# Predicting Economic Growth from Business News

**Siu-Hang Chan**
chansiuhang@hotmail.com

**Mustafa Panbiharwala**
mustafa.shabbir10@gmail.com

## Abstract

This paper tests the hypothesis that signals in the form of sentiment from newspaper text can be used to materially improve the forecasts of macroeconomic variables including infaltion, GDP and unemployment rate. Our news corpus is drawn from GDELT database which contains news from all popular US newspapers. A financial sentiment model trained on financial news headlines to predict sentiments of daily news articles from 2013 to 2021. A timeseries of these sentiments is constructed which is then used in an auto-regressive model to predict the inflation and GDP at various time lags. This methodology should allow researchers in policy-making institutions to produce more accurate assessment of the macroeconomy of United States.

## 1 Introduction

Macroeconomics forecasting, e.g. economic growth, inflation, and unemployment rate, is one of the major tasks in macroeconomics, and it matters a lot for policymaker market participants. Traditionally, various economics factors are used as independent variables / features, e.g. survey data, for forecasting, most probably with an autoregressive model.

The recent advance in natural language technology and text analytics techniques have made the use of text information more practical on a large scale. As a result, researchers and market practitioners are looking into whether large text corpus, e.g. those from newspapers, can be helpful in macroeconomics forecasting or not.

Here, we will test if the hypothesis: text corpus from news articles are helpful in macroeconomics forecasting, is true or not. We expect the sentiment scores, which are estimated from text corpus from new articles using our own-trained sentiment classification model, will boost the prediction accuracy

of the variables of concern.

In the following, we aim to build a pipeline that extracts information from news articles to help make assessments about the USA's economics (namely the GDP, CPI and unemployment rate). There will be two main steps there: 1) extract sentiment scores from unstructured news articles 2) using the sentiment scores to forecast a macroconomic indicator such as Inflation or growth in the USA. For step (1), we plan to build a sentiment classification model using labelled data (text passage with labelled sentiment), and then use it on a larger unlabelled news dataset to construct a time series of sentiment scores. For step (2), the constructed time series of sentiment scores will be applied to forecasting of macroeconomic indicators and various models will be compared.

## 2 Related Work

There are various literatures about using text information for macroeconomics forecasting. While they have the same goal of assessing how sentiments extracted from the unstructured business news can be used to forecast macroeconomic indicators of a country's health, each of them have a slightly different approach in how they collect news data, how they create a sentiment analysis model and finally how they use the model to forecast the macroeconomic indicator.

There are different ways to construct the sentiment score. All papers use a lexicon based model for the task of sentiment analysis on news data. For example, in (Shapiro and Wilson, 2020) (Trajanov, 2020) (Ardia and Boudt, 2019) (Larsen and Zhulanova, 2021) use predefined lexicons and human judgement to create daily sentiment scores across all business news articles and (Tilly and Livan, 2021) used a Bi-LSTM model trained on a knowledge graph to extract sentiment scores. A

completely different approach of skipping the sentiment score is also investigated in (Eleni Kalamara and Kapadia, 2020).

The data source varies quite a lot too. Some of them are based on US newspapers (Shapiro and Wilson, 2020) (Trajanov, 2020), some on UK ones (Eleni Kalamara and Kapadia, 2020), while (Tilly and Livan, 2021) are based on a global database (GDELT). This shows that text information can be used to improve macroeconomics forecasts regardless of the location of the economy. Majority of the papers use a proprietary data source to extract news articles. (Larsen and Zhulanova, 2021) uses Dow Jones Newswires Archives to extract news articles from 1990 to 2016 and (Ardia and Boudt, 2019) uses LexsNexis' proprietary smart indexing technology to extract news articles as well as topics associated with those articles.

While considering the macroeconomics forecasting part, the most common benchmark is some form of autoregressive linear time series model, like the one used in (Shapiro and Wilson, 2020) and (Tilly and Livan, 2021). Some other machine learning approaches like Lasso, Ridge, Neural Networks, SVM, and Random Forest are also discussed in (Eleni Kalamara and Kapadia, 2020). A dynamic factor model approach is tested in (Ardia and Boudt, 2019) and (Larsen and Zhulanova, 2021) which uses dimensionality reduction technique on the sentiment scores generated from the text data to forecast the macroeconomic indicators. Also, although both (Ardia and Boudt, 2019) and (Larsen and Zhulanova, 2021) use to create sentiment scores for certain news topics, the way they extract these topics are very different. (Larsen and Zhulanova, 2021) use LDA on unstructured news data to extract and create their own news topics whereas (Ardia and Boudt, 2019) uses built in topics from their data source and skips this step of topic modeling. But both of the papers stress on the fact that sentiment scores based on news topics do have a high predictive power to forecast macroeconomic indicators.

## 3 Data

In this project, for the sentiment analysis task we use 2 publicly available datasets which have been labeled by financial experts and thus provide a good source of sentiment labels across news articles to train our machine learning models. We frame the sentiment classification task as a ternary classification problem (positive, negative and neutral). For sentiment score calculation and macroeconomics forecasting, another two dataset are used.

### 3.1 Financial Phrase Bank

The financial phrase bank dataset (Pekka Malo, 2014) consists of 4846 english sentences selected from financial news dataset and labeled by 16 financial experts. This data is constructed by randomly selecting sentences from financial news articles found on the LexisNexis database. The sentences are then labeled by experts by considering how that sentence would influence a company's stock price. All sentences are annotated with 3 labels: Positive, Negative and Neutral. The distribution of sentiment labels is presented in Table 1.

### 3.2 SemEval 2017 TASK 5

The second dataset used in this paper is provided by the SemEval-2017 task "Fine-Grained Sentiment Analysis on Financial Microblogs and News" (K. Cortis and Davis, 2017). This dataset consists of 1633 English sentences taken from financial news headlines which are sourced from different publicly available datasets such as Yahoo Finance. Each instance is labeled by financial experts in the range -1 to 1 where -1 is negative sentiment and +1 is positive sentiment. These sentiment scores were converted to positive, negative and neutral labels where a score greater than 0 is a positive sentiment, score less than 0 is assigned a negative sentiment and a score equal to 0 is assigned a neutral sentiment. The distribution of sentiment labels is presented in Table 1.

The dataset used for creating the financial sentiment model is a combination of both of the datasets. For creating shallow machine learning models the data was divided into 2 parts: training and test. The training data consisted of 80% population from the full dataset and test data had the remaining 20% sentences. For creating deep learning models, we further divided the training into two parts: training and validation, where 80% of the training data was assigned to training and the rest 20% went to validation set.

### 3.3 Global Database of Events, Language, and Tone (GDELT)

GDELT version 1.0, from 1-Apr-2013 to 30-Sept-2021, totally 3101 days of news are used for calculating the sentiment score. In this database, news on the web from all around the world is collected,

| Dataset | Neutral | Positive | Negative |
|---|---|---|---|
| Financial Phrases | 2879 | 1363 | 604 |
| SemEval 2017-Task5 | 529 | 653 | 451 |
| Total | 3408 | 2016 | 1055 |

Table 1: Sentiment Label statistics.

but for the purpose of predicting macroeconomics variables of the United States, only US news is filtered for processing.

In the GDELT database, the URL of the website of each news item is stored in a CSV table, which contains the news of the same date. In order to read the text in the news, we first unzip the zip file that contains the CSV file, and then load the URL one by one to get the text in the news from the website. Unfortunately, this process is quite slow, and thus we can only sample some of the news each day due to time constraints of the project.

### 3.4 Macroeconomics Data from FRED

Three macroeconomics variables are of concern: GDP, CPI and Unemployment Rate, which are important ones that reflect the status of economic growth, price level, and the job market. The corresponding data from Federal Reserve Economic Data (FRED) are downloaded from Federal Reserve Bank of St. Louis: Real Gross Domestic Product (from Q1-1947, quarterly), Consumer Price Index for All Urban Consumers: All Items in U.S. City Average (from Jan-1913, monthly) and Unemployment Rate (from Jan-1948, monthly). In order to make the performance comparable between the model with sentiment scores calculated using GDELT, only the data from Apr-2013 / Q2-2013 to Sep-2021 / Q3-2021 is used.

### 4 Data Processing

Financial headlines and financial news text are contain errors and be inconsistent. To prepare the data we perform initial pre-processing that includes lower-casing the text, punctuation removal, removing numbers from text, clearing white spaces, stemming and finally tokenization. These pre-processing steps are crucial for building good machine learning models when working with text data. These steps were used in shallow machine learning when building Logistic Regression and Support Vector Machine classifier.

When pre-processing data for LSTM model, we BertTokenizer from HugginFace library to tokenize text which tokenizes and lower cases the text in a way that is consistent with the pre-trained BERT model.

### 5 Sentiment Classification

The first step in forecasting macroeconomic indicators such as inflation or GDP growth is creating a sentiment analysis model that can predict sentiment of financial news articles. First we create machine learning models on the text data to learn sentiment weights for words and entire phrases in order to measure the sentiment of an entire expression. We train the machine learning model on training data and use k-fold cross validation to evaluate performance of each possible set of hyper-parameters. Based on the cross validation results, the optimal model is created using the best hyper-parameters and that model is evaluated against a test set for final results which we report here. The first model we consider is a "bag-of-words" model where the features are counts of each word (unigrams) or count of word phrases (bi/tri-grams) which result in very high dimensional and sparse feature vectors. To reduce dimensionality to a certain extent we use a feature selector which selects the top features based on a chi-squared statistic which measures the degree of association between the response and the count features. The bag-of-words models are simple and easy to implement and thereby are considered as our baseline models.

The next set of models we consider are deep learning based contextual models which use word embeddings to encode words into vectors which bring additional knowledge about words while training. We use a bidirectional LSTM as our neural network architecture because of its ability to learn long term dependencies in text. The input for the LSTM are word embeddings which are extracted from the final layer of the BERT model, FinBERT, which is a pre-trained transformer based model trained on financial text data. FinBERT is widely used in finance sentiment analysis. Table 2 shows the performance of different classification algorithms on test data.

Performance of the model is assessed using precision, recall for each of the three classes along with f1-macro score. F1-macro is considered as main classification metric for choosing the best

3

| Classifier | F1-Macro |
|---|---|
| LR Bag-of-words | 0.64 |
| SVM Bag-of-words | 0.63 |
| LSTM with FinBERT | 0.75 |

Table 2: Classifier Performance.

model because it treats all classes equally strikes a balance precision and recall. All three classes (Positive, Negative and Neutral Sentiment) are equally import to our problem and f1-macro is low for models that only perform well on common classes while performing poorly on the rare classes. Thus, we select LSTM model with FinBERT embeddings because it gives us the highest f1-macro of 0.75 on the test data.

## 6 Macroeconomics Forecasting

In order to investigate if the text information from newspapers are useful in macroeconomics forecasting, we leverage the GDELT, and estimate the sentiment from the US news. Various sentiment scores are then constructed based on this, and they will be tested if they are statistically significant in macroeconomics variable construction.

### 6.1 Models

Firstly, for each news, the sentiment is estimated by the sentiment classification models (LR and Bi-LSTM model respectively), and they are classified to positive, neutral and negative. For each month, the total number of positive, neutral, and negative news are counted, and the proportional of that are stored. We have computed three sentiment scores: the raw proportion of positive, neutral and negative news, the difference between positive and negative news, and the log of the ratio between the proportion of positive and negative news.

Secondly, the scores are added into the benchmark model one by one to see if they are statistically significant for prediction of the macroeconomics model. Due to autocorrelation in the three macroeconomics variables, an autoregressive similar to that adopted in (Eleni Kalamara and Kapadia, 2020) is used. Based on the ACF and PACF graphs, an AR(1) model is used as the benchmark to be compared with.

### 6.2 Experiment

Totally three baseline models are fitted, one for each of the GDP, CPI and unemployment. In addi-

tion, for each of the 3 macroeconomics variables, and for each of the 2 sentiment classification models (LR and LSTM), 3 linear models are trained for each method of the sentiment scores: the raw sentiment proportion, the difference, and the log ratio respectively, are trained, which make up to a total of 18 models.

Benchmark Model:

$$y_t = \alpha + \beta * y_{t-1} + \epsilon$$

with Sentiment Scores:

$$y_t = \alpha + \beta_1 * y_{t-1} + \beta_2 * sentiment_t + \epsilon$$

where yt is the macroeconomics variable of concern at time t.

In order to conclude if the sentiment scores are useful or not, we check if the coefficients are statistically significant or not, and check if the adjusted R-square are improved or not.

### 6.3 Results and Analysis

From the table below, we can see that the results are positive for most macroeconomics variables: namely the proportion of negative news are significant in prediction of unemployment rate, the proportion of positive news are strongly significant in prediction of CPI, and the proportion of positive news are weakly significant in prediction of GDP. In all cases, the adjusted R-square has also been improved.

Comparing the models, it seems that the more simple LR model is performing slightly better than the more complicated LSTM one in macroeconomics forecasting. This can be the effect of better generalization to other dataset of a more simple model, or simply because of the way of how the sentiment scores are constructed.

While among the various ways of sentiment scoring, it seems that the raw proportion is the best among the three, possible due to the fact that it preserves the most information among them.

## 7 Conclusion, Limitation and Further Work

We have successfully built two sentiment classification models that predict the sentiment of news articles well. In addition, we have also found that the sentiment score calculated can be useful in predicting the unemployment rate and CPI, and also potentially useful for predicting the GDP.

As mentioned before, the sentiment scoring part is done on a sampling of the subset of the database due to time constraint, and the result can be more
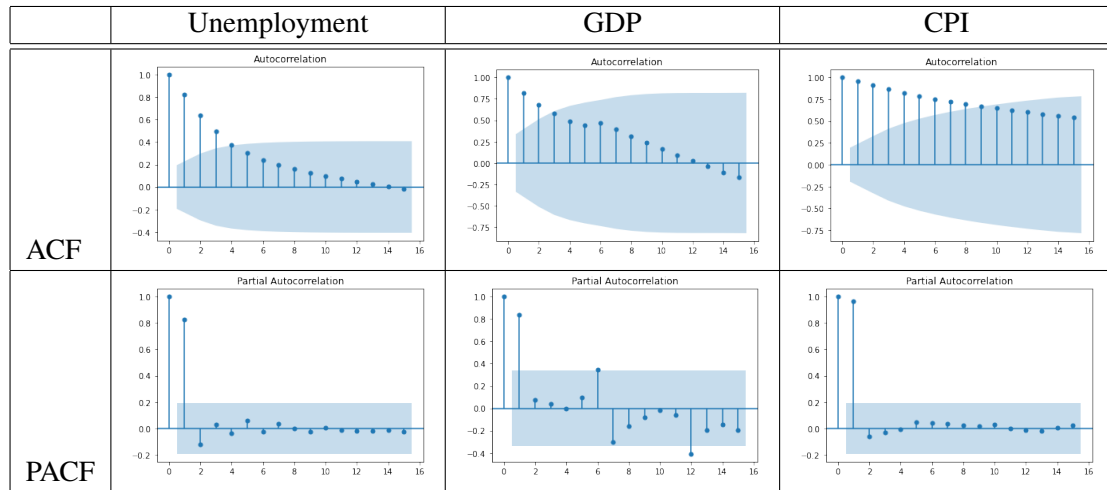
Table 3: ACT and PACF graph for macroeconomics variables

| LR model | | Unemployment Rate | | GDP | | CPI | |
|---|---|---|---|---|---|---|---|
| | | coeff. | s.e. | coeff. | s.e. | coeff. | s.e. |
| Benchmark | const | 0.9325 | 0.321*** | const 1,797.2 | 1,347.2 | const −4.7618 | 1.691*** |
| | lag_UNRATE | 0.8239 | 0.056*** | lag_GDPC1 0.9046 | 0.075*** | lag_CPIAUCNS 1.0209 | 0.007*** |
| | Adj. R-square | 0.683 | | 0.814 | | 0.995 | |
| Raw Proportion | const | 0.7069 | 0.239*** | const −1,060.9 | 1,601.7 | const −4.0379 | 1.515*** |
| | lag_UNRATE | 0.7069 | 0.239*** | lag_GDPC1 1.0842 | 0.119*** | lag_CPIAUCNS 1.0237 | 0.008*** |
| | positive_lr | −0.2076 | 1.711 | positive_lr 2,976.0 | 1638.220* | positive_lr −3.0473 | 1.163*** |
| | neutral_lr | −2.6617 | 1.890 | neutral_lr −164.2 | 1,654.5 | neutral_lr 1.9249 | 1.304 |
| | negative_lr | 3.5762 | 1.747** | negative_lr −3,872.7 | 2,349.3 | negative_lr −2.9155 | 1.781 |
| | Adj. R-square | 0.689 | | 0.823 | | 0.996 | |
| Difference | const | 0.9230 | 0.320*** | const −1,419.8 | 2,100.9 | const −4.2965 | 2.022** |
| | lag_UNRATE | 0.8256 | 0.056*** | lag_GDPC1 1.0843 | 0.117*** | lag_CPIAUCNS 1.0191 | 0.008''' |
| | sentiment_lr_diff | −1.8377 | 1.452 | sentiment_lr_diff 3,431.8 | 1767.010* | sentiment_lr_diff −0.5129 | 1.212 |
| | Adj. R-square | 0.685 | | 0.828 | | 0.995 | |
| Log Ratio | const | 1.3659 | 0.588** | const −1,205.5 | 2,841.5 | const −3.8099 | 2.300 |
| | lag_UNRATE | 0.8272 | 0.056*** | lag_GDPC1 1.0272 | 0.127*** | lag_CPIAUCNS 1.0182 | 0.008*** |
| | sentiment_lr_ratio | −0.4423 | 0.502 | sentiment_lr_ratio 791.9 | 661.097 | sentiment_lr_ratio ;−0.2568 | 0.419 |
| | Adj. R-square | 0.682 | | 0.816 | | 0.995 | |

***: significant at 1%, **: significant at 5%, *: significant at 10%.

Figure 1: Regression from Benchmark model, and Model with Sentiment (under LR model)

| LSTM model | | Unemployment Rate | | GDP | | CPI | |
|---|---|---|---|---|---|---|---|
| | | coeff. | s.e. | coeff. | s.e. | coeff. | s.e. |
| Benchmark | const | 0.9325 | 0.321*** | const 1,797.2 | 1,347.2 | const −4.7618 | 1.691*** |
| | lag_UNRATE | 0.8239 | 0.056*** | lag_GDPC1 0.9046 | 0.075*** | lag_CPIAUCNS 1.0209 | 0.007*** |
| | Adj. R-square | 0.683 | | 0.814 | | 0.995 | |
| Raw Proportion | const | 4.8532 | 1.954** | const 900.8 | 2,030.9 | const −5.2482 | 2.297** |
| | lag_UNRATE | 0.7840 | 0.058*** | lag_GDPC1 0.8199 | 0.144*** | lag_CPIAUCNS 1.0214 | 0.010*** |
| | positive_lstm | 2.3351 | 3.015 | positive_lstm 3,910.5 | 5,111.0 | positive_lstm −0.8027 | 3.400 |
| | neutral_lstm | −5.0419 | 2.358** | neutral_lstm 2,595.6 | 2,038.0 | neutral_lstm 0.7878 | 1.613 |
| | negative_lstm | −5.0419 | 2.358 | negative_lstm −5,605.2 | 5,842.1 | negative_lstm −5.2333 | 3.755 |
| | Adj. R-square | 0.694 | | 0.809 | | 0.995 | |
| Difference | const | 0.7531 | 0.402* | const 2,658.9 | 2,268.0 | const −5.2333 | 3.755** |
| | lag_UNRATE | 0.8257 | 0.056*** | lag_GDPC1 0.8478 | 0.142*** | lag_CPIAUCNS 1.0190 | 0.009*** |
| | sentiment_lr_diff | 2.3355 | 3.132 | sentiment_lr_diff 2,150.2 | 4,522.0 | sentiment_lr_diff 0.8985 | 3.000 |
| | Adj. R-square | 0.682 | | 0.809 | | 0.995 | |
| Log Ratio | const | 1.0399 | 0.435** | const 3,788.7 | 1970.439* | const −3.9617 | 1.932** |
| | lag_UNRATE | 0.8203 | 0.057*** | lag_GDPC1 0.7699 | 0.123*** | lag_CPIAUCNS 1.0170 | 0.008*** |
| | sentiment_lr_ratio | −0.0826 | 0.225 | sentiment_lr_ratio 398.2 | 290.879 | sentiment_lr_ratio 0.1584 | 0.184 |
| | Adj. R-square | 0.680 | | 0.819 | | 0.995 | |

***: significant at 1%, **: significant at 5%, *: significant at 10%.

Figure 2: Regression from Benchmark model, and Model with Sentiment (under LSTM model)

5

reliable if we can skip this sampling step. In addition, since the US economy can be affected by news in other countries, it will be also worth testing if including news elsewhere will help too.

Lastly, in the macroeconomics forecasting part, it would be contributive to test if other statistical model, e.g. Lasso, Ridge, Decision Tree will perform better too, just like what (Eleni Kalamara and Kapadia, 2020) did.

## References

Keven Bluteau Ardia, David and Kris Boudt. 2019. Questioning the news about economic growth: Sparse forecasting using thousands of news-based sentiment values. *International Journal of Forecasting*, pages 1370–1386.

Chris Redl George Kapetanios Eleni Kalamara, Arthur Turrell and Sujit Kapadia. 2020. Making text count: economic forecasting using newspaper text. *Bank of England*.

T. Daudert M. Huerlimann M. Zarrouk S. Handschuh K. Cortis, A. Freitas and B. Davis. 2017. Semeval-2017 task 5: Fine-grained sentiment analysis on financial microblogs and news. *Proc. 11th Int. Workshop Semantic Eval. (SemEval-), 2017, pp. 519–535*.

Leif Anders Thorsrud Larsen, Vegard H. and Julia Zhulanova. 2021. News-driven inflation expectations and information rigidities. *Journal of Monetary Economics*, pages 507–520.

Pyry Takala Pekka Korhonen Jyrki Wallenius Pekka Malo, Ankur Sinha. 2014. Good debt or bad debt: Detecting semantic orientations in economic texts. *Journal of the Association for Information Science and Technology*, pages 782–796.

Moritz Sudhof Shapiro, Adam Hale and Daniel J. Wilson. 2020. Measuring news sentiment. *Journal of Econometrics*.

Markus Ebner Tilly, Sonja and Giacomo Livan. 2021. Macroeconomic forecasting through news, emotions and narrative. *Expert Systems with Applications*.

Kostadin Mishev; Ana Gjorgjevikj; Irena Vodenska; Lubomir T. Chitkushev; Dimitar Trajanov. 2020. Evaluation of sentiment analysis in finance: from lexicons to transformers. *IEEE Access 8*.