance 62(3):1139–1168

Tetlock PC, Saar-Tsechansky M, Macskassy S (2008) More than words: quantifying language to measure firms fundamentals. J Finance 63(3):1437–1467

Tibshirani R (1996) Regression shrinkage and selection via the lasso. J R Stat Soc Ser B (Methodol) 58(1):267–288

Tillmann P, Walter A (2018) ECB vs Bundesbank: diverging tones and policy effectiveness. MAGKS Joint Discussion Paper Series in Economics 20-2018, Marburg

Tobback E, Naudts H, Daelemans W, de Fortuny EJ, Martens D (2018) Belgian economic policy uncertainty index: improvement through text mining. Int J Forecast 34(2):355–365

Varian HR (2014) Big data: new tricks for econometrics. J Econ Perspect 28(2):3–28

Zipf GK (1949) Human behavior and the principle of least effort. Addison-Wesley, Cambridge