Linking Words in Economic Discourse: Implications for Macroeconomic Analysis

J. Daniel Aromí

May 11, 2018

Abstract

Indicators of unstructured information in the press are developed using a word vector representation model. The information content of these indicators is assessed through business cycle prediction tasks. Word vector representations are trained using the GloVe model (Pennington et al. 2014) and a corpus covering 90 years of press content. The representations are shown to learn meaningful associations in economic context. These associations are exploited to develop indicators of uncertainty. In-sample and out-of-sample forecasting exercises show that the indicators contain valuable information regarding future economic activity. The combination of indices associated to different subjective states (e.g. uncertainty, fear, pessimism) is shown to result in further gains in information content. Alternative text analysis techniques previously proposed in the literature are not seen to capture as much information.

1 Introduction

A large quantity of unstructured economic data is generated and disseminated everyday through multiple channels. For example, this type of data is found in corporate and government documents, expert reports, mass media and social media. Improvements in data availability and processing capacity have allowed for studies that summarize and evaluate the information provided by unstructured data. Multiple works have demonstrated the relevance of unstructured data in macroeconomic and financial contexts (Tetlock 2007, Loughran & MacDonald 2011, Alexopoulus & Cohen 2015, Baker et al. 2016, Aromi 2017a 2017c, Hansen et al. 2017). These contributions typically compute interpretable indicators that are based on a small set of keywords or predefined dictionaries. The resulting indicators are interpreted as metrics of uncertainty, pessimism or the level of attention allocated to a topic of interest (e.g. recession).

One relevant question is whether natural language processing tools can be used to learn to interpret information in a more efficient manner. For example, can valuable indicators of uncertainty be built using these tools? How does the performance of these indices compare with the performance observed under more traditional methods? The extent to which gains are realized is a function of the efficiency of learning algorithms and the informativeness of the training corpus. Positive results would be relevant in macroeconomic analysis. More precise metrics can lead to the discovery of empirical regularities or the revision of previously estimated associations. In this work, the performance of an specific natural language processing tool is evaluated in the context of business cycle prediction tasks.

More specifically, this work considers the use of word vector representations (WVRS), an unsupervised learning tool that has been successfully tested in natural language processing applications (Collobert et al. 2011, Mikolov et al. 2013, Pennington et al. 2014). Word vector representations are trained to learn meaning in economic context using a large corpus of information in the press. The resulting structure of meaning allows for indicators with a straightforward interpretation. Following Pennington (2014), the GloVe (global vectors) model is used to compute the representations. The similarity between words, as indicated by their vector representations, is used to generate quantitative indicators of information in the press.

Preliminary evaluations show that the resulting vectors are able to learn meaning in economic context. Vectors are shown to resolve word ambiguity and recognize relationships between economic entities. For example, the word "vice" can be thought to refer to immoral behavior or to corporate positions (e.g. Vice President). Trained vectors are seen to resolve this ambiguity in favor of the second option. According to the distance between vectors, the closest words are "executive" and "president". More relevant for the current work, WVRs are shown to identify sets of related words (e.g. words related to uncertainty). In this way, it is suggested that WVRs can allow for indicators in which, instead of using predefined dictionaries or subjective judgment, relevant words are identified through quantitative information generated by unsupervised learning algorithms.

In the first set of exercises, a metric of uncertainty is evaluated. This choice reflects the prominent role assigned to the concept of uncertainty in the analysis of business cycles (see for example Jurado et al. 2015, Baker et al. 2016, Rossi & Sekhposyan 2015). In-sample, the indicators are shown to provide information on future levels of employment, industrial production, investment and GDP. A one standard deviation increment in the uncertainty index anticipates, on average, a 0.40 standard deviations drop in GDP growth over the next year. Out-of-sample exercises are implemented using Bayesian model averaging (BMA). The forecast combination tool allow for data-driven learning of efficient specifications of indicators from the press. These exercises show the indicators that exploit word vector representations allow for significant gains in forecast accuracy.

Beyond indicators approximating uncertainty, complementary explorations consider indices capturing manifestations of different subjective states that are conjectured to be relevant. This additional subjective states are identified inspecting previous literature and recurring to subjective judgment. The resulting indices summarize manifestations linked to pessimism, fear and anxiety. Out of sample forecast exercises show that these alternative indices contain additional information that can be combined to attain higher accuracy.

The extent to which indices based on natural language processing techniques are more precise than more traditional methods is unknown. Traditional methods are based on knowledge in the form dictionaries and subjective judgment. As a result, they incorporate information that might allow for precise metrics. In the final set of exercises, the information content that results from traditional methods is compared to information content that results from the use of word vector representations. A new set of business cycle prediction exercises are implemented with that purpose.

Four traditional text analysis methodologies, as proposed in Baker et al. (2016), Loughram & McDonald (2011), Tetlock (2007) and The Economist[1], are considered. In Baker et al. (2016) and The Economist's R-word index small sets of words are selected to generate indices of press content. In Tetlock (2007) and Loughran & MacDonald (2011) the indices are based on large lists built using pre-defined dictionaries and expert judgment. Forecasting exercises show that the performance of indices based on word vector representations compares favorably with that verified under alternative techniques previously proposed in the literature. For example, in the case of one-year-ahead GDP forecasts, the predictions based on traditional indices are less accurate than baseline forecasts. In contrast, uncertainty indices based on word vector representations are able to generate highly significant improvements in forecast accuracy.

The rest of the paper is organized as follows. The next section presents the methodology and the data. The properties of trained word vectors are preliminary explored in section 3. Forecasting exercises associated to the metric of uncertainty are presented in section 4. The next section evaluates indices approximating alternative subjective states. In section 6, comparisons with other methodologies are presented. Section 7 concludes.

2 Methodology and data

The construction of the indicators proposed in this work involves two steps. In the first step, word vector representations are trained using a corpus covering ninety years of economic press content. In this way, a structure of meaning is built. In the second step, indicators that summarize relevant aspects of information in the press are computed. This involves identifying a relevant keyword and exploiting associations in trained word vector representations. More specifically, having selected a relevant keyword or set of keywords (e.g. uncertainty), closely associated words are identified computing the distance between their respective word vector representations. The indicator is given by the frequency of these closely associated words. In forecasting exercises, the indices are computed using a second, non-overlapping corpus. In this section, the methodology is outlined in more detail and a description of the training corpus is provided.

2.1 Word vector representations

As previously indicated, the first step involves learning to represent words through vectors using the GloVe model (Pennington et al. 2014). This type of representation

[1] "The R-word" (2001, Apr, 5th), The Economist. Retrieved from http://www.economist.com/

has been shown to efficiently summarize word semantic (and syntactic) information (Collobert et al. 2011, Mikolov et al. 2013, Pennington et al. 2014). As suggested by the authors, it can be understood as a linear structure of meaning. This quantitative representation can be used to assess relatedness between different words. Relatedness is established computing the distance between the respective vectors. Also, information provided by multiple words can be combined through simple algebraic operations. While GloVe is not the only method that computes vector representations of words, it has been shown to perform better than alternative methods in multiple natural language processing tasks (see Pennington 2014).

In the GloVe model, word vectors are trained to capture information on word co-occurrences in the training corpus. The method is global in the sense that all vectors are computed in a single optimization exercise. Let W denote a dictionary and let X_{ij} denote the number of times word i occurs in the context of word j. The loss function of the GloVe model is given by:

$$\sum_i \sum_j f(X_{ij}) \left[v_i \cdot \tilde{v}_j + b_i + \tilde{b}_j - log(X_{ij}) \right]^2$$

Where v_i and $\tilde{v}_j$ are word vectors, $f(X_{ij})$ is a weighting function and b_i and $\tilde{b}_j$ are word biases.[2] This is a log-bilinear regression model. The weighting function $f(X_{ij})$ is increasing and concave.[3] The vector representations are trained using stochastic gradient descent (Duchi et al. 2011). More details can be found in Pennington et al. (2014).

Following parameter values that are in line with those used in the natural language processing literature, the vector dimensionality is 100 and the window size used to compute term co-occurrence is 5. The vocabulary used in the implementation is given by words with a frequency of 100 or higher in the training corpus. Robustness analyses indicate that the results are not sensible to variations in the value of these parameters. Vector representations of words are computed using package text2vec in platform R. The same package was used in other related computations (e.g. tokenization, term co-occurrence matrix).

2.2 Quantitative indicators

In the second step, word vector representations are used to construct quantitative indicators of information in the press. The intention is to generate indicators that exploit knowledge captured by word vectors and can be interpreted in a straightforward manner. The procedure involves, first, identifying a keyword representing a relevant aspect of press content (e.g. "uncertainty"). Next, the set of K most closely related terms are found based on the cosine distance between the respective vectors. Finally, the indicator is defined by the frequency of selected words.

[2] The vector representations used in applications are typically given by the sum of the two fitted word vectors: v_i and $\tilde{v}_j$. This practice is followed in the current implementation.

[3] More specifically, following Pennington et al. (2014), the weighting function equals $f(x) = (x/100)^{3/4}$ if $x < 100$, otherwise $f(x) = 1$.

More formally, given keyword $k \in W$ the set of K closest words is identified computing the cosine distance: $\frac{v_w \cdot v_j}{||v_w||||v_j||}$. This results in a set of words $K \subset W$. The index computed for a selected set of text C is given by:

$$I_C^k = \frac{\sum_{w \in K} c_w}{\sum_{w \in W} c_w}$$

where c_w indicates the number of times word w is observed in the selected set of text C. The set of selected text in the exercises below is given by economic press content over a specific time window.

Given the high level of attention placed on the concept of uncertainty (see for example Jurado et al. 2015, Baker et al. 2016, Rossi & Sekhposyan 2015), an indicator for "uncertainty" will be computed and evaluated. Beyond uncertainty, other indices approximating related but different manifestations in press content will be constructed and evaluated. More specifically, these manifestations are: pessimism, fear and anxiety. This choice reflect views and evidence reported in previous contributions and subjective judgments regarding relevance in macroeconomic contexts.

2.3 Data

The corpus used to train the vectors is given by text published in the Wall Street Journal between 1900 and 1989. The selected text corresponds to the content of a public webpage.[4] For each article published in the newspaper, this website provides access to the headline, the lead and a fraction of the body. To avoid concerns regarding forward looking biases, the training corpus is constructed using a time period that predates the period of the forecasting exercises that are presented in the next section. Table 1 shows information on the corpus used to train the word vector representations and the corpus used to compute the indicators.

Following common practice, numbers and punctuation marks are deleted from the text. Also, all text is converted to lower case and stop words are filtered.[5] After applying the minimum frequency filter, the dictionary of the training dataset is given by 28296 words. This is the number of 100 dimensional vectors computed in the GloVe model implementation.

The training corpus is relatively small compared to some databases used in the field of natural language processing.[6]. On the other hand, the training corpus is focused on economic discussions and can be conjectured to follow a relatively stable natural language. Additionally, the corpus used to compute the indices, the test corpus, shares the theme and style of the training corpus. As a result, there are reasons to remain optimistic regarding the present implementation's ability to learn

[4]The text was extracted from: http://pqasb.pqarchiver.com/djreprints/.

[5]The list of stop words can be found in the appendix.

[6]For example, in Pennington et al. (2014) word vector representations are trained using corpora with sizes that range from 1 billion tokens to 42 billion tokens.

word meaning in economic context.

Beyond text, a second set of data used in this study is given by real-time economic activity indicators for years 1966 through 2017. Four variables were selected: employment (Nonfarm Payroll Employment), industrial production (Industrial Production Index: Manufacturing), investment (Real Gross Private Domestic Investment: Nonresidential) and GDP (Real Gross Domestic Product)[7] The information is from the Federal Reserve Bank of Philadelphia's Real-Time Data Research Center.[8] The database built for the exercises below, preserves the real time nature of the original data. More specifically, for each sample quarter t, the values of the economic activity indicators, current values and lagged values, are given by information available at the time data corresponding to quarter t is first released. For example, real GDP data for the third quarter of year 1999 is given by information released in 28 October 1999. In particular, the figure for one-quarter-lagged real GDP (the level of activity in the second quarter) is 8778.6 billion dollars as expressed in the October 1999 release.

Table 1: Description of training corpus and test corpus

Corpus	**Number of articles**	**Number of tokens**
Training (1900-1989)	3,233,481	134,797,611
Test (1990-2017)	1,241,706	98,979,322

Table 2: Descriptive statistics - Quarterly growth rates

Activity Indicator	**Mean**	**St. Dev.**	**Min**	**Max**
Employment	0.0039	0.0051	-0.0202	0.0166
Industrial Production	0.0060	0.0186	-0.1098	0.0598
Investment	0.0085	0.0228	-0.1190	0.0531
GDP	0.0061	0.0074	-0.0274	0.0264

Note: Figures correspond to first releases. Sample period is 1966-2017.

[7]For National Income and Product Accounts, the information reported in the real-time dataset is the quarterly advance release.

[8]The data can be downloaded from: https://www.philadelphiafed.org/research-and-data/real-time-center/real-time-data/

3 Preliminary analysis of trained word vectors

Before proceeding to the forecasting exercise, preliminary evaluations of the information captured by word vectors are presented. First, some selected associations between word vectors will be used to demonstrate that trained vectors are able to learn word meaning in economic contexts. Second, an indicator that intends to capture manifestations of "uncertainty" will be evaluated in terms of its contemporaneous association with relevant macroeconomic events.

3.1 Vectors and meaning in economic context

The extent to which word vectors representations capture meaning is an empirical matter that depends on the informativeness of the training corpus and the efficiency of the learning model. One reason for concern is that, as previously indicated, the training corpus is relatively small compared to corpora typically used in the field of natural language processing. A variety of tasks can be used to evaluate trained vectors. The evaluations are based on associations between trained vectors. Three tasks are considered below: resolution of ambiguity in word meaning, entity identification through vector composition and identification of words indicative of tone or topic. The last task is the most relevant for the construction of indicators that reflect information in the press.

Ambiguous words are a common challenge in natural language processing applications. In particular, it is a problem for indicators based on predefined dictionaries. For example, Tetlock (2017), Garcia (2013) and Aromi (2017a 2017c) have shown that words in the negative category of Harvard IV dictionary can be used to anticipate financial and macroeconomic dynamics. Nevertheless, this category includes ambiguous words such as "capital", "tire" and "vice". In economic contexts, these words are not likely to transmit negative information. The presence of the word "capital" typically reflects discussions regarding financial issues, not discussions regarding the death penalty. Similarly, the word "tire" typically refers to the manufacturing of rings of rubber, not the need of rest or sleep. In the case of "vice" the most likely use is linked to the title of a corporate or bureaucratic position (e.g. vice chairman). The use of this word to refer to immoral or wicked behavior can be conjectured to be less likely.

Table 3 shows, for each of the mentioned ambiguous terms, the set of words with the closest vectors. The selected words suggest that word vectors are able to identify the most likely meaning. For example, in the case of "tire" the closest terms are related to the manufacturing of rings of rubber: goodyear, firestonoe and akron. A final example of an ambiguous word is given by "default". The set of closest terms (payment, debt and obligations) suggests that the identified meaning points to failures to fulfill an obligation not to preselected options. These examples are suggestive of efficiency gains associated to the use of unsupervised learning algorithms instead of pre-defined dictionaries. In addition, it is observed that the ambiguous, context dependent ways of natural language require the acquisition of knowledge through field-specific collections of text.

Word vector representations have been shown to learn relationships between words (Mikolov 2013). For example, in natural language processing tasks, computed vectors have been shown to learn associations such as: “king”-“male”+“female”= “queen”. This type of associations can be used to identify related entities in economic settings. A couple of examples are shown in table 3. The results suggest that vectors trained using economic press content learn to identify relationships between government entities and manufacturing corporations.

Finally, and more relevant for the current analysis, vectors are shown to identify groups of words related to the tone or the topic in a collection of texts. Suggesting valuable associations are learned, the word “uncertainty” is identified as close to other words that manifest negative, forward-looking, emotional and cognitive states. This evidence indicates that these associations between words can be used to construct indices that approximate uncertainty as communicated by economic press content. As an additional example, showing that word vector representations can be used to identify topics, it is observed that vectors also learn to identify words related to “debt”.

Table 3: Sample evidence on unsupervised learning of word meaning

Selected vector	**5 closest word vectors**
Ambiguous words:	
tire	goodyear, firestone, akron, tires, rubber
capital	par, authorized, outstanding, shares, common
vice	executive, president, elected, director, manager
default	payment, debt, obligations, overdue, waiver
Vector compositions:	
bundesbank-germany+us	fed, regulators, intervention, analysts, agency
gm-cars+planes	boeing, northrop, lockheed, aircraft, fighter
Tone/topic keywords:	
uncertainty	confusion, nervousness, apprehension, uneasiness, anxiety
debt	funding, longterm, financing, subordinated, restructure

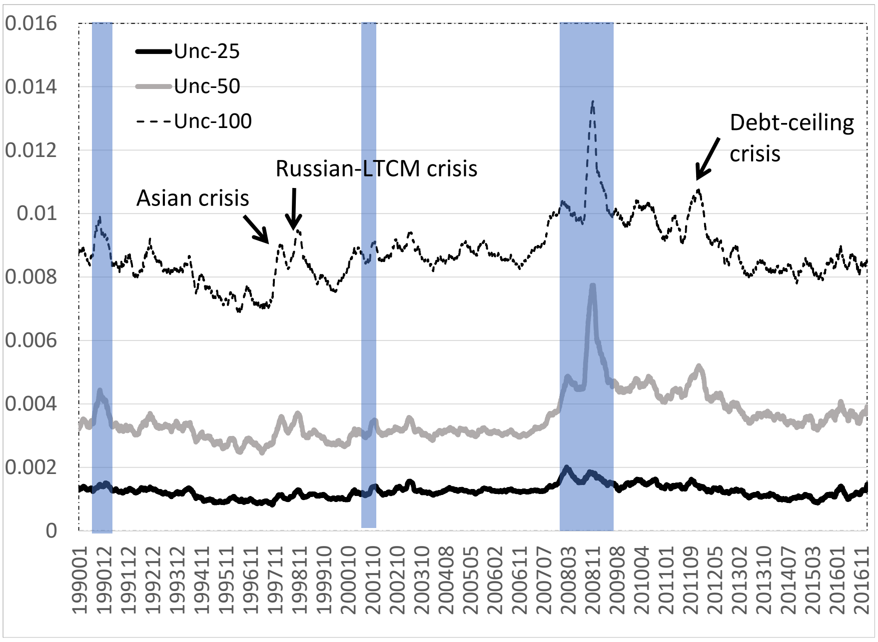
Note: The distance is computed using cosine similarity. Closest words exclude words with the same root.

3.2 An indicator of uncertainty

In this subsection, an indicator of uncertainty is presented. As previously indicated, the concept of uncertainty has received substantial attention in the analysis of business cycles (Jurado et al. 2015, Baker et al. 2016, Rossi & Sekhposyan 2015). Choosing ”uncertainty” as a keyword is a natural choice given this related literature.

An additional rationale is given by a study that exploits word vector representations to construct an uncertainty index that is shown to anticipate changes in expected stock market volatility (Aromi 2017b). In the first step of the computation, the vectors associated to the words "uncertainty", "uncertain" and "uncertainties" are added to compute a new vector w_U. The distance between this new vector and all words in the vocabulary W is computed. As previously described, the set of K words whose vectors are closest to w_U are selected. Finally, the index is given by the frequency of these K words. The indices are computed forming the second corpus that covers material published between January 1990 and February 2017. This second dataset contains approximately 98 million tokens.

Figure 1: Uncertainty Indices.



Notes: The figure shows the average of the indices for 90 day moving windows. Horizontal bars indicate recessions.

Figure 1 shows the values of three specifications of the uncertainty indicator. Each index is computed using a different number of uncertainty related words. Increments in the indices can be observed in the three recessions that took place during the sample period. This increment is particularly clear in the case of the recession linked to the 2008 Global Financial Crisis. Interestingly, in the case of the 2007-2009 recession, the indices show increments several months before the start of the recession in December 2007. Additionally, spikes in the indices are observed around three well-known crisis episodes: the Asian Crisis of 1997, the Russian Crisis of 1998 and the 2011 Debt-ceiling crisis. These associations suggest that meaningful information is captured by the index. Its ability to anticipate economic activity is evaluated in the next sections.

4 Macroeconomic forecasts

In this section, the information content of the indicators of uncertainty is assessed through business cycle prediction tasks. Beyond its intrinsic value, these exercises can serve as a general gauge of the relevance of these indicators in macroeconomic analysis. Positive results would suggest that academics and policymakers can benefit from the application of natural language processing techniques to large collections of unstructured data.

The first group of forecasting tasks involves in-sample evidence. In this case, the focus is placed on characterizing statistically and economically significant associations between indicators of information in the press and subsequent business cycle trajectories. A second group of exercises involve more demanding out-of-sample exercises. In this second case, gains in forecast accuracy are evaluated.

The forecasting model is given by an autoregressive specification complemented with an indicator of lagged press content. The growth rate of each economic activity indicator over the following h quarters is modeled as a function of lagged quarterly growth rates. The number of lags is selected minimizing the Bayesian Information Criterion.

More formally, let a_t be the value of an economic indicator in quarter t. The growth rate computed on quarter t is given by $\Delta a_t = log(a_t) - log(a_{t-1})$. Let $\Delta^h a_t$ represent the growth rate computed on quarter t for a window of size h, that is $\Delta^h a_t = log(a_t) - log(a_{t-h})$. The baseline autoregressive model satisfies:

$$\Delta^h a_{t+h} = \alpha + \sum_{s=0}^{p} \beta_s \Delta a_{t-s} + u_t \tag{1}$$

To evaluate the predictive ability of indicators based on press content, this baseline model is modified incorporating as predictor an indicator of press content. Let I_t represent the value of an indicator of press content corresponding to quarter t. Then, the forecasting model is given by the following equation:

$$\Delta^h a_{t+h} = \alpha + \sum_{s=0}^{p} \beta_s \Delta a_{t-s} + \beta_I I_t + u_t \tag{2}$$

The parameter of interest is β_I. Also, the relative metric of model fit, as indicated by increments in adjusted R^2's, will be analyzed to assess the in-sample forecasting performance of the indicator. The models are estimated for the period 1990-2017. In this way, word vector representations trained with press content published before year 1990 do not contain any forward looking information.

In the first of evaluations, the indicator of uncertainty is computed using the set of 100 words most closely related words. That is, the index is given by lagged quarterly frequency of words in this set. While the optimal specification of the indicator is unknown, this specification is used to advance with a first evaluation of information content. In out-of-sample exercises developed below, acknowledging uncertainty regarding optimal index specification, a Bayesian model averaging framework will be used to learn convenient specifications.

When in-sample forecasting exercises are implemented, by design, the estimated parameters reflect all information in the dataset, including future information. Beyond this feature, the exercise has been designed to secure that no other forward looking element is incorporated.[9] In particular, the forecasting exercise has been carefully designed taking into account the schedule of economic data release. For each instance of the forecasting exercise, any information used to produce the forecast must have been available at the time the forecast is generated. Each forecast exercise is simulated to occur on the day in which a new quarterly figure is released. All information released on that day is incorporated to the information set. The indicator of information in the press I_t summarizes lagged press content up 90 days before the release of the respective economic activity indicator. In other words, the forecasting exercise evaluates the predictive value of indicators of press content, right after having incorporated the news proceeding from the first release of quarterly economic activity data.[10]

In the case of industrial production, the release dates are as reported by the Board of Governors of the Federal Reserve System.[11] In the case of payrolls, new data becomes available in the first days that follow the month for which data is reported for the first time. In a cautious approach, in this case, the index is based on information published up to the last day of the most recent month for which information is available. In the case of National Income and Product Accounts, starting in 1996, release dates are available from the Bureau of Economic Analysis webpage.[12] For earlier dates, release dates are not available. The release date was assumed to be 28th day of the month of the release, that is, one day earlier than the average release day observed in the 1966-2017 period.

Table 4 shows evidence on the information content of the selected indicator. Four forecast horizons are considered: $h \in \{1, 2, 4, 8\}$. Adjusted R^2's show impor-

[9] In the next subsection, the implementation of out-of-sample exercises will eliminate forward-looking information in estimated parameters.

[10] In the case of data that is published on a monthly basis (payrolls and industrial production), the exercises are carried out four times a year. More specifically, the exercises are carried out in: January, April, July and October.

[11] The list can be found visiting https://www.federalreserve.gov/releases/g17/

[12] https://www.bea.gov/newsreleases/releasearchive.htm

tant gains in explanatory ability. This is specially noticeable in the case of longer forecasting horizons. For example, in the case of one-year ahead GDP forecasting models, the adjusted R^2 increases from 0.113 to 0.239 as the press content indicator is incorporated as a predictor.

In all cases, the estimated coefficient is negative. The estimated models point to a consistent economically significant association. A one standard deviation increment in the information metric anticipates a mean drop in economic activity growth that ranges from 0.18 to 0.40 standard deviations. For short forecast horizon models, statistically significant associations are observed. As forecast horizon grows, the number of statistically significant associations decreases. The indicator of press content is seen to be consistently informative in the case of GDP forecasts. In contrast, when industrial production forecasts are considered, the associated parameter is statistically significant only in the case of the shortest forecast horizon.

This evidence suggests that indices that exploit word vector representations have information regarding future levels of economic activity. In particular, this can be inferred from increments in adjusted R^2's as these indices are incorporated in the forecasting models. At the same, it is observed that the estimated associations are not always statistically significant. This is could be the result of inefficient specification of the index reflecting information in the press. For example, it is not known which is the appropriate weight that should be assigned to words characterized by different levels of associations with the concept of uncertainty. In the current specification, zeros and ones are assigned based on a arbitrary threshold. Also, it is reasonable to conjecture that more recent information should be allocated heavier weights. These issues are dealt with through Bayesian model averaging in implementations of out-of-sample forecast exercises shown below.

Table 4: Estimated forecast models

	h=1	h=2	h=4	h=8
Employment				
$\hat{\beta}_I$	-0.263**	-0.342**	-0.408**	-0.389
t-stat.	[2.24]	[2.25]	[2.05]	[1.29]
Adj. R^2	0.710	0.659	0.523	0.315
Baseline adj. R^2	0.666	0.583	0.412	0.214
Industrial Production				
$\hat{\beta}_I$	-0.280*	-0.341	-0.298	-0.180
t-stat.	[1.68]	[1.40]	[0.77]	[0.43]
Adj. R^2	0.414	0.290	0.155	0.029
Baseline adj. R^2	0.352	0.194	0.084	0.009
Investment				
$\hat{\beta}_I$	-0.342***	-0.396**	-0.354	-0.297
t-stat.	[2.96]	[2.00]	[1.23]	[0.94]
Adj. R^2	0.328	0.358	0.287	0.114
Baseline adj. R^2	0.230	0.225	0.180	0.041
GDP				
$\hat{\beta}_I$	-0.320***	-0.387***	-0.387***	-0.370**
t-stat	[3.85]	[3.01]	[2.44]	[2.06]
Adj. R^2	0.299	0.280	0.239	0.151
Baseline adj. R^2	0.217	0.156	0.113	0.035

Notes: significance levels: "*" 0.10, "**" 0.05 and "***" 0.01. Standard errors are estimated following Newey & West (1987, 1994). Parameter estimates are standardized; absolute t-statistics in brackets.

4.1 Out of sample exercises

The previous evidence regarding in-sample predictive ability is extended implementing a set of out-of-sample forecast exercises. Forecasts are generated using models fitted with real-time data. Four different forecast horizons are considered: $h \in \{1, 2, 4, 8\}$. The test sample starts in year 1990. Models are trained using expanding windows of historic data that begin in 1966 and end h quarters before the date in which the respective prediction exercise is implemented.

The analysis implements forecast combinations to acknowledge uncertainty regarding optimal specification of indicators summarizing information in the press. First, it must be noted that the choice of 100 words used in the previous forecasting exercises was arbitrary. It is reasonable to conjecture that a more efficient approach would allow for larger weights being allocated to more closely related words. In the exercise below, a data driven selection of weights is implemented. Also, optimal indicators would probably weight more heavily more recent information. Considering these concerns, indices associated to different number of words and alternative lagged windows are incorporated in the exercise. The forecasts associated to each of those indices are combined using Bayesian model averaging (BMA) techniques. In this way, forecasts combinations are used as a strategy to deal with risks associated to unknown models (Timmermann 2006).

Baseline forecasts correspond to those generated by the autoregressive model. As in the previous exercises, the number of lags is selected to minimize the Bayesian information criterion. Forecasts generated by the baseline model are compared to the combination of forecasts that are informed by different indicators of uncertainty in the press. To contemplate variation in the informativeness of more closely related words, indices with different number of related words are considered. The number of words used to construct the index equals 100, 50 or 25. Also, keeping in mind that more recent news flows might be more informative, two window sizes for lagged information are considered. In addition to the previously proposed 90 day window specification, indices based on 30 day windows are considered. These alternative specifications result in six indicators of information in the press. Let $\{I_t^i\}_{i=1}^6$ represent the indices computed the alternative specifications. Then, given a variable measuring economic activity a_t and a forecast horizon h, each indicator of uncertainty defines a forecasting model given by:

$$\Delta^h a_{t+h} = \alpha^i + \sum_{s=0}^{p} \beta_s^i \Delta^s a_t + \beta_I^i I_t^i + u_t^i \tag{3}$$

Where u_t^i is normally distributed with mean 0 and variance σ_i^2 . The BMA exercise incorporates an additional model given by the baseline autoregressive specification. Under BMA, forecast combinations involve computing weighted averages of the forecasts generated under each model. The weights are given by the posterior probability that the corresponding model is the true model. The current implementation follows the specification proposed in Faust et al. (2012).

More formally, let $\{M_i\}_{i \in N}$ be a collection of models. Also, let θ_i represent the parameters $\{\alpha^i, \beta_1^i, ..., \beta_p^i, \beta_I^i, \sigma_i^2\}$, and let D be the observed data. Then, the posterior probability is given by:

$$P(M_i|D) = \frac{P(D|M_i)P(M_i)}{\sum_{j \in N} P(D|M_j)P(M_j)} \tag{4}$$

where

$$P(D|M_i) = \int P(D|\theta_i, M_i)P(\theta_i|M_i)d\theta_i \tag{5}$$

is the marginal likelihood of the i-th model; $p(\theta_i|M_i)$ is the prior density of the parameter vector θ_i and $P(D|\theta_i, M_I)$ is the likelihood function. Following usual practice, it is originally assumed that prior probabilities are the same for all models. Also, it is assumed that the prior density of the parameters $\{\alpha^i, \beta_1^i, \beta_p^i, \beta_I^i, \sigma_i^2\}$ is uninformative and proportional to $1/\sigma_i$. The prior for parameter β_I^i follows Zellner (1986) g-prior specification: $\beta_I^i \sim N(0, \phi\sigma_i^2(I_i'I_i)^{-1})$. The parameter $\phi > 0$ controls the strength of the prior. Following previous literature, this parameter value is set to $\phi = 4$.[13] After computing the forecasts associated to each model, $\hat{a}_{t+h}^i$, and updating beliefs, the forecast combination is given by:

$$\Delta^h \hat{a}_{t+h} = \sum_{i=0}^{N} P(M_i|D)\Delta^h \hat{a}_{t+h}^i \tag{6}$$

Table 5 shows the information on the accuracy of forecasts that incorporate information from the press.[14] More specifically, the table shows the ratio between the root mean square error that results from the BMA approach and the root mean square error of the baseline model. The table also shows the p-values for the test of the null hypothesis that the ratio is equal to one. Given the presence of nested models, p-values are based on bootstrap methods as implemented in Faust et al. (2013). Gains in forecast accuracy are observed for most activity indicators and forecast horizons. For short forecast horizons (h=1 and h=2), gains in accuracy are statistically significant with p-values below 0.01. The case of GDP forecasts presents the most consistent gains associated to lagged information from the press. In contrast, in the case of industrial production, one-year-ahead and two-year-ahead forecasts are not seen to improve when compared to baseline forecasts.

The reported gains in forecast accuracy are consistent with the positive results observed in previously reported in-sample prediction exercises. Additionally, these results are indicative of gains associated to data driven selection of the specification of indicators of press content.

[13] See for example Fernandez et al. (2001) and Faust et al. (2012). The results are not sensible to changes in this parameter. These robustness exercises are available from the author upon request.

[14] The BMA implementation was estimated using R's package BMS.

Table 5: Out-of-sample predictive accuracy

	h=1	h=2	h=4	h=8
Employment	0.921	0.907	0.967	0.996
	[0.00]	[0.00]	[0.08]	[0.44]
Industrial Production	0.941	0.954	1.004	1.011
	[0.00]	[0.00]	[0.39]	[0.59]
Investment	0.943	0.939	0.911	0.945
	[0.00]	[0.00]	[0.00]	[0.07]
GDP	0.945	0.936	0.925	0.875
	[0.00]	[0.00]	[0.00]	[0.02]

NOTE: Relative RMSPEs; bootstrapped p-values for the test of the null hypothesis that the ratio of the RMSPEs is equal to one are reported in square brackets.

5 Indices measuring other subjective states

So far the analysis has focused on indicators that capture expressions associated to uncertainty. Forecasting exercises shown above indicate that these indices contain information regarding the future evolution of the business cycle. The choice of uncertainty related indices is a natural choice given the theoretical and empirical contributions that have focused on this concept on the context of business cycle studies (Jurado et al. 2015, Baker et al. 2016, Rossi & Sekhposyan 2015). Also, previous work showed that this type of uncertainty metric is able to anticipate stock market implied volatility (Aromi 2017b).

On the other hand, the evaluation of indicators associated to alternative aspects communicated in the press is a logical extension of the previous exercise. It is likely that the uncertainty proxy does not capture all relevant factors in an appropriate manner. Hence, increases in information content can result from the consideration of additional indicators. In particular, proxies of alternative subjective states could be considered. In the exercises shown below, three types of related but different indicators are incorporated. The choice of relevant additional subjective states is guided by previous literature and subjective judgment. While it would be desirable to have a more systematic approach for feature selection, this is beyond the scope of the current exercise and is left for future explorations.

First, considering the current prediction task associated to business cycles, manifestations in the press related to ”pessimism”, that is, expectations of negative scenarios, are conjectured to be relevant. Suggesting potential complementarities, pessimism can be viewed as a first moment feature of subjective states while uncertainty could be linked to second moment features. Second, pointing to a very prominent emotion, expressions related to "fear" are used as another potentially informative indicator. The perception of fear can be linked to the detection of dangers that are likely to have behavioral correlates. In this direction, it can be observed that the intensity of web searches related to "fear" was shown to have predictive value regarding investment decisions and stock market volatility (Da et al. 2014). Finally, in Nyman et al. (2018) it is suggested that expressions related to "anxiety" capture important information regarding subjective states and associated behaviors. Following this perspective, indices associated to this concept are also incorporated in the evaluated described below. [15]

Words most closely related to the selected keywords are shown in table 6. It is observed that most closely associated words are, for the most part, consistent with expected associations. Most of these words point to negative emotional and cognitive states. It also worth observing that some selected words do not contain a subjective element. For example, cause, situation and trouble are words that, in principle, do not refer to emotions or other subjective states. Finally, it can be observed that “optimism” is the word most closely associated to pessimism. This outcome suggests that the words “optimism” and “pessimism” are used in very similar contexts. After inspecting the other words associated to “pessimism”, it can be conjectured that these contexts are predominantly negative.

Beyond specific observations, overall, the associations suggest that the selected keywords allow for the construction of indices that extract relevant information from unstructured data. This exploratory evidence also shows that these concepts are closely linked. For example, words such as uncertainty, uneasiness and anxiety appear in table 6 in multiple occasions. As a result, indices associated to these concepts are expected to have an important common component. At the same time, differences in these indices might allow for data driven identification of the most informative indicator. Beyond rankings, complementarities between these closely related indicators could also be conjectured. These possibilities are formally evaluated through a new set of business cycle prediction exercises.

[15] In related explorations, indices associated to positive words such as optimism and excitement were evaluated. No predictive value was observed in this case. This can be linked to the "Pollyanna Hypothesis", according to which positive words are used more diversely and do not carry as much information as negative words. In other words, negative words are used in a more discriminatory manner (Bouchard & Osgood 1969, Garcia et al. 2012). Relatedly, Tetlock (2007) and Aromi(2017a) observe that, in contrast to negative words, positive words do not provide any information regarding future stock market returns.

Table 6: Words related to selected keywords

Selected keyword	**10 closest word vectors**
uncertainty	uncertainties, confusion, nervousness, uncertain, apprehension, uneasiness, anxiety, feeling, fears, situation
fear	fears, worry, feared, causing, danger, worried, cause, trouble, talk, worries,
pessimism	optimism, feeling, prevalent, anxiety, uneasiness, apprehension, gloom, discouragement, prevails, persists
anxiety	uneasiness, apprehension, nervousness, causing, confusion, uncertainty, pessimism, disappointment, excitement, feeling

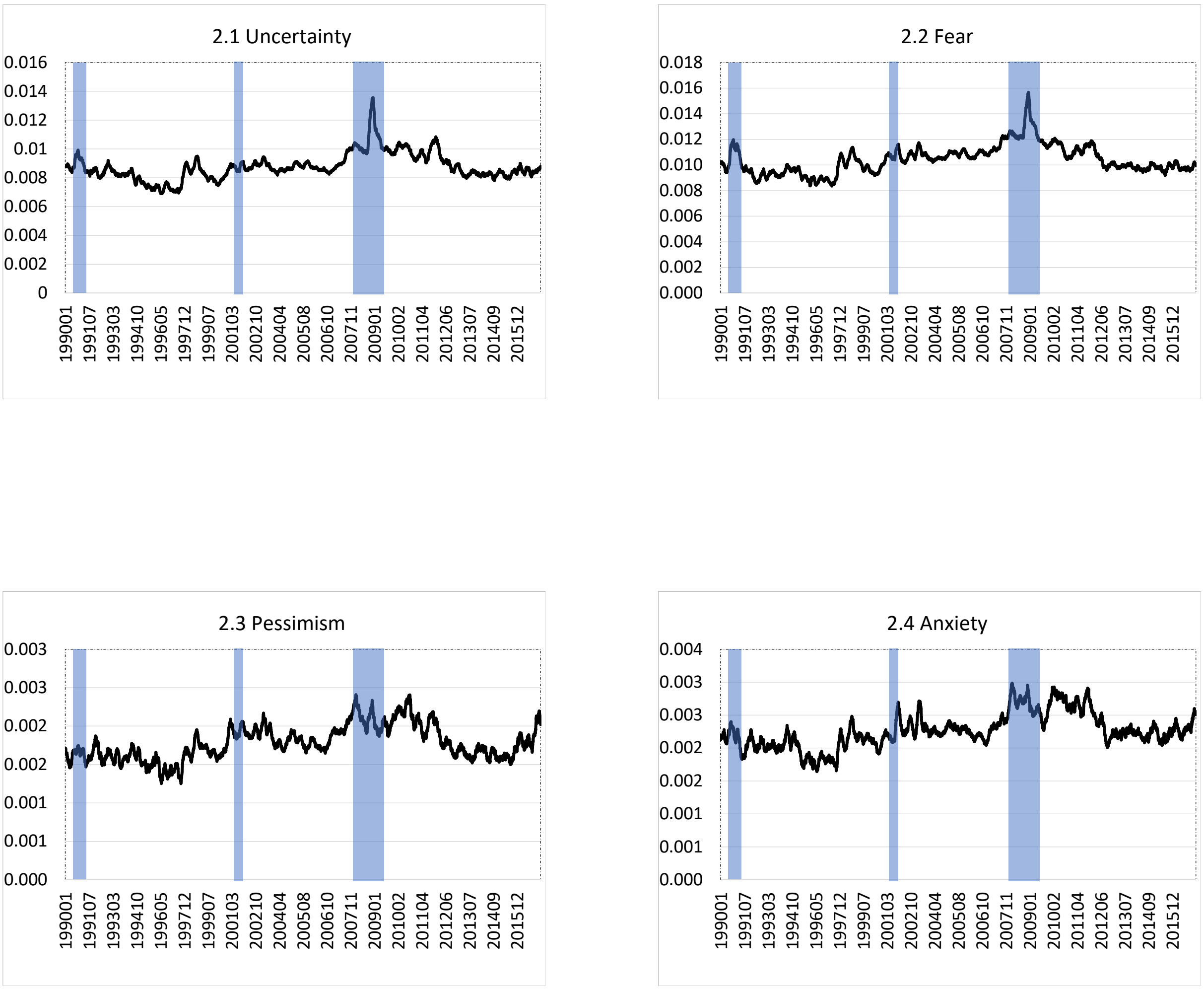
Notes: Distance computed using cosine distance.

Table 6: Subjective States Indices - Correlations

	Uncertainty	Fear	Pessimism	Anxiety
Uncertainty	1	0.9181	0.795	0.844
Fear	-	1	0.780	0.809
Pessimism	-	-	1	0.885
Anxiety	-	-	-	1

Notes: Indices computed using the 100 most closely related words and 90-day windows.

Figure 2: Subjective States Indices



Notes: Indices computed using the 100 most closely related words and 90-day windows.

The close association across the different indicators can verified inspecting correlations statistics shown in table 6. The indices were computed using 90-day windows. Correlation coefficients range from 0.78 to 0.92. The strongest correlation corresponds to the indicators approximating uncertainty and fear. The lowest, but still high, correlation corresponds to the indices measuring pessimism and fear. Additional preliminary evaluations result from inspecting figure 2. In consistence with the computed correlations, indices are seen to co-move. In all cases, the lowest figures are observed during the mid-nineties; increments are detected around recessions; the 2008-2009 crisis is associated to important and persistent increments. On the other hand, some differences that could prove meaningful can be distinguished. For example, the anxiety index presents distinctive spikes around 9/11 and the 2nd Gulf War. The increments around those events are not as prominent in the case of the other indices. In another example, the fear index and the uncertainty index present acute spikes around the most severe stage of the 2008-2009 crisis that are not observed in the case of the pessimism index and the anxiety index.

The first set of formal exercises involves the individual evaluation of the indices through dynamic regressions that incorporate one of the indicators as predictor. In these evaluations, as in the previous case of the index approximating uncertainty, the indices reflect the frequency of the set of 100 most closely related words. In the case of pessimism and fear, multiple keywords are used to construct the respective indices. In the case of pessimism, the adjective "pessimistic" is also used as a keyword. Also, in addition to the word "fear", the associated index is built using the words "fears" and "feared". In these cases, the vector representations of keywords are added and associated words are identified using this composite vector.[16]

Table 8 shows information for in-sample forecasting exercises. It is observed that, in all cases, the estimated coefficients are negative and the adjusted R^2's increase. Additionally, in most cases, the estimated coefficients are significantly different from zero. The indices approximating uncertainty and fear are seen to contain more information regarding future levels of economic activity. Suggesting that forecasts combinations might allow for more precise predictions, the best performing index is not always the same. While the uncertainty index seems to generate the most informative forecasts in the case of GDP and investment forecasts, the index associated to fear shows the strongest performance when employment and industrial activity are considered. The coefficients associated to indices approximating pessimism and anxiety are in most cases statistically significant. Nevertheless, the associations are clearly weaker as indicated by the p-values and the absolute value of standardized estimated coefficients.

As indicated in the previous section, this evidence might understate the information that can be inferred using indicators of press content based on word vector representations. The failure to identify statistically significant associations might reflect inefficiencies in the selected specifications for the indicators of information in the press. In the next subsection out of sample exercises are implemented allowing

[16]In the case of anxiety, the adjective "anxious" was not incorporated considering it is an ambiguous word that can be linked to positive content. In consistence with this choice, according to the cosine distance, the computed vector representations for "anxiety" and "anxious" are not similar.

for a data driven process that is designed to learn advantageous ways to summarize information.

Table 8: Estimated forecast models - h=1

	Uncertainty	**Fear**	**Pessimism**	**Anxiety**
Employment (baseline adj. R^2 =0.666)				
$\hat{\beta}_I$	-0.263**	-0.296***	-0.109**	-0.128**
t-stat.	[2.24]	[3.16]	[2.02]	[2.09]
Adj. R^2	0.710	0.723	0.672	0.677
Ind. Prod. (baseline adj. R^2 =0.352)				
$\hat{\beta}_I$	-0.280*	-0.314**	-0.139	-0.153
t-stat.	[1.68]	[2.18]	[1.65]	[1.57]
Adj. R^2	0.414	0.429	0.364	0.369
Investment (baseline adj. R^2 =0.230)				
$\hat{\beta}_I$	-0.342***	-0.332***	-0.158*	-0.168*
t-stat.	[2.96]	[3.31]	[1.73]	[1.81]
Adj. R^2	0.328	0.312	0.240	0.243
GDP (baseline adj. R^2 =0.217)				
$\hat{\beta}_I$	-0.320***	-0.292***	-0.139*	-0.175**
t-stat	[3.85]	[4.09]	[1.94]	[2.49]
Adj. R^2	0.299	0.286	0.229	0.239

Notes: significance levels: "*" 0.10, "**" 0.05 and "***" 0.01. Standard errors are estimated following Newey & West (1987, 1994). Parameter estimates are standardized; absolute t-statistics in brackets.

5.1 Out of sample forecasts

The methodology implemented in this subsection is the same as that implemented in the case of the indicators of uncertainty. To evaluate the information content of indices, forecast combinations based on BMA are evaluated against the baseline autoregressive model. Associated to each concept (e.g. fear), six indices are constructed. Also, as in the previously reported out-of-sample exercises, the alternative indices correspond to different number of related words (100, 50 and 25) and different periods (30-day and 90 -day windows). As a result, the BMA exercise involves 25 models. One model is the baseline model and 24 models correspond to autoregressive models that incorporate one of the 24 indices reflecting information in the press.

The results are shown in table 9. Suggesting complementarities between indicators that proxy different subjective states, accuracy is seen to increase in all forecasting tasks. Compared to forecasts that only exploit indices summarizing manifestations of uncertainty, improvements are specially noticeable in the case of payrolls and industrial production forecasts. A representative example is given by the case of one-year-ahead industrial production forecasts. In this case, the metric of relative forecast accuracy drops from 1.004, in the case of forecast based on uncertainty indicators, to 0.938, when additional indicators of subjective states are incorporated. In other words, there is an improvement from a scenario of no information gain to a statistically significant 6% improvement in the accuracy metric.

Table 10 shows information on posterior probabilities associated to the BMA exercise. In consistence with the evidence from in-sample forecasts, BMA is seen to assign a high posterior probability to models that incorporate indices related to fear and uncertainty. In the early stages of each forecasting exercise, optimal forecast combinations are not seen to be exclusively concentrated on a particular type of index. The only exemption is given by employment forecasts. In this case, the sum of posterior probabilities assigned to models that incorporate an indicator of fear is 0.85. Interestingly, by the end of the sample (2016-III), almost all of the posterior probability mass is placed on models that incorporate indices that approximate manifestations of fear.

Table 9: Out-of-sample predictive accuracy

	h=1	h=2	h=4	h=8
Employment	0.885	0.859	0.897	0.919
	[0.00]	[0.00]	[0.01]	[0.06]
Industrial Production	0.906	0.916	0.938	0.967
	[0.00]	[0.00]	[0.02]	[0.13]
Investment	0.928	0.931	0.886	0.934
	[0.00]	[0.00]	[0.00]	[0.06]
GDP	0.935	0.924	0.910	0.860
	[0.00]	[0.00]	[0.00]	[0.01]

NOTE: Relative RMSPEs; bootstrapped p-values for the test of the null hypothesis that the ratio of the RMSPEs is equal to one are reported in square brackets.

Table 10: Posterior probabilities - h=4.

Employment	1990-I	2003-II	2016-III
Uncertainty	0.076	0.055	0.006
Fear	0.854	0.654	0.977
Pessimism	0.044	0.248	0.010
Anxiety	0.018	0.035	0.001

Industrial Production	1990-I	2003-II	2016-III
Uncertainty	0.217	0.148	0.039
Fear	0.494	0.604	0.952
Pessimism	0.111	0.124	0.004
Anxiety	0.113	0.078	0.003

Investment	1990-I	2003-II	2016-III
Uncertainty	0.402	0.390	0.054
Fear	0.421	0.533	0.945
Pessimism	0.081	0.013	0.000
Anxiety	0.061	0.003	0.000

GDP	1990-I	2003-II	2016-III
Uncertainty	0.254	0.250	0.223
Fear	0.446	0.519	0.730
Pessimism	0.113	0.094	0.009
Anxiety	0.125	0.084	0.013

Notes: The table shows, for each economic activity indicator and each set of subjective indicators, the sum of the posterior probabilities assigned to models that incorporate indicators of the respective subjective state. This information is shown for the beginning, the middle and the final period of the out-of-sample forecast exercises. Forecast tasks

These exercises suggest that the combination of indicators capturing multiple subjective states is an advantageous strategy in business cycle prediction tasks. In particular, indicators of uncertainty and fear seem to provide the most valuable information. The selection of keywords was informed by subjective judgment and related contributions. A systematic procedure for keyword selection is a desirable feature that is beyond the scope of the current study.

6 Comparison with alternative methodologies

This work focuses on the ability to automatically learn meaning in economic contexts. It is conjectured that natural language processing tools might allow for efficient extraction of information in unstructured data. On the other hand, indices based on a small set of keywords or predefined dictionaries contain a significant amount of information. These traditional methods reflect expert judgment regarding convenient categorizations or keywords. The relative performance of these alternative methodologies is unknown, and needs to be evaluated empirically.

The first indication of informational gains associated to the use of word vector representations will be based on a simple benchmark. More specifically, a simple indicator of uncertainty based on the frequency of the words “uncertainty”, “uncertain” and “uncertainties” (Unc-3) is proposed. The information content of this indicator is compared to the information provided by the index based on the set of 100 most closely related words as indicated by word vector representations (Unc-WVR).

In addition, the information content of four traditional methodologies proposed in the literature are compared to information captured by uncertainty indices that exploit word vector representations. Following Tetlock (2007) multiple contributions have exploited the list of words categorized as negative in the Harvard IV dictionary.[17] Suggesting that the previous dictionary needs to be adapted to the context, Loughran & MacDonald (2011) propose a list of words that transmit a negative tone in financial contexts.[18]. In a simple, yet potentially valuable approach, the monthly publication "The Economist" has proposed the R-Index, a metric of the frequency with which the word "recession" is found in the economic press.[19]. Finally, an influential metric based on press content is proposed in Baker et al. (2016). The metric is known as the Economic Policy Uncertainty (EPU) index. This index computes the fraction of news articles that refer to economic policy and uncertainty.[20] These articles are identified using a small set of words.[21]

The performance of the uncertainty metric based on word vector representation is compared to the performance of indices associated to the previously described alternative methods. In the case of the first three alternative methods, the indices are computed using the test corpus of WSJ content used in this contribution. In the case of the Economic Policy Uncertainty index, the index was downloaded from the website created by the authors. The EPU is constructed searching text content for a large collection of publications.

Table 11 shows results for in-sample forecast exercises for the five alternative indices and for the index approximating uncertainty using the set of 100 words most closely related uncertainty. Estimated parameters, p-values and adjusted R^2s indicate that the index based on word vector representations is the most informative indicator. This conclusion is valid independently of the economic activity metric under consideration.

Among alternative indices, there is no clearly superior methodology. The index based on the Harvard-IV dictionary dominates in the case of GDP forecast. The EPU seems to be the most valuable indicator in the case of employment and industrial production forecasts. Nevertheless, in the case of investment forecasts, the R-word index shows the strongest performance among alternative indicators.

[17]The list can be downloaded from http://www.wjh.harvard.edu/ inquirer/homecat.htm.

[18]The list can be downloaded from www3.nd.edu/ mcdonald/

[19]The index was first proposed in 2001 (http://www.economist.com/node/566293)

[20]More precisely, this metric of press content is one out of three elements used to compute the EPU index.

[21]Details on methodology and data can be found visiting http://www.policyuncertainty.com/index.html

Table 11: Estimated forecast models - h= 1

	Unc-WVR	Unc-3	Harvard-IV	L&M(2011)	R-word	EPU
Employment						
$\hat{\beta}_I$	-0.263**	-0.098	-0.095	-0.132*	-0.114**	-0.182*
t-stat.	[2.24]	[1.54]	[1.34]	[1.77]	[2.26]	[1.97]
Adj. R^2	0.710	0.671	0.670	0.676	0.668	0.693
Ind. Prod.						
$\hat{\beta}_I$	-0.280*	-0.157*	-0.115	0.108*	0.076	-0.174*
t-stat.	[1.68]	[1.86]	[1.10]	[1.72]	[1.55]	[1.72]
Adj. R^2	0.414	0.369	0.357	0.356	0.350	0.373
Investment						
$\hat{\beta}_I$	-0.342***	-0.212	-0.243	-0.202	-0.326**	-0.237**
t-stat.	[2.96]	[1.59]	[1.59]	[1.42]	[2.57]	[2.11]
Adj. R^2	0.328	0.257	0.265	0.252	0.293	0.271
GDP						
$\hat{\beta}_I$	-0.320***	-0.228***	-0.302**	-0.263**	-0.263*	0.188**
t-stat	[3.85]	[2.95]	[2.24]	[2.51]	[1.96]	[2.18]
Adj. R^2	0.299	0.255	0.278	0.264	0.255	0.241

Notes: significance levels: "*" 0.10, "**" 0.05 and "***" 0.01. Standard errors are estimated following Newey & West (1987, 1994). Parameter estimates are standardized; absolute t-statistics in brackets.

The absence of a clear ranking between alternative methodologies implies that forecasts combinations could be used to find an efficient way to incorporate information provided by the respective indicators. Out of sample forecasts are generated through Bayesian model averaging. In the first exercise, the predictive ability associated to the four alternative methodologies proposed in the literature is evaluated jointly through BMA. In the second exercise, alternative methodologies are considered jointly with the uncertainty indices based on word vector representations.

In the first set of exercises, nine models are considered. One is associated to the baseline autoregressive specification. For each alternative methodology, indices are built using 30-day and 90-day lagged windows. In this way, eight additional models are added to the baseline specification. The results are shown in panel A of table 12. Compared to the baseline specification, gains in forecast accuracy are observed in the case of the shortest forecast horizon. For longer forecast horizons, no significant gain in accuracy is observed. Additionally, these alternative specifications are not seen to match the forecasting performance observed in the case of forecasts that exploit word vector representations. For example, in the case of one-year-ahead GDP forecasts, forecasts based on the alternative indices generate a metric of accuracy that is 3.2% worse that the baseline model. In contrast, uncertainty indices informed by word vector representations generate forecasts that are significantly better that baseline forecasts. The metric of accuracy improves by 7.5%.

While alternative indices do not provide a strong forecasting performance, they can be conjectured to capture information that could be advantageously used through forecast combination. In other words, the indices proposed in this work might be complementary with alternative text summarizing techniques. To evaluate this hypothesis, the indices that exploit word vector representations to proxy manifestations of uncertainty are incorporated to the previous BMA exercise. As in the previous sections, alternative specifications associated to number of words and time windows result in six indices. As a result, these forecast combination exercises involve identifying the weights assigned to 15 models.

The results, shown in panel B of table 12, suggest that there are no gains associated to incorporating the alternative indicators. Forecast accuracy as indicated by relative RMSPEs is worse than that observed when the uncertainty metrics based on word vector representations are the only indicator of information in the press (see table 5).

Table 12: Out-of-sample predictive accuracy - Alternative indicators

[A] Alternative indicators

	h=1	h=2	h=4	h=8
Employment	0.950	0.952	0.991	0.975
	[0.00]	[0.01]	[0.20]	[0.18]
Industrial Production	0.986	1.015	1.044	1.009
	[0.01]	[0.68]	[0.47]	[0.54]
Investment	0.992	1.003	0.989	1.019
	[0.31]	[0.88]	[0.42]	[0.54]
GDP	0.980	0.987	1.032	0.954
	[0.01]	[0.14]	[0.19]	[0.11]

[B] Combination WVR + Alternative indicators

	h=1	h=2	h=4	h=8
Employment	0.924	0.913	0.982	0.990
	[0.00]	[0.00]	[0.15]	[0.30]
Industrial Production	0.953	0.982	1.037	1.009
	[0.00]	[0.03]	[0.58]	[0.54]
Investment	0.946	0.944	0.914	0.953
	[0.00]	[0.00]	[0.00]	[0.08]
GDP	0.944	0.951	0.960	0.902
	[0.00]	[0.00]	[0.03]	[0.02]

NOTE: Relative RMSPEs; bootstrapped p-values for the test of the null hypothesis that the ratio of the RMSPEs is equal to one are reported in square brackets.

7 Conclusions

This study proposes a method for the quantification of unstructured information in the press. It it is based on word vector representations, a tool developed in the field of natural language processing. It is shown that trained representations learn meaningful relationships between words in economic contexts. These associations are exploited to build indicators of uncertainty and other subjective states in press content.

Using real-time data on economic activity, the indices are shown to capture valuable information. Indicators of uncertainty anticipate business cycles dynamics under in-sample and out-of-sample prediction exercises. Bayesian model averaging implementations show that there are benefits associated to combining information from indices linked to different subjective states. Also, their information content compares favorably to that resulting from alternative text processing techniques considered in the literature. In this way, this work shows how novel machine learning tools can generate interpretable and informative indicators that can be used in macroeconomic analysis.

The are several directions in which the current work can be extended. A natural path is associated to implementations based on larger training and testing corpora. While larger collections of text do not necessarily mean more information, a careful selection of additions to the corpus could result in more precise indicators. As previously indicated, automated methods for the selection of relevant features in unstructured data can also be explored. In the field of natural language processing, word vector representations are used as inputs in neural network applications (Kim 2014). Hence, while the property of straightforward interpretation would be lost, another possible extension involves exploring nonlinear forecasting models.

References:

Alexopoulos, M., & Cohen, J. (2015). The power of print: Uncertainty shocks, markets, and the economy. International Review of Economics & Finance, 40, 8-28.

Ardia, D., Bluteau, K., & Boudt, K. (2017). Aggregating the Panel of Daily Textual Sentiment for Sparse Forecasting of Economic Growth.

Aromí, J. D. (2017a). Conventional views and asset prices: What to expect after times of extreme opinions?. Journal of Applied Economics, 20(1), 49-73.

Aromí, J. D. (2017b). Measuring uncertainty through word vector representations. Económica, 63(1), 135-156.

Aromí, J. D., (2017c), GDP growth forecasts and information flows: is there evidence of overreaction?, International Finance,-(-), -.

Baker, S. R., Bloom, N., & Davis, S. J. (2016). Measuring economic policy uncertainty. The Quarterly Journal of Economics, 131(4), 1593-1636.

Bernanke, B. S., Gertler, M., & Gilchrist, S. (1999). The financial accelerator in a quantitative business cycle framework. Handbook of macroeconomics, 1, 1341-1393.

Boucher, J., & Osgood, C. E. (1969). The pollyanna hypothesis. Journal of Verbal Learning and Verbal Behavior, 8(1), 1-8.

Chen, H., De, P., Hu, Y. J., & Hwang, B. H. (2014). Wisdom of crowds: The value of stock opinions transmitted through social media. The Review of Financial Studies, 27(5), 1367-1403.

Collobert, R., Weston, J., Bottou, L., Karlen, M., Kavukcuoglu, K., & Kuksa, P. (2011). Natural language processing (almost) from scratch. Journal of Machine Learning Research, 12(Aug), 2493-2537.

Corsi, F. (2009). A simple approximate long-memory model of realized volatility. Journal of Financial Econometrics, 7(2), 174-196.

Diebold, F. (2015). Comparing Predictive Accuracy, Twenty Years Later: A Personal Perspective on the Use and Abuse of Diebold-Mariano Test, Journal of Business and Economic Statistics, Vol 33, 1.

Duchi, J., Hazan, E., & Singer, Y. (2011). Adaptive subgradient methods for online learning and stochastic optimization. Journal of Machine Learning Research, 12(Jul), 2121-2159.

Faust, J., Gilchrist, S., Wright, J. H., & Zakrajšsek, E. (2013). Credit spreads as predictors of real-time economic activity: a Bayesian model-averaging approach. Review of Economics and Statistics, 95(5), 1501-1519.

Fernandez, C., Ley, E., & Steel, M. F. (2001). Benchmark priors for Bayesian model averaging. Journal of Econometrics, 100(2), 381-427.

Garcia, D. (2013). Sentiment during recessions. The Journal of Finance, 68(3), 1267-1300.

Garcia, D., Garas, A., & Schweitzer, F. (2012). Positive words carry less information than negative words. EPJ Data Science, 1(1), 3.

Gentzkow, M., Kelly, B. T., & Taddy, M. (2017). Text as data (No. w23276). National Bureau of Economic Research.

Gilchrist, S., & Zakrajšek, E. (2012). Credit spreads and business cycle fluctuations. American Economic Review, 102(4), 1692-1720.

Hansen, S., McMahon, M., & Prat, A. (2017). Transparency and deliberation within the FOMC: a computational linguistics approach. The Quarterly Journal of Economics, 133(2), 801-870.

Jermann, U., & Quadrini, V. (2012). Macroeconomic effects of financial shocks. American Economic Review, 102(1), 238-71.

Jurado, K., Ludvigson, S. C., & Ng, S. (2015). Measuring uncertainty. American Economic Review, 105(3), 1177-1216.

Kim, Y. (2014). Convolutional neural networks for sentence classification. arXiv preprint arXiv:1408.5882.

Loughran, T., & McDonald, B. (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. The Journal of Finance, 66(1), 35-65.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In Advances in neural information processing systems (pp. 3111-3119).

Newey WK & West KD (1987), A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. Econometrica, 55,pp. 703-

708.

Newey WK & West KD (1994), Automatic Lag Selection in Covariance Matrix Estimation. Review of Economic Studies, 61,pp. 631-653.

Nyman, R. & Kapadia, S., Tuckett, D., Gregory, D., Ormerod, P. and Smith, R.(2018). News and Narratives in Financial Systems: Exploiting Big Data for Systemic Risk Assessment. Bank of England Working Paper No. 704.

Pennington, J., Socher, R., & Manning, C. D. (2014). Glove: Global vectors for word representation. In EMNLP (Vol. 14, pp 1532-1543).

Rossi, B., & Sekhposyan, T. (2015). Macroeconomic uncertainty indices based on nowcast and forecast error distributions. American Economic Review, 105(5), 650-55.

Schularick, M., & Taylor, A. M. (2012). Credit booms gone bust: Monetary policy, leverage cycles, and financial crises, 1870-2008. American Economic Review, 102(2), 1029-61.

Tetlock, P. C. (2007). Giving content to investor sentiment: The role of media in the stock market. The Journal of finance, 62(3), 1139-1168.

Timmermann, A. (2006). Forecast combinations. Handbook of economic forecasting, 1, 135-196.

Zellner, A. (1986). On assessing prior distributions and Bayesian regression analysis with g-prior distributions. Bayesian inference and decision techniques, ed. by P.K. Goel & A. Zellner, Amsterdam, The Netherlands: North-Holland, 233-243.

Appendix A: List of stop words

a, an, and, at, are, been, by, between, by, can, could, for, has, have, is, in, of, on, or, since, that, the, these, this, those, to, was, were, will, with, without.