Central Bank narratives and macroeconomic forecasting: using textual analysis from machine learning

Abstract

The purpose of the paper is to verify whether the tone/sentiment of those contained in the reports produced by the Central Bank of Brazil contain information that can be used to improve the accuracy of forecasts of macroeconomic indicators for one quarter ahead. Thus, we built predictors for inflation and GDP growth that were obtained from textual analyzes of the Copom Minutes and the Inflation Report. For the creation of sentiment scores we use a traditional fixed-lexicon dictionary approach and a new approach that uses machine learning to generate a time-varying dictionary. We then test the predictive power of the new variables for macroeconomic indicators for one period ahead. We also test whether these new predictors are able to improve the performance of prediction models. The results show that the best predictions were obtained with the models that used the time-varying dictionary textual score series. The fact happened because this type of dictionary is capable of incorporating new terms that appear in the reports. We also find that market average GDP growth forecasts can be improved with sentiment scores. But, this has not been verified for inflation.

Keywords: Macroeconomic forecasting. Machine learning. Textual analysis. Copom Minutes. Inflation Reports.

JEL Classification: C01, C22.

1 Introduction

Forecasting macroeconomic variables, in particular key indicators such as GDP growth, inflation and interest rates, are fundamental inputs for government budget planning, central bank policy formulation and business decisions. The use of time series approaches for macroeconomic forecasting gained momentum in the 1970s and 1980s, as the predictions of the univariate ARIMA (Box et al., 2015) and vector autoregressive (VAR) (Sims, 1980) models showed superior performance than structural macroeconomic models footnoteFor further discussion, see Diebold (1998) and its references. During this time, the sets of information used to form predictions often contained only a small number of variables. However, such methods have an important limitation, in that these models only support a small number of predictors. This situation changed in the early 2000s, when researchers began using large-scale macroeconomic data. In this scenario, models capable of dealing with a large number of predictors began to gain prominence. Therefore, the macroeconomic forecasting literature started to use factor models and shrinkage models of machine learning more frequently.

Two examples that can be found in the literature are the US dataset containing 149 variables measured with a monthly frequency shown in Stock and Watson (2002) and the euro area dataset containing 447 variables measured with a monthly frequency shown in Forni et al. (2003). In both studies, the use of a large number of predictors in a dynamic factor modeling framework (see Geweke (1977); Sargent, Sims et al. (1977)) performed better in industrial production forecasts compared to traditional reference models . An important factor in the popularity of this approach is its simplicity, where the principal components provide consistent estimates of dynamic factors and can subsequently be used in auxiliary predictive regressions. There is an extensive literature showing that when the factor model is used with a large number of predictors, it produces good predictions for macroeconomic variables such as GDP and inflation for many different economies (Eickmeier and Ziegler, 2008).

Despite its success, the dynamic factor model is not the only forecasting framework with a large number of predictors. Advances in machine learning statistics and literature were also explored in the macroeconomic context. For example, Mol, Giannone and Reichlin (2008) consider the regression of ridge and *least absolute shrinkage and selection operator* LASSO (Tibshirani, 1996) for data from Stock and Watson (2002) and obtained predictions with similar performance to that obtained with the dynamic factors model. Bai and Ng (2008) use LARS regression (Efron et al., 2004) to select a set of predictors. Predictions were produced using these predictors selected by the model. Bai and Ng (2008) show that, at least for some periods of the data, LARS-based methods produce better forecasts of inflation, income, retail sales, industrial production, and total employment compared to the principal factors model.

Studies that use methods that account for model uncertainty, such as bootstrap aggregation or bagging (see Breiman (1996); Bühlmann, Yu et al. (2002); Lee and Yang (2006)) have also appeared for inflation forecasting. Finally, in the multivariate forecast class, there was a focus on "large" VARs estimated using Bayesian techniques. Examples include Kadiyala and Karlsson (1997) and, more recently, Bańbura, Giannone and Reichlin (2010), Carriero, Kapetanios and Marcellino (2011) and Koop (2013) which use shrinkage priors.

In the case of Brazil, in recent years, a growing body of literature on macroeconomic forecasting has emerged, with emphasis on machine learning models. Medeiros and Mendes (2016) considered different high-dimensional models to predict Brazilian inflation. The authors showed that LASSO-based techniques have the smallest forecast errors for short-term forecasts. For longer horizons, the AR benchmark is the best model for predicting points, even if there are no significant differences between them. Factor models also produce good long-term forecasts in some cases. More recently, Garcia, Medeiros and Vasconcelos (2017) and Medeiros et al. (2019) used high-dimensional models to predict inflation in real time in Brazil and showed that the performance of shrinkage models is superior compared to more traditional techniques.

Barbosa, Ferreira and Silva (2020) analyzed the performance of high-dimensional factor models to predict four Brazilian macroeconomic variables: two real variables, unemployment rate and industrial production index, and two nominal variables, IPCA and IPC. The authors used three types of statistical learning techniques that were applied: shrinkage methods, combinations of predictions, and selection of predictors. Factors were extracted in a supervised and unsupervised manner. The results indicated that statistical learning methods improve the predictive performance of Brazilian economic variables.

Araujo and Gaglianone (2020) performed an out-of-sample prediction exercise using a variety of machine learning techniques and traditional econometric models. The results found by the authors corroborate recent conclusions in favor of non-linear automated procedures, indicating that machine learning algorithms (in particular, Random Forest) can outperform traditional prediction methods in terms of mean squared error.

However, there were some works in the international macroeconomic forecast literature that turned their focus to the construction of sentiment scores[1] from the textual information contained in reports published by the Central Bank of England (Jones, Sinclair and Stekler, 2019);(Clements and Reade, 2020), which they are used as predictors for the inflation rate, GDP and other macroeconomic indicators. These Central Bank polarity indicators can be used as direct predictors for the macroeconomic variables for the next period. They can also be used in more general models (eg VAR models, machine learning models, deep learning models, etc.) to improve the performance of such prediction models.

According to Jones, Sinclair and Stekler (2019) the evaluation of economic forecasts has typically focused on the quality of numerical forecasts. However, this focus on quantitative forecasts neglects the substantial amount of text that often accompanies them, particularly in central bank publications. Text is included to provide context and nuance and can be evaluated using textual analysis. In addition, numerical forecasts are not always publicly available, but the text of reports, minutes of meetings, and speeches can reveal information about the assessment of current and future economic conditions.

The literature has established the value of textual analysis as well as a general methodology for converting text into quantitative scores that primarily assess text polarities. According to Gentzkow, Kelly and Taddy (2019), information encoded in the text is a rich complement to the more structured data types traditionally used in empirical research. In fact, in recent years, there has been an intense use of textual data in different areas of research.

There is a rapidly expanding body of literature on the use of textual information, such as the subtle narratives of the sort that appear in inflation reports and central bank minutes. A key article is Stekler and Symington (2016) which investigates the "narratives" that constitute the Federal Open Market (FOMC) minutes between 2006 and 2010. His study quantifies the qualitative statements in the FOMC minutes about the current and future economic trends and compares the resulting indices to predictions from the Greenbook and the US Professional Forecast Survey.

Dossani (2019) analyzes how the tone of US Central Bank press conferences affects risk premiums in the foreign exchange market. He measures pitch as the difference between the number of hawkish and dovish phrases made during a press conference. He used four currency futures contracts traded on the Chicago Mercantile Exchange (CME) and found that implied risk aversion increases when central banks are hawkish and decreases when central banks are dovish.

Jones, Sinclair and Stekler (2019) performed qualitative analyzes by sentiment analysis of the texts of the inflation reports of the Central Bank of England in the period 2005-2014. They also built polarity indices out of the sample. They then compared the scores to real-time output growth data and the corresponding quantitative projections published by the Central Bank. They concluded that the general evolution of the UK economy was accurately represented in the text of the Inflation Report. In addition, efficiency regressions suggested that there is information in the text that could improve the Bank of England's quantitative reports and forecasts for one quarter ahead.

Clements and Reade (2020) analyzed the narratives accompanying the numerical predictions in the Bank of England's quarterly inflation reports for the period 1997–2018. The work focused on whether the narratives contain useful information about the future course of key macroeconomic variables in addition to point predictions, in terms of whether the narratives can

[1] According to Medhat, Hassan and Korashy (2014) sentiment analysis is the computational study of opinions, attitudes and emotions contained in written texts. In general, opinion mining helps to collect information about the positive and negative aspects of a specific topic.

be used to improve the accuracy of numerical predictions. It was also considered whether the narratives are able to predict future changes in numerical predictions. The authors concluded that a measure of sentiment derived from the narratives can predict errors in numerical forecasts of output growth but not inflation. They found no evidence that past shifts in sentiment predict subsequent changes in point forecasts of output growth or inflation, but they found that adjustments to numerical forecasts of output growth have a systematic element.

This literature for emerging countries like Brazil is scarce. Therefore, this paper seeks to fill this gap in the literature. In this context, the main objective is to build sentiment scores by textual analysis of the Copom Minutes (CM) and the Inflation Report (IR) produced by the Central Bank of Brazil (BCB). This paper will verify whether the narratives in the CM and in the IR contain useful textual information that can be used to improve the accuracy of forecasts of macroeconomic indicators, such as the inflation rate and GDP growth. So, to accomplish these goals we will perform two exercises. In the first one, we are going to make forecasts for the inflation rate and for the GDP growth through traditional models and machine learning models. In this case, we intend to know if the inclusion of sentiment scores in these models is able to improve the prediction accuracy. Let's compare the performance of our forecasts with the performance of the market average, represented by the Focus Survey. The second exercise will directly relate the textual information from the CM and IR with the Focus forecasts. In this case, let's check if the textual information is able to improve the Focus predictions.

Thus, the first step will be to build time series that represent the polarity in the texts of the CM and the IR using textual analysis methods. At this stage, we use the traditional Loughran and McDonald (2016) dictionary with fixed lexicons, which is widely used and disseminated in works that analyze financial texts. Additionally, we employ the Lima, Godeiro and Mohsin (2019) approach which allows dictionary content to vary over time.

After obtaining the sentiment scores series, we will verify if these new indicators are able to improve the performance of multivariate forecasting models. The multivariate models used in this exercise are the Vector Auto-Regressive (VAR), LASSO, Random Forest (RF), Factor Model (FM) and Support Vector Machines (SVM) models. As a benchmark we will use the forecasts provided by the Focus Survey. The forecasts will be carried out for each period ahead out of the sample (π_{t+h}) and will only take into account the information available so far X_t. Also, similar to Jones, Sinclair and Stekler (2019), let's test the efficiency of Focus's prediction by the extended version of the Mincer and Zarnowitz (1969) equation. In this case, it will be possible to know if the predictions made by Focus can be improved with the incorporation of all the information and nuances contained in the BCB narratives.

This research has some contributions. The first is to incorporate in the forecast performance comparison the use of a machine learning model that none of the previously mentioned work in the literature used, in this case the SVM. The second contribution is to use a time-varying dictionary method, as all the works in the literature cited above use fixed dictionaries. The third is to fill some gaps in the Brazilian macroeconomic forecasting literature through an alternative and innovative approach that is at the frontiers of the world literature. Finally, the fourth contribution is to show academic and market agents that the texts of the reports produced by the institutions, especially those from the Central Bank, contain relevant textual information for the forecasting exercise that cannot be ignored and should be incorporated into the framework methodology of the main agents that forecast macroeconomic variables. The results obtained show, in general, that the forecast models that use the series of sentiment scores present the smallest forecast errors, and in some cases obtaining forecast errors lower than Focus. The best predictions were obtained with the models that used the series of textual scores from the time-varying dictionary. This fact occurred because this type of dictionary is capable of incorporating new terms that appear in the reports. Thus, the series of scores constructed by variant dictionary were able to capture, for example, the Covid-19 pandemic, as the terms related to the pandemic emerged in 2019 and are not included in the fixed dictionaries, so the feeling of the reports is more realistic with the Time-varying dictionary.

Finally, the Mincer and Zarnowitz (1969) equation for the efficiency of Focus forecasts showed that sentiment scores are able to predict errors in Focus forecasts for GDP, so on average the market for GDP forecasts does not take taking into account all the information and textual nuances contained in the reports. Thus, the GDP growth forecasts carried out by Focus can be improved with sentiment scores. But, this was not verified for inflation. This fact is in line with the results previously found, as the Focus forecast errors for inflation are much smaller than for GDP. At the same time, our sentimental forecasts managed to outperform Focus only on forecasts made for GDP.

2 Data

2.1 Copom Minutes and Inflation Report

According to Filho and Rocha (2010) the Copom Minutes are one of the main communication tools of the Central Bank of Brazil, presenting economic projections for the national and international scenario, inflation control, decisions regarding interest rates, etc. It is through it that the monetary authority explains the procedures used to make decisions on monetary policy, with the aim of making communication more transparent and keeping expectations under control.

As in the work of Silva et al. (2020) who also used the Copom minutes for an exercise of textual analysis, it is necessary to take into account a change regarding the publication frequency, during the period 2000 to 2005 the meetings were held on a monthly basis. Like the publications, from 2006 onwards, meetings are held every forty-five days, with eight minutes being published a year. Another point that deserves to be highlighted is the fact that in 2002 thirteen minutes were published instead of twelve. The polarity series, after being estimated, needed to be converted into quarterly data with the aim that the entire sample would have the same frequency.

According to the BCB, the Inflation Report (IR) presents the policy guidelines adopted by the Copom, considerations about the recent evolution of the economic scenario and projections for inflation. The projections are presented in scenarios with conditions for some economic variables. The Copom uses a wide range of models and scenarios to guide its monetary policy decisions. By exposing some of these scenarios, the Copom seeks to make monetary policy decisions more transparent, contributing to their effectiveness in controlling inflation, which is its main objective.

The IR frequency is quarterly, however, the report has a "gap" between the years 2012 and 2015. In these years, the Copom only provides the IR executive summary, which is a summary of the document. Therefore, we chose to use the executive summary as a proxy for the IR during the years 2012 and 2015.

Thus, we chose to compose the sample with a temporal window starting in the first quarter of 2005 until the second quarter of 2020, totaling 62 observations. It was possible to obtain all CM and the IR by web scraping, in which we imported and worked with the documents in PDF format in the English language version.

2.2 Macroeconomic Variables

For the exercise of macroeconomic forecasting in machine learning models, a large set of macroeconomic and financial variables were used as predictors. The variables of interest to be predicted are the inflation rate and the GDP growth rate. For the inflation rate, we chose to choose the IPCA index, which is the official inflation index adopted by Copom for monetary policy purposes. The frequency and periodicity of the variables are identical to those in the inflation reports in the previous section. Table 1 and Table 2 illustrate all the variables.

Table 1 – List of macroeconomic and financial variables

Series	Category	Name	Source	Original
1	Inflation	IPCA (consumer price index)	IBGE	%
2	Inflation	IPCA (consumer price index, market prices)	IBGE	%
3	Inflation	IPCA (consumer price index, regulated and monitored prices)	IBGE	%
4	Inflation	IPCA (consumer price index, tradables)	BCB	%
5	Inflation	IPCA (consumer price index, nontradables)	BCB	%
6	Inflation	IPC-Fipe (consumer price index)	Fipe	%
7	Inflation	IPC-Br (consumer price index)	FGV	%
8	Inflation	IPA-DI (wholesale price index)	FGV	%
9	Inflation	IGP-DI (general price index)	FGV	%
10	Inflation	IGP-M (general price index)	FGV	%
11	Inflation	IGP-10 (general price index)	FGV	%
12	Inflation	INCC (national index of building costs)	FGV	%
13	Inflation	Core IPC-Br (core inflation)	FGV	%
14	Inflation	Core IPCA Exclusion EX0 (core inflation)	BCB	%
15	Inflation	Core IPCA Exclusion EX1 (core inflation)	BCB	%
16	Inflation	Core IPCA Double Weight (core inflation)	BCB	%
17	Inflation	Core IPCA Trimmed Means Smoothed (core inflation)	BCB	%
18	Inflation	Break Even Inflation (IPCA, 1 year)	Anbima	%
19	Inflation	Break Even Inflation (IPCA, 2 years)	Anbima	%
20	Inflation	Break Even Inflation (IPCA, 5 years)	Anbima	%
21	Interest rates	Nominal policy interest rate (Selic)	BCB	%
22	Interest rates	Nominal policy interest rate (longterm-interest rate, TJLP)	BCB	%
23	Interest rates	Nominal market interest rate (prefixed, 1 year)	Anbima	%
24	Interest rates	Nominal market interest rate (prefixed, 2 years)	Anbima	%
25	Interest rates	Nominal market interest rate (prefixed, 5 years)	Anbima	%
26	Interest rates	Nominal market interest rate (Swap Pré-DI, 1 year)	BCB	%
27	Interest rates	Real market interest rate (Swap Pré-DI, year, deflator: Focus 12m infl.expect.)	BCB	%
28	Interest rates	Real market interest rate (indexed IPCA, 1 year)	Anbima	%
29	Interest rates	Real market interest rate (indexed IPCA, 2 years)	Anbima	%
30	Interest rates	Real market interest rate (indexed IPCA, 5 years)	Anbima	%
31	Money	Monetary base	BCB	R$ thousand
32	Money	Money supply (currency outside banks)	BCB	R$ thousand
33	Money	Demand deposits	BCB	R$ thousand
34	Money	Savings deposits	BCB	R$ thousand
35	Money	M1	BCB	R$ thousand
36	Money	M2	BCB	R$ thousand
37	Money	M3	BCB	R$ thousand
38	Money	M4	BCB	R$ thousand
39	Banking sector	Credit spread (nonearmarked credit rate - Selic rate)	BCB	basis points
40	Banking sector	Non-Performing Loans (NPL) of total credit	BCB	%
41	Banking sector	Loan-to-Deposit ratio (LTD)	BCB	Units
42	Banking sector	Reserve requirements ratio (financial inst. reserve requirements / total deposits)	BCB	Units
43	Banking sector	Real growth of nonearmarked credit operations outstanding	BCB	R$ million
44	Banking sector	Initial Public Offers (IPOs) accumulated in 12 months (Brazil)	BCB	R$ million
45	Banking sector	Net equity of stock funds (Brazil)	BCB	R$ million
46	Banking sector	Net equity of financial investment funds (Brazil)	BCB	R$ million
47	Capital markets	Ibovespa (Brazil)	Reuters	Index
48	Capital markets	MSCI emerging countries (EM, US$)	Reuters	Index
49	Capital markets	MSCI developed countries (World, US$)	Reuters	Index
50	Risk premium	Embi+Br (Emerging Markets Bond Index Plus Brazil, spread)	Reuters	basis points
51	Risk premium	Embi+composite (average spread of 16 emerging countries)	Reuters	basis points
52	Risk premium	CDS (Credit Default Swap, Brazil 5 years)	Reuters	basis points
53	Exchange rates	FX-rate (nominal exchange rate, R$/US$)	BCB	Units
54	Exchange rates	REER (Real effective exchange rate, IPA-13 currencies)	Funcex	Index
55	Global Economy	U.S. dollar index (média geométrica de 6 em relação ao US$)	Reuters	Index
56	Global Economy	U.S. Treasury 2 years (Treasury nominal interest rates)	Reuters	%
57	Global Economy	U.S. Treasury 10 years (Treasury nominal interest rates)	Reuters	%
58	Global Economy	U.S. Treasury 5 years TIPS (Treasury Inflation Protected Securities)	Reuters	%
59	Global Economy	CRB all commodities index	Reuters	Index
60	Global Economy	Oil price (WTI, OklahomaUSA)	Reuters	US$/barrel
61	Global Economy	VIX CBOE volatility index (30 day expected volatility of the S&P500)	Reuters	Index

Table 2 – List of macroeconomic and financial variables - B

Series	Category	Name	Source	Original
62	Exterior	Import price index	Funcex	Index
63	Exterior	Import quantum index	Funcex	Index
64	Exterior	Export price index	Funcex	Index
65	Exterior	Export quantum index	Funcex	Index
66	Exterior	Imports (FOB, total)	MDIC/Secex	US$
67	Exterior	Exports (FOB, total)	MDIC/Secex	US$
68	Exterior	Exports (FOB, primary goods)	MDIC/Secex	US$
69	Exterior	International reserves (total)	BCB	US$ million
70	Exterior	Current account (monthly, net)	BCB	US$ million
71	Exterior	Current account (accumulated in 12 months, in relation to GDP)	BCB	%
72	Exterior	FDI (Foreign Direct Investment, accumulated in 12 months)	BCB	US$ million
73	Exterior	FPI (Foreign Portfolio Investment, accumulated in 12 months)	BCB	US$ million
74	Economic activity	IBC-BR (central bank economic activity index)	BCB	Index
75	Economic activity	GDP (accumulated in the last 12 months, current prices)	BCB	R$ million
76	Economic activity	Consumer confidence index	Fecomercio	Index
77	Labor	Unemployment rate (open)	IBGE	%
78	Labor	Registered employees index (wholesale and retail trade)	MTE	Index
79	Labor	Registered employees index (construction sector)	MTE	Index
80	Labor	Hours worked in production (São Paulo)	Fiesp	Index
81	Labor	Real overall wages (industry, São Paulo)	Fiesp	Index
82	Industry	Industrial production (total)	IBGE	Index
83	Industry	Industrial production (mineral extraction)	IBGE	Index
84	Industry	Industrial production (manufacturing industry)	IBGE	Index
85	Industry	Industrial production (capital goods)	IBGE	Index
86	Industry	Industrial production (intermediate goods)	IBGE	Index
87	Industry	Industrial production (consumer goods)	IBGE	Index
88	Industry	Industrial production (durable goods)	IBGE	Index
89	Industry	Industrial production (semidurable and nondurable goods)	IBGE	Index
90	Industry	Installed capacity utilization (São Paulo)	Fiesp	%
91	Industry	Capacity utilization (manufacturing industry, FGV)	FGV	%
92	Industry	Steel production	BCB	Index
93	Industry	Vehicles production (total)	BCB	Units
94	Industry	Passenger cars and light commercial vehicles production	BCB	Units
95	Industry	Truck production	BCB	Units
96	Industry	Bus production	BCB	Units
97	Industry	Production of agricultural machinery (total)	BCB	Units
98	Sales	Sales volume index in the retail sector (total)	BCB	Index
99	Sales	Sales volume index in the retail sector (fuel and lubricants)	BCB	Index
100	Sales	Sales volume index in the retail sector (hyperm., superm., food, bever. and tobacco)	BCB	Index
101	Sales	Sales volume index in the retail sector (textiles, clothing and footw ear)	BCB	Index
102	Sales	Sales volume index in the retail sector (furniture and white goods)	BCB	Index
103	Sales	Sales volume index in the retail sector (vehicles and motorcycles, spare parts)	BCB	Index
104	Sales	Sales volume index in the retail sector (hyperm. and supermarkets)	BCB	Index
105	Sales	Vehicle sales (total)	BCB	Units
106	Sales	Domestic vehicle sales	BCB	Units
107	Energy	Electric energy consumption (commercial)	Eletrobras	GWh
108	Energy	Electric energy consumption (residential)	Eletrobras	GWh
109	Energy	Electric energy consumption (industrial)	Eletrobras	GWh
110	Energy	Electric energy consumption (other)	Eletrobras	GWh
111	Energy	Electric energy consumption (total)	Eletrobras	GWh
112	Public sector	Primary result of consolidated public sector (current monthly flows)	BCB	R$ million
113	Public sector	Primary result of consolidated public sector (flow s accum. in 12 months)	BCB	R$ million
114	Public sector	Primary result of consolidated public sector (flow accum. in 12 months, % GDP)	BCB	%
115	Public sector	Net public debt (total, federal government and central bank, % GDP)	BCB	%
116	Public sector	Net public debt (internal, federal government and central bank, % GDP)	BCB	%
117	Public sector	Net public debt (external, federal government and central bank, % GDP)	BCB	%
118	Public sector	Net public debt (total, consolidated public sector, balances in reais)	BCB	R$ million
119	Public sector	Net public debt (internal, consolidated public sector, balances in reais)	BCB	R$ million
120	Public sector	Net public debt (external, consolidated public sector, balances in reais)	BCB	R$ million

3 Methodology

This section describes the methods used in this papers for forecasting. Thus, as Garcia, Medeiros and Vasconcelos (2017) we consider a direct forecasting approach in which the variable to be forecasted is h periods ahead, y_{t+h}, is modeled as a function of a set of measured predictors in time t, such as:

$$y_{t+h} = T(x_t) + u_{t+h} \tag{1}$$

where $T(x_t)$ is a mapping of a set of q predictors, u_{t+h} is the forecasting error, and $x_t = (x_{1t}, \ldots, x_{qt})' \in \mathbb{X} \subseteq \mathbb{R}^q$ may include weakly exogenous predictors, lagged values of the variable of interest and various factors calculated from a large number of potential covariates. It is important to point out that x_t contains only variables observed and available at the moment t. Note that considering direct prediction models for each horizon avoids the need to estimate a model for the evolution of x_t.

For most of the methods considered in this paper, the T(·) mapping is linear, so:

$$y_{t+h} = \beta' x_t + u_{t+h} \tag{2}$$

where $\beta \in \mathbb{R}^q$ is a vector of unknown parameters.

3.1 Sentiment scores

In this section, the methodology for constructing sentiment scores (S) from the texts of the CM and IR will be presented. Each S_t is intended to capture some of the narrative information in the report at the moment t, for each document in our sample. This measure turns thousands of words into a single number. To obtain each sentiment series S_t we use two approaches: one that measures sentiments from dictionaries with fixed lexicons and another that uses machine learning models to build a time-varying dictionary.

Before performing lexicographical analysis on the documents, we perform a series of transformations on the original text. The text is first split into a sequence of substrings (tokens) whose characters are all lowercase. We remove English stop words and clear the text using R package tolower.

According to Shapiro, Sudhof and Wilson (2020) there are two general methodologies to quantify the sentiment in the text. The first is known as lexical methodology. This approach is based on predefined lists of words, called lexicons or dictionaries, with each word assigned a score for the emotion of interest. Generally, these scores are simply 1, 0, and -1 for positive, neutral, and negative, but some lexicons have more than three categories. Typical applications of this approach measure the emotional content of a given corpus of text based on the prevalence of negative vs. negative words in the corpus. These word-matching methods are called bag-of-words (BOW) methods because of the contextual characteristics of each word, such as its order in the text, grammatical class, co-occurrence with other words, and other contextual characteristics specific to the text in which the word appears, are ignored.

Among this type of method, the dictionary created by Loughran and McDonald (2011) (LM) stands out. The authors constructed lists of negative and positive words that are selected to be appropriate for the financial text. They show that their dictionaries are superior for classifying economic and financial texts than other dictionaries, for example, Apel and Grimaldi (2012) and the Harvard Psychosociological Dictionary, which tends to miscategorize neutral words in a financial/economic context (eg, taxes, costs, capital, expense, liability, risk, excess and depreciation). There are 2,355 negative words and 354 positive words in the LM dictionaries. Therefore, for the construction of sentiment scores via the fixed dictionary approach, we use the LM dictionary.

Shapiro, Sudhof and Wilson (2020) states that the second, more incipient approach employs machine learning techniques to build complex models to probabilistically predict the sentiment of a given set of text. One of the applications of machine learning models is in the construction of time-varying dictionaries. Lima, Godeiro and Mohsin (2019) used this approach to create a variant dictionary method.

According to Lima, Godeiro and Mohsin (2019) the assumption of a time-invariant dictionary does not appear to be realistic in documents that introduce new words over time or whether the vocabulary used in periods of recession differs from that used in periods of economic expansion. The authors emphasize that even if vocabulary were constant over time, the predictive power of some words may vary, that is, the relevance of words changes over time, but the existing literature does not explain this effect and, therefore, the resulting predictors do not reflect the most predictive textual information found in documents at any given time. Therefore, we apply the approach developed by Lima, Godeiro and Mohsin (2019) for the construction of sentiment indices via time-varying dictionaries.

Using the methodology proposed by the authors to build the time-varying dictionary, we first create a time series vector, X_t, in which each element of the vector shows time series observations of the frequency at which each word (or combination of words) appears in CM and IR up to time t. Therefore, this step transforms the words into numeric values without using a pre-specified (fixed) dictionary. This numerical representation is large and sparse; therefore, dimensionality reduction should be employed in the next step. In the second step, we use supervised machine learning to select the most predictive time series (words) $X_t^* \subset X_t$.

The elastic net model was chosen to perform the second stage:

$$y_{t+h} = W_t' \boldsymbol{\beta}_h + X_t' \phi_h + \epsilon_{t+h} \tag{3}$$

where $h \geq 0$ is the forecast horizon, $\widehat{\beta}_h$ and $\widehat{\phi}_h$ are estimated by minimizing the following objective function:

$$\min_{\beta_h, \phi_h} \sum_t \left(y_{t+h} - W_t' \beta_h - X_t' \phi_h\right)^2 + \lambda_1 \left\|\phi_h\right\|_{\ell_1} + \lambda_2 \left\|\phi_h\right\|_{\ell_2} \tag{4}$$

where W_t is a vector $k\ x\ 1$ of predetermined predictors, such as lags of y_t as well as traditional structured data predictors and $\| \cdot \|_{ell_1}$ and $\| \cdot \|_{\ell_2}$ are the standard ℓ_1 and ℓ_2, respectively. Then, from the selection of words with greater predictive power, we have for each period t a set of words that serve as a lexical dictionary to obtain the series of feelings S. In this paper, we will apply this method to inflation and GDP.

Finally, both dictionary approaches calculate the sentiment index by the difference between positive and negative words, divided by the sum of positive and negative words, as proposed by Hubert and Labondance (2018):

$$S_t = \frac{\text{PositiveWords}_{\text{ t}} - \text{ NegativeWords}_{\text{ t}}}{\text{PositiveWords}_{\text{ t}} + \text{ NegativeWords}_{\text{ t}}} \tag{5}$$

Therefore, we obtain the measure of sentiment score, S, which varies between -1 and 1.

3.2 Forecasting models

3.2.1 ARMA

One of the most common statistical models used for time series prediction is the autoregressive moving average (ARMA) model, which assumes that future observations are mainly guided by recent observations. Information that generally exhibits persistent behavior is largely consistent with this assumption. The simplest representation is the AR (1), described below:

$$y_t = \alpha + \beta y_{t-1} + \varepsilon_t \tag{6}$$

where the estimated parameters $[\widehat{\alpha}; \widehat{\beta}]'$ can be calculated using a sample with $t = 1,...,T$ observations:

$$y_{t+h} = \widehat{\beta}^h y_T + \sum_{i=0}^{h-1} \widehat{\alpha}\widehat{\beta}^i \tag{7}$$

where y_{t+h} is the h-steps ahead prediction.

3.2.2 Factor Model

The idea that time variations in a large number of variables can be summarized by a small number of factors is empirically attractive and is employed in a large number of studies in economics and finance (Forni et al., 2000); (Stock and Watson, 2002). Let $x_{i,t}$ be the observed data for the i-th unit of cross-section in time t, for $i = 1,...,N$ and $t = 1,...,T$, and consider the following factorial representation of the data:

$$x_{i,t} = \lambda_i' F_t + e_{i,t} \tag{8}$$

where F_t is a vector of common factors, λ_i is a vector of factor loadings associated with F_t and $e_{i,t}$ is the indiosyncratic component of $x_{i,t}$. Note that F_t, λ_i and $e_{i,t}$ are unknown, as only $x_{i,t}$ is observable. Here, we estimate the factors and their loads using principal component analysis (PCA), which is a well-established technique for dimension reduction in time series. The number of components is determined by the Bai and Ng (2002) criterion. After estimating the PCA of common factors F_t, we employ the direct forecast approach to model the inflation rate and GDP growth over time $t + h$, as follows:

$$y_{t+h} = \beta_h F_t + \varepsilon_{t+h} \tag{9}$$

Therefore, the information prediction of the direct factor model approach, y_{t+h}, using a sample of $t = 1,...,T$ observations, is given by:

$$y_{t+h} = \widehat{\beta_h} \widehat{F_T}, \quad \text{for } h = 1, \ldots, H \tag{10}$$

where $\widehat{\beta_h}$ and $\widehat{F_T}$ are the parameters to be estimated.

3.2.3 LASSO

In this type of shrinkage method, the idea is to reduce the parameters that correspond to irrelevant variables to zero. Under some conditions, it is possible to estimate models with more variables than observations.

Among the shrinkage methods, LASSO, introduced by Tibshirani (1996) has received special attention. It has been shown that LASSO can handle more variables than observations, and the correct subset of relevant variables can be selected (Efron et al., 2004); (Meinshausen, Yu et al., 2009); (Zhao and Yu, 2006).

The LASSO estimator is defined as:

$$\hat{\boldsymbol{\beta}} = \arg\min_{\hat{\beta}} \left[\sum_{t=1}^{T} (y_{t+h} - \boldsymbol{\beta}' \boldsymbol{x}_t)^2 + \lambda \sum_{j=1}^{k} |\beta_j| \right] \tag{11}$$

where λ controls the amount of shrinkage and is determined by data-driven techniques such as cross-validation or use of information criteria. The model works even when the number of variables increases faster than the number of observations and when the errors are non-Gaussian and heteroskedastic.

3.2.4 Random Forest

The random forest (RF) methodology was initially proposed by Breiman (2001) as a way to reduce the variation of regression trees and is based on bootstrap aggregation (bagging) of randomly constructed regression trees.

According to Garcia, Medeiros and Vasconcelos (2017) a regression tree is a nonparametric model based on the recursive binary partitioning of the covariate space $\mathbb{X}$, where the function $\mathrm{T}(\cdot)$ is a sum of local models (usually just a constant), each of which is determined in $K \in \mathbb{N}$ different regions (partitions) of $\mathbb{X}$. The model is usually displayed in a graph that has the format of a binary decision tree with nodes $N \in \mathbb{N}$ parent (or split) and nodes $K \ \ in\ \mathbb{N}$ terminals (also called leaves) and that grows from the root node to the terminal nodes. Typically, partitions are defined by a set of hyperplanes, each orthogonal to the axis of a particular predictor variable, called a split variable. Therefore, conditional on knowledge of the subregions, the relationship between y_{t+h} and x_t in Equation 1 is approximated by a constant model in parts, in which each leaf (or node terminal) represents a distinct regime.

A complex regression tree model can be mathematically represented by introducing the following notation. The root node is at position 0 and a parent node at position j generates left and right child nodes at positions $2j+1$ and $2j+2$, respectively. Every parent node has an associated split variable $x_{s_j t} \in \mathbf{x}_t$, where $s_j \in \mathrm{S} = \{1, 2, \ldots, q\}$. Furthermore, if we allow $\mathbf{J}$ and $\mathbf{T}$ to be the sets of indexes of the parent and terminal nodes, respectively, a tree architecture can be determined completely from $\mathbf{J}$ and $\mathbf{T}$.

The prediction model based on regression trees can be represented mathematically as:

$$y_{t+h} = H_{\mathbb{JT}}\left(\mathbf{x}_t; \psi\right) + u_{t+h} = \sum_{i \in \mathbb{T}} \beta_i B_{\mathbb{J}i}\left(\mathbf{x}_t; \boldsymbol{\theta}_i\right) + u_{t+h} \tag{12}$$

where:

$$\begin{aligned} B_{\mathbb{J}\mathrm{i}}\left(\mathbf{x}_t; \boldsymbol{\theta}_i\right) = \prod_{j \in \mathbb{J}} I\left(x_{s_j,t}; c_j\right)^{\frac{n_{i,j}\left(1+n_{i,j}\right)}{2}} \\ \times \left[1 - I\left(x_{s_j,t}; c_j\right)\right]^{(1-n_{i,j})(1+n_{i,j})} \end{aligned} \tag{13}$$

$$I\left(x_{s_j,t}; c_j\right) = \begin{cases} 1 & \text{if } x_{s_j,t} \leq c_j \\ 0 & \text{otherwise} \end{cases} \tag{14}$$

$$n_{i,j} = \begin{cases} -1 & \text{if the path to the leaf } i \text{ does not include the parent node} \\ 0 & \text{if the path to the leaf } i \text{ include} \\ & \text{the right child node of the parent node } j \\ 1 & \text{if the path to the leaf } i \text{ include} \\ & \text{the left child node of the parent node } j \end{cases}$$

Let $\mathbb{J}_i$ be the subset of $\mathbb{J}$ that contains the indices of the parent nodes that form the path to the leaf i; then θ_i is the vector that contains all the parameters c_k, so $k \in \mathbb{J}_i$, $i \in \mathbb{T}$. Note that $\sum_{j \in \mathbb{J}} B_{\mathbb{J}i}\left(\mathbf{x}_t; \boldsymbol{\theta}_j\right) = 1, \forall \mathbf{x}_t \in \mathbb{R}^{q+1}$.

A random forest is a collection of regression trees, each specified in an initialization subsample of the original data. Suppose there are subsamples with initialization B and denote the regression tree estimated for each of the subsamples by $H_{\mathbb{J}_b \mathbb{T}_b}(\cdot; \psi_b)$. The final forecast is defined as:

$$\widehat{y}_{t+h} = \frac{1}{B} \sum_{b=1}^{B} H_{\mathbb{J}_b \mathbb{T}_b}(x_t; \psi_b) \tag{15}$$

A regression tree is estimated for each of the initialization subsamples, repeating the following steps recursively for each terminal node of the tree until the minimum number of observations at each node is reached.

1. Randomly select m from q covariates as possible division variables. variáveis.

2. Choose the best variable/split point among the m candidates.

3. Divide the node into two child nodes.

The random forest models can handle a very large number of explanatory variables, and the predicted model is highly non-linear. It is important to note that bootstrap samples are calculated using block bootstraps as we are dealing with time series.

3.2.5 Support Vector Machine

Since SVM was introduced from statistical learning theory by Vapnik (1995), a number of studies have been announced about its theory and applications. Compared with most other learning techniques, SVM leads to increased performance in pattern recognition, regression estimation, financial time series prediction, among other applications. The following brief description of SVM focuses entirely on the pattern recognition problem in the field of classification. The detailed explanation and proofs of SVM can be contained in the books (Vapnik, 1995) and (Vapnik, 1999).

According to Shin, Lee and Kim (2005) the SVM produces a binary classifier, the so-called optimal separation hyperplanes, by means of the extremely non-linear mapping of the input vectors in the space of high-dimensional resources. SVM builds a linear model to estimate the decision function using nonlinear class bounds based on support vectors. If the data is separated linearly, the SVM trains linear machines to an ideal hyperplane that separates the data without error and at the maximum distance between the hyperplane and the nearest training points. The training points closest to the ideal separation hyperplane are called support vectors. All other training examples are irrelevant to determining the limits of binary classes. In general cases where the data is not separated linearly, the SVM uses non-linear machines to find a hyperplane that minimizes the number of errors in the training set.

3.3 Forecast accuracy

Formal tests of predictive power can be done using approaches such as those popularized by Diebold and Mariano (2002) (see, for example, Clark and McCracken (2010) and Giacomini (2010), for recent analyses) to see if differences in the RMSE reflect differences statistically significant differences between predictions. Thus, as in the work of Clements and Reade (2020), we apply a test based on regression (Diebold and Mariano, 2002), where we construct the term e_t^{DM} as:

$$e_t^{DM} = L\left(y_t - \hat{y}_{t|t-h}\right) - L\left(y_t - \hat{y}_{t|t-h}^{EM}\right) \tag{16}$$

where $\hat{y}^{EM}_{t|t-h}$ is the reference forecast, in our case the reference is the Focus forecasts. Conventionally, the loss function L is the squared error loss, that is, $L(e) = e^2$. Equal prediction performance implies $E\left(e_t^{DM}\right) = 0$, and this test can be implemented using the regression model:

$$e_t^{DM} = \alpha + u_t \tag{17}$$

with null hypothesis of equal precision, $H_0 : \alpha = 0$. The significance of α implies a difference in prediction performance. If $\alpha > 0$, this implies that the benchmark forecast is 'better' than the realized forecast, while $\alpha < 0$ implies the opposite, that is, that the realized forecast is different from the forecast of the benchmark.

3.4 Forecast efficiency

At this stage, we will compare the projections made by the market, represented by the Focus, and the qualitative information contained in the text of the CM minutes and the IR, regressing the Focus forecast errors as a function of sentiment scores, and we will repeat the same procedure Focus forecast errors for one quarter ahead. As with Jones, Sinclair and Stekler (2019), for this comparison exercise we will use an extended version of a regression of Mincer and Zarnowitz (1969). If Focus predictions contain all of the relevant information included in the text, then Focus prediction errors should not be predicted by sentiment scores. The equation can be adapted as follows:

$$e_{t+h}^{Focus} = \alpha + \beta S_t + u_t \tag{18}$$

where e_{t+h}^{Focus} is the Focus forecast error which is defined as the difference between the realized values and the Focus forecast values, S_t the sentiment index and u_t is the residue. Testing the null hypothesis that the sentiment index coefficient is zero allows us to determine whether or not Focus's predictions were improved by incorporating information from the text.

4 Results

Three scores were built from the CM and three from the IR, totaling six sentiment scores. The indexes and their respective definitions can be seen in Table 3.

Table 3 – Definition of sentiment scores

Score	Definition
CM_Index	Copom Minutes sentiment index using the fixed dictionary
IR_Index	Inflation Report sentiment index using the fixed dictionary
IPCA_CM_Index	Copom Minutes sentiment index using the variant dictionary selecting the most predictive words for the IPCA
IPCA_IR_Index	Inflation Report sentiment index using the variant dictionary selecting the most predictive words for the IPCA
GDP_CM_Index	Copom Minutes sentiment index using the variant dictionary selecting the most predictive words for the GDP
GDP_IR_Index	Inflation Report sentiment index using the variant dictionary selecting the most predictive words for the GDP

The variables CM_Index and IR_Index were built using the Loughran and McDonald (2011) dictionary, so they are sentiment scores of a list of pre-selected words that do not change over time. The IPCA_CM_Index and IPCA_IR_Index are created by a set of words that vary at each period t, that is, for each report, the dictionary consists of the most predictive words for the IPCA that are selected via machine learning. The same happens for the GDP_CM_Index and GDP_IR_Index, the difference lies in the answer variable, in this case, the GDP. The main advantage of feeling indices from the variant dictionary is that they are able to represent a more realistic feeling, as they can more accurately measure the feeling of the document when extreme events occur, such as the Covid-19 pandemic.

The Figure 7 and Figure 8, in appendix, illustrate this point well. It is possible to see in Figure 7 the 10 most positive and 10 most negative coefficients for the IPCA in the CM and IR, that is, the most predictive words for the IPCA in these documents. In the CM for the IPCA, "budget", "certain" and "gas" were the three most positive words and "coronavirus", "covid" and "basket" were the most negative. In the IR, the three most positive words are "vaccines", "arabia" and "fuels", but the most negative are "election", "cigarretes" and "surprises", but terms that represent scenarios of environmental disaster such as "brumadinho" also appear. " and the current Covid-19 scenario as "epidemic". For GDP we also have that the terms "coronavirus", "covid", "infections" and "vaccines" are the most predictive. In the case of fixed dictionary indices, it is impossible for these series to accurately capture the Covid-19 pandemic, as words such as "covid" and "coronavirus" are not present in such a dictionary.

When we analyze the size of documents we notice that they vary with time, with the number of words decreasing over the quarters, this is evident in the Figure 5, in appendix. The CM in the first quarter of 2005 contained approximately 5,328 words and in the second quarter of 2020, an amount of 1,797 words. The number of words in the CM was drastically reduced from 2014 onwards. In the first quarter of 2005, the IR had 57,091 words and 36,664 at the end of the sample. Still on IR, we can see from Figure 5 that between 2012 and 2015 there is an extreme reduction in the number of words. As explained in the database section, this was because during these years the BCB did not produce the RI, it only provided an executive summary of the document.

The most frequent words in the CM and in the IR can be seen in Figure 6, in appendix, which shows the word cloud of each document. Words such as "growth", "goods", "rate", "credit", "industrial", "exchange", "expectation" and "sales" are the most common terms found in both documents. The word clouds also help to show that despite the IR being a document with a much higher number of words compared to the CM, the main terms that appear in these documents are similar.

Table 9 shows the correlations between sentiment scores with the IPCA and GDP growth. It is possible to notice that the fixed dictionary sentiment scores, CM_Index and IR_Index, have low correlations with the macroeconomic variables. This fact is not entirely abnormal, as similar works such as Catalfamo et al. (2018) and Mathy and Stekler (2018) also found low correlations with data from the United States.

With respect to variant dictionary sentiment scores, correlations increase significantly. We can also see that indices that use GDP growth in their construction present correlations with GDP greater than the correlation with the IPCA indices that use the IPCA. Furthermore, when we look at the variant dictionary indices, we see that those from the CM produce higher correlations.

The highest correlation of all was between GDP_CM_Index and GDP growth (0.75). This value is lower than that found in works by Jones, Sinclair and Stekler (2019) and Clements and Reade (2020) that used data from the Central Bank of England and used a fixed dictionary. However, the authors did not use the "raw" sentiment index, rather using them as predictors they improved them to be closer to GDP and inflation by scaling. So if they used the original fixed dictionary scores, they would get much lower correlations.

Another interesting fact is that time-varying sentiment scores have even greater correlations with the IPCA and GDP for one quarter ahead. This shows that the textual sentiment of the BCB reports may contain information that already anticipates movements in GDP growth and inflation.

Figure 1 – Quarterly trajectory of sentiment scores and IPCA

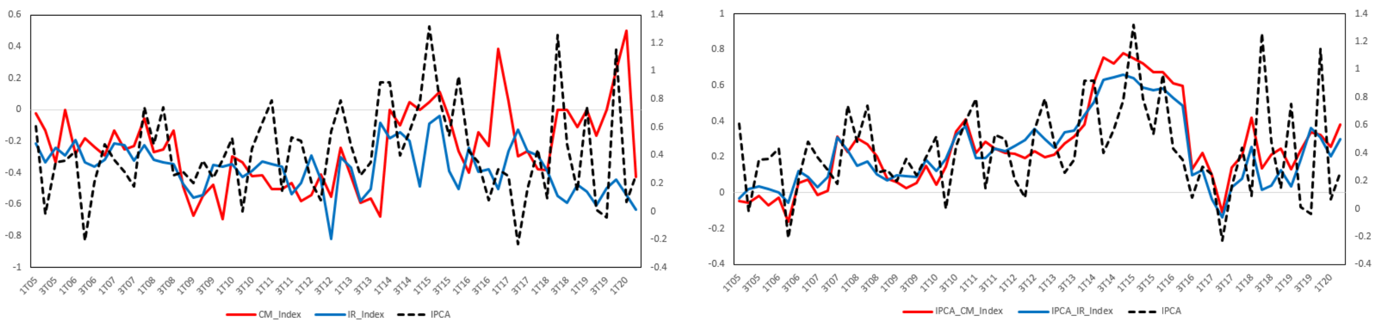


Figure 2 – Quarterly trajectory of sentiment scores and GDP growth

Figure 1 and Figure 2 graphically illustrate the movement of sentiment scores along with GDP growth and IPCA over time. Despite the different scales, we can see that the variant dictionary sentiment scores better fit the macroeconomic variables. Then, the graphical analysis visually confirms what has already been seen in the correlations.

The IPCA_CM_Index and IPCA_IR_Index indexes follow with greater precision the peaks and valleys of the IPCA in relation to the CM_Index and IR_Index indicators and as already mentioned, due to the low correlation between the sentiment scores with fixed dictionary. However, CM_Index and IR_Index track GDP movements more closely than the IPCA. It is visible that these variables are able to capture the sharp drop in output due to the Subprime crisis and are also able to track the fall in GDP due to fiscal problems in 2015, but fail to capture the abrupt reduction in GDP in the first and second quarter of 2020 from the Covid-19 pandemic. This fact was already expected, as the Loughran and McDonald (2011) financial dictionary has terms capable of detecting financial and fiscal crises, but they are incapable of directly measuring crises such as the Coronavirus. CM_Index is still able to follow the GDP fall late from the second quarter of 2020, as the Coronavirus had already significantly deteriorated the Brazilian economy, thus the documents began to have a pessimistic feeling due to the fact of negative terms appear later more often.

When we compare the textual scores from the perspective of the source report, we see that, in general, the sentiment series from the CM are more optimistic in relation to the IR series, including the IR_Index does not have any positive value. In addition, we note that the CM scores of origin have strong peaks of optimism after 2014, while the RI sentiment indices are more pessimistic. This must be due to the different function and structure between the documents, as the CM are a document with a more communicative and objective nature, whereas the IR is a more analytical document. So, as the BCB uses the Copom Minutes to communicate the monetary policy decision, the document tends to be more optimistic, anchoring market expectations and we note that from 2014 onwards, the tone of the CM started to have a more optimistic pattern. The indices built by Jones, Sinclair and Stekler (2019) and Clements and Reade (2020) that use the IR of the Central Bank of England also follow the line of our IR origin indices, that is, they tend to have a more pessimistic tone.

4.1 Forecast accuracy

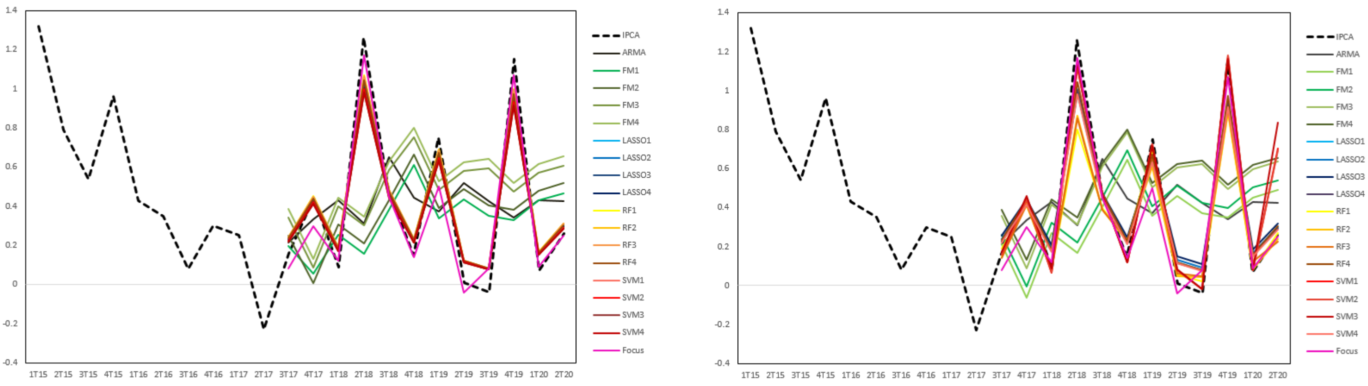
For our prediction exercise we use five types of prediction models: ARMA, Factor Model, LASSO, Random Forest and SVM. Our benchmark is the average forecasts made by the market, represented by the Focus Survey. In order to verify whether the inclusion of sentiment scores in forecasting models are able to improve the accuracy of forecasts for the IPCA and GDP, we estimated multivariate models with different categories. In this case, the first category is models that include all sentiment scores (with fixed dictionary and variant dictionary), the second uses only variant dictionary scores, third uses only fixed dictionary scores, and finally the fourth category it is from models with no sentiment index. Table 4 below illustrates the different categories of models.

Table 4 – Forecasting Models

Model	Definition
LASSO1	LASSO model with all macroeconomic variables + sentiment score with variant dictionary + sentiment score with fixed dictionary
LASSO2	LASSO model with all macroeconomic variables + sentiment score with variant dictionary
LASSO3	LASSO model with all macroeconomic variables + sentiment score with fixed dictionary
LASSO4	LASSO model with all macroeconomic variables
RF1	Random Forest model with all macroeconomic variables + sentiment score with variant dictionary + sentiment score with fixed dictionary
RF2	Random Forest model with all macroeconomic variables + sentiment score with variant dictionary
RF3	Random Forest model with all macroeconomic variables + sentiment score with fixed dictionary
RF4	Random Forest model with all macroeconomic variables
SVM1	Support Vector Machines model with all macroeconomic variables + sentiment score with variant dictionary + sentiment score with fixed dictionary
SVM2	Support Vector Machines model with all macroeconomic variables + sentiment score with variant dictionary
SVM3	Support Vector Machines model with all macroeconomic variables + sentiment score with fixed dictionary
SVM4	Support Vector Machines model with all macroeconomic variables
FM1	Factor Model with all macroeconomic variables + sentiment score with variant dictionary + sentiment score with fixed dictionary
FM2	Factor Model with all macroeconomic variables + sentiment score with variant dictionary
FM3	Factor Model with all macroeconomic variables + sentiment score with fixed dictionary
FM4	Factor Model with all macroeconomic variables

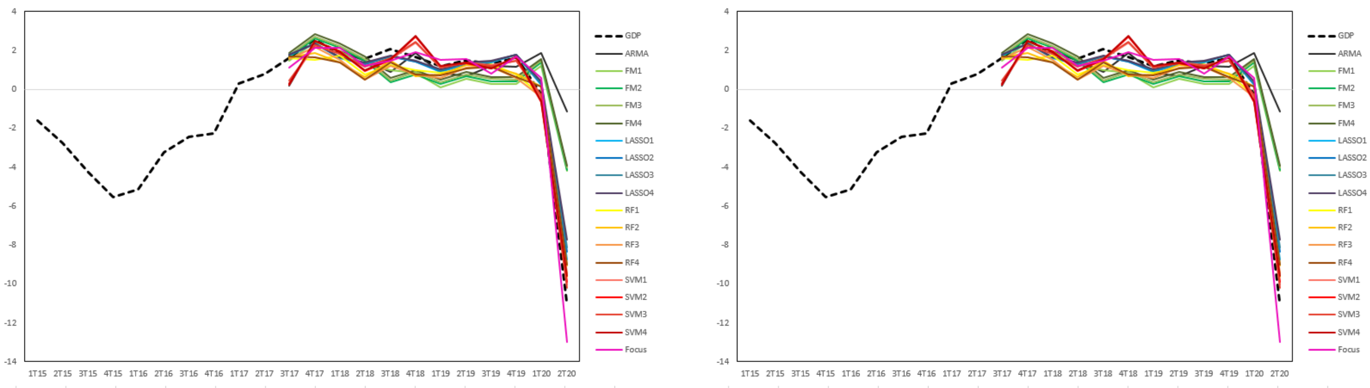
Figure 3 and Figure 4 illustrates the IPCA forecasts and GDP growth[2], respectively, outside the sample and considering h = 1, that is, one quarter ahead. The total sample goes from the first quarter of 2005 to the second quarter of 2020, totaling 62 observations and 80% of the total sample was used to compose the model training sample and 20% to compose the validation sample. In both figures we can see that the Focus forecasts are very close to the actual data and in some quarters the machine learning models are able to surpass the Focus.

Figure 3 – IPCA forecasts for h = 1 - Copom Minutes and Inflation Report



Copom Minutes   Inflation Report

Figure 4 – GDP forecasts for h = 1 - Copom Minutes and Inflation Report



Copom Minutes   Inflation Report

Table 5 and Table 6 illustrates the forecast error (MSE) of each model for up to four quarters ahead, together with the Diebold-Mariano test. In this case, for the DM test we consider the Focus as a reference, so that the null hypothesis is the same prediction as the Focus and the alternative hypothesis is the prediction different from the Focus.

It was possible to verify some facts. The first is that, as reported in the works of ((Medeiros and Mendes, 2016), (Medeiros et al., 2019), (Araujo and Gaglianone, 2020)) the machine learning models presented a satisfactory performance in terms of accuracy for one period ahead, as the forecasts were close to the Focus forecasts and some even surpassing the benckmark, however, for more periods ahead, the forecasts are already starting to move away from the Focus. In other words, short-term forecasts of up to one quarter ahead, we were able to get closer and even surpass Focus, but for quarters ahead, our forecasts are far from Focus. The second is that the MSEs reported for the IPCA are lower than the MSEs for the GDP. So,

[2] In Figure 3 and Figure 4, we chose to show the series from the first quarter of 2015 onwards, for a better visualization.

as expected, both for us and for the market average, it is easier to predict inflation than output growth.

The third is that models that included sentiment indices performed better than models that did not incorporate the polarity series. The best models are those that use both fixed dictionary and variant dictionary sentiment scores (category 1) at the same time, followed by models that consider only variant dictionary sentiment indices. Another fact is that the models that used the series of sentiments from the CM obtained a lower MSE, both for the IPCA and for GDP growth.

Table 5 – Mean Squared Error (MSE) and Diebold-Mariano test for IPCA predictions

	Copom Minutes				Inflation Report			
Modelo	h = 1	h = 2	h = 3	h = 4	h = 1	h = 2	h = 3	h = 4
ARMA	0.2147***	0.2404***	0.2833***	0.3222***	0.2147***	0.2404***	0.2833***	0.3222***
FM1	0.2550***	0.2707***	0.2908***	0.3587***	0.2583***	0.2894***	0.3003***	0.3597***
FM2	0.2500***	0.2916***	0.3130***	0.3693***	0.2541***	0.2917***	0.3237***	0.3612***
FM3	0.2649***	0.3034***	0.3317***	0.3713***	0.2741***	0.3372***	0.3904***	0.4481***
FM4	0.2745***	0.3341***	0.3850***	0.4399***	0.2745***	0.3341***	0.3850***	0.4399***
LASSO1	0.0124	0.0130*	0.0142**	0.0151**	0.0147*	0.0176***	0.0195***	0.2310***
LASSO2	0.0126	0.0164***	0.0180***	0.0191***	0.0131	0.0192***	0.0199***	0.2050***
LASSO3	0.0143*	0.0203***	0.0222***	0.0236***	0.0151**	0.0195***	0.0211***	0.0236***
LASSO4	0.0196***	0.0204***	0.0223***	0.0235***	0.0196***	0.0204***	0.0223***	0.0255***
RF1	0.0190***	0.0259***	0.0263***	0.0291***	0.0208***	0.0242***	0.0264***	0.0268***
RF2	0.0202***	0.0267***	0.0279***	0.0317***	0.0218***	0.0257***	0.0279***	0.0345***
RF3	0.0204***	0.0241***	0.0260***	0.0320***	0.0215***	0.0285***	0.0299***	0.0336***
RF4	0.0231***	0.0259***	0.0280***	0.0332***	0.0231**	0.0259***	0.0280***	0.0332***
SVM1	0.0160***	0.0188***	0.0224***	0.0262***	0.0168***	0.0183***	0.0229***	0.0266***
SVM2	0.0156***	0.0161***	0.0199***	0.0230***	0.0175***	0.0189***	0.0231***	0.0280***
SVM3	0.0173***	0.0183**	0.0222***	0.0259***	0.0196***	0.0311***	0.0364***	0.0401***
SVM4	0.0224***	0.0300***	0.0348***	0.0388***	0.0224***	0.0300***	0.0348***	0.0388***
Focus	0.0102	0.0126	0.0130	0.0140	0.0102	0.0126	0.0130	0.0140

Table 6 – Mean Squared Error (MSE) and Diebold-Mariano test for GDP predictions

	Copom Minutes				Inflation Reports			
Modelo	h = 1	h = 2	h = 3	h = 4	h = 1	h = 2	h = 3	h = 4
ARMA	9.5203***	16.7311***	18.5564***	20.0584***	9.5203***	16.7311***	18.5564***	20.0584***
FM1	4.9172***	5.8518***	6.5147**	6.7372***	5.0423***	5.9473***	6.5682***	6.7749***
FM2	5.0715***	5.8099***	6.7179***	8.1984***	5.1795***	5.9147***	6.7597***	8.2218***
FM3	5.1721***	5.8951***	6.9611***	8.4422***	5.2792***	5.9792***	7.0146***	8.4620***
FM4	5.3333***	6.0396***	7.1901***	8.6615***	5.3333***	6.0396***	7.1901***	8.6615***
LASSO1	0.6455***	0.7017***	0.7613***	0.8332***	0.6614***	0.7864***	0.8207***	1.0336***
LASSO2	0.7778***	1.0629***	1.1532***	1.2622***	0.6844***	0.8448***	0.8991***	1.0093***
LASSO3	0.6594***	0.9342***	1.0136***	1.1094***	0.6962***	0.9621***	1.0517***	1.1483***
LASSO4	0.8564***	1.3453***	1.4595***	1.5975***	0.8564***	1.3453***	1.4595***	1.5975***
RF1	0.4359	0.6776***	0.7161***	0.8406***	0.4708	0.7752***	0.8179***	0.9265***
RF2	0.4591	0.5727***	0.7945***	0.9457***	0.4824	0.6832***	0.7225***	1.1748***
RF3	0.4933	0.6627***	0.7390***	0.8564***	0.5026	0.6584***	0.8925***	1.0644***
RF4	0.5524**	0.6832***	0.7985***	1.1748***	0.5524**	0.6832***	0.7985***	1.1748***
SVM1	0.4404	0.5472**	0.7406***	1.0012***	0.4634	0.5534**	0.7902***	0.9431***
SVM2	0.4892	0.6563***	0.8178***	0.8609***	0.4949	0.7499***	0.8495***	1.0112***
SVM3	0.4939	0.6829***	0.8320***	0.9143***	0.5042	0.7794***	0.9307***	1.1990***
SVM4	0.5361**	0.7001***	0.8917***	1.0632***	0.5361**	0.7001***	0.8917***	1.0632***
Focus	0.4670	0.5469	0.6321	0.7444	0.4670	0.5469	0.6321	0.7444

Table 5 illustrates that, for the IPCA forecast, the LASSO model stood out, both with the sentiment scores of the CM and for those of the IR. For inflation, only this model managed not to reject the null hypothesis of the DM test of equality with Focus, the other models rejected the hypothesis. As for GDP forecasts, the RF and SVM models performed better, even some of them managed to surpass the Focus accuracy, such as the RF1 and RF2 with the sentiment scores of the CM and SVM1 in both documents. All these models use polarity series, so for output growth it was only possible to surpass the market average with the inclusion of sentiment scores in the forecasts.

Although Medeiros et al. (2019) and Araujo and Gaglianone (2020) were also successful in overcoming the Focus with IPCA forecasts for one period ahead without using sending rates, it is noteworthy that the authors used a monthly frequency and this paper used a quarterly frequency. In addition, as noted in this paper, the inclusion of textual information from the Copom Minutes and the Inflation Report was only essential to overcome the Focus in GDP growth forecasts. For the IPCA forecasts, textual information was able to improve the performance of our models, but it did not make it possible to beat the market average. This fact is not strange, as as we mentioned, it is easier to obtain fewer forecast errors with inflation than with GDP, so with the IPCA there is less room for improvement in performance using sentiment indices.

Additionally, we check whether the best forecasting models illustrated above are statistically different from each other. Then, as can be seen in Table 8 and Table 9, we again performed the DM test for the three best forecast models for each forecast horizon. In general it is possible to see that the null hypothesis of equality of accuracy between the models cannot be rejected. So, there is no statistical difference in the performance of the best models for each horizon, report and macroeconomic variable.

4.2 Forecast efficiency

We compared the market average forecasts for the IPCA and GDP with qualitative information from the CM and IR texts, regressing Focus's short-term forecast errors into sentiment scores and Focus's forecast errors one quarter ahead with sentiment scores, using an extended version of a Mincer and Zarnowitz (1969) regression. If the Focus forecasts contain all the relevant information that is included in the text of the CM and IR, then the forecast errors should not be predicted by the textual scores.

Table 7 – Mincer–Zarnowitz regression for IPCA

Forecast error in (t)				
	Copom_Index	IR_Index	IPCA_CM_Index	IPCA_IR_Index
Constant	-0.024	-0.037	-0.005	-0.015
	(0.016)	(0.030)	(0.017)	(0.018)
Coefficient of	0.054	0.070	0.024	0.015
Sentiment Score	(0.046)	(0.077)	(0.050)	(0.059)
R^2	0.023	0.013	0.040	0.011
Forecast error in (t+1)				
	Copom_Index	IR_Index	IPCA_CM_Index	IPCA_IR_Index
Constant	-0.013	-0.016	-0.009	-0.013
	(0.016)	(0.031)	(0.017)	(0.018)
Coefficient of	0.005	0.012	0.011	0.006
Sentiment Scores	(0.047)	(0.080)	(0.051)	(0.060)
R^2	0.018	0.038	0.084	0.165

Table 8 – Mincer–Zarnowitz regression for GDP

Forecast error in (t)				
	CM_Index	IR_Index	GDP_CM_Index	GDP_IR_Index
Constant	-0.243	-0.497	-0.088	-0.388**
	(0.186)	(0.361)	(0.181)	(0.130)
Coefficient of	0.624	0.281	1.367*	1.577**
Sentiment Scores	(0.533)	(0.908)	0.564	(0.553)
R^2	0.022	0.016	0.089	0.079
Forecast error in (t+1)				
	CM_Index	IR_Index	GDP_CM_Index	GDP_IR_Index
Constant	-0.330	-0.368	-0.079	-0.389**
	(0.189)	(0.369)	(0.198)	(0.122)
Coefficient of	0.274	0.076	1.345*	2.604***
Sentiment Scores	(0.545)	(0.940)	(0.623)	(0.650)
R^2	0.043	0.011	0.073	0.214

The regression results show that all sentiment index coefficients do not reject the null hypothesis that they are zero for IPCA forecast errors, that is, the textual information in the CM and IR minutes does not affect the forecast errors of the Focus on inflation. So, there is no evidence that the textual information in these documents can improve the inflation forecasts carried out by Focus. The same was observed for the IPCA forecasts by Focus for one quarter ahead. These results are in line with what was found by Clements and Reade (2020) which also did not find a statistically significant coefficient of the sentiment score in explaining the inflation forecast errors of the Central Bank of England.

As for GDP growth forecasts, we found evidence that textual information contains relevant information that can improve Focus forecasts, as forecast errors in real time and for one quarter ahead of GDP are explained by sentiment scores. However, this result is only valid for the variant dictionary sentiment scores (GDP_CM_Index and GDP_IR_Index) and the highest and most significant coefficients were those of GDP_IR_Index. Jones, Sinclair and Stekler (2019) found statistical significance of the sentiment index coefficient only for real-time forecast errors, while forecast errors for one quarter ahead of output growth were not related to the sentiment scores. On the other hand, Clements and Reade (2020) obtained similar results in terms of statistical significance of the coefficient of the sentiment scores, both for the forecast error in the same polarity period and for the forecast errors in *t+1*. As in this article, Clements and Reade (2020) found positive signs of the coefficients, while Jones, Sinclair and Stekler (2019) estimated negative signs.

As forecast errors are defined as less-predicted realized values, this means that Focus' forecasts over this period were overly pessimistic, and that the more positive tone of the text is a more accurate assessment that provides a reduction in forecast errors.

Therefore, these results show that taking into account the textual information in the CM and IR is useful to obtain a more complete picture of current and future economic conditions.

5 Conclusions

The purpose of this paper was to verify whether the textual information contained in the Copom Minutes and in the Inflation Report published by the Central Bank of Brazil are capable of improving the forecasts of macroeconomic variables, specifically, the inflation rate and the GDP growth.

From the results obtained, we can conclude some points. We found that the use of a textual analysis approach via machine learning to select the words that will make up the dictionary produces sentiment scores that more realistically capture the sentiment existing in the texts of the BCB publications, in addition to producing indices more correlated with the macroeconomic variables. The correlations between sentiment indices and macroeconomic variables reveal that scores, in this case those of the variant dictionary, which in addition to showing high correlations in real time, also have a high correlation with macroeconomic variables for one quarter ahead. This reveals that the textual information in the BCB documents may contain relevant information about the future of the IPCA and, mainly, about GDP growth.

It was also possible to verify that machine learning models provide a high degree of accuracy compared to traditional models such as ARMA and factor models. Another interesting point is that the models that incorporated sentiment scores as predictors had the lowest MSEs and that the best models were those that used both variant dictionary and fixed dictionary indices, followed by models that used dictionary indices variant and the worst models were those that did not consider any sentiment variable in their set of predictors. In addition, the Diebold-Mariano test illustrated that short-term forecasts for one quarter ahead of machine learning models were close to the forecasts carried out by Focus, and for GDP growth, some of our forecasts were able to surpass the Focus.

A revealing result that was obtained was that the variant dictionary scores from both the CM and the IR are able to explain the Focus forecast errors in real time for GDP growth. The same was found for forecast errors for one quarter ahead. This fact means that the text of the CM and IR can reduce the Focus forecast error and, in this case, improve the Focus forecasts. This analysis also points out that the coefficient of sentiment scores was positive, indicating that Focus' forecasts for GDP were excessively pessimistic during the sample's time window. As for the IPCA, the same result was not found because the estimated coefficients are not statistically significant. These results make sense, as Focus had lower MSE in the forecasts for the IPCA and higher for the GDP. Furthermore, we were only able to overcome the Focus with the use of sentiment indices on GDP growth.

Finally, we can conclude that the textual information in the CM and IR managed to improve the forecasts of macroeconomic variables, especially forecasts for the GDP, and that it is important for the market to consider such textual information in its short-term forecasts. Furthermore, such information can be useful to show the future conditions of the economy.

Declaration of interest

The authors declare that they have no conflict of interest.

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Bibliography

APEL, M.; GRIMALDI, M. The information content of central bank minutes. *Riksbank Research Paper Series*, n. 92, 2012.

ARAUJO, G. S.; GAGLIANONE, W. P. Machine learning methods for inflation forecasting in brazil: new contenders versus classical models. 2020.

BAI, J.; NG, S. Determining the number of factors in approximate factor models. *Econometrica*, Wiley Online Library, v. 70, n. 1, p. 191–221, 2002.

BAI, J.; NG, S. Forecasting economic time series using targeted predictors. *Journal of Econometrics*, Elsevier, v. 146, n. 2, p. 304–317, 2008.

BAŃBURA, M.; GIANNONE, D.; REICHLIN, L. Large bayesian vector auto regressions. *Journal of applied Econometrics*, Wiley Online Library, v. 25, n. 1, p. 71–92, 2010.

BARBOSA, R. B.; FERREIRA, R. T.; SILVA, T. M. d. Previsão de variáveis macroeconômicas brasileiras usando modelos de séries temporais de alta dimensão. *Estudos Econômicos (São Paulo)*, SciELO Brasil, v. 50, n. 1, p. 67–98, 2020.

BOX, G. E. et al. *Time series analysis: forecasting and control.* [S.l.]: John Wiley & Sons, 2015.

BREIMAN, L. Bagging predictors. *Machine learning*, Springer, v. 24, n. 2, p. 123–140, 1996.

BREIMAN, L. Random forests. *Machine learning*, Springer, v. 45, n. 1, p. 5–32, 2001.

BÜHLMANN, P.; YU, B. et al. Analyzing bagging. *The Annals of Statistics*, Institute of Mathematical Statistics, v. 30, n. 4, p. 927–961, 2002.

CARRIERO, A.; KAPETANIOS, G.; MARCELLINO, M. Forecasting large datasets with bayesian reduced rank multivariate models. *Journal of Applied Econometrics*, Wiley Online Library, v. 26, n. 5, p. 735–761, 2011.

CATALFAMO, E. et al. French nowcasts of the us economy during the great recession: A textual analysis. *Research Program on Forecasting Working Paper*, n. 2018-001, 2018.

CLARK, T. E.; MCCRACKEN, M. W. Testing for unconditional predictive ability. *Federal Reserve Bank of St. Louis Working Paper No*, 2010.

CLEMENTS, M. P.; READE, J. J. Forecasting and forecast narratives: The bank of england inflation reports. *International Journal of Forecasting*, Elsevier, 2020.

DIEBOLD, F. X. The past, present, and future of macroeconomic forecasting. *Journal of economic Perspectives*, v. 12, n. 2, p. 175–192, 1998.

DIEBOLD, F. X.; MARIANO, R. S. Comparing predictive accuracy. *Journal of Business & economic statistics*, Taylor & Francis, v. 20, n. 1, p. 134–144, 2002.

DOSSANI, A. Central bank tone and currency risk premia. *Available at SSRN 3304785*, 2019.

EFRON, B. et al. Least angle regression. *The Annals of statistics*, Institute of Mathematical Statistics, v. 32, n. 2, p. 407–499, 2004.

EICKMEIER, S.; ZIEGLER, C. How successful are dynamic factor models at forecasting output and inflation? a meta-analytic approach. *Journal of Forecasting*, Wiley Online Library, v. 27, n. 3, p. 237–265, 2008.

FILHO, A. E. C.; ROCHA, F. Como o mercado de juros futuros reage à comunicação do banco central? *Economia aplicada*, SciELO Brasil, v. 14, n. 3, p. 265–292, 2010.

FORNI, M. et al. The generalized dynamic-factor model: Identification and estimation. *Review of Economics and statistics*, MIT Press, v. 82, n. 4, p. 540–554, 2000.

FORNI, M. et al. Do financial variables help forecasting inflation and real activity in the euro area? *Journal of Monetary Economics*, Elsevier, v. 50, n. 6, p. 1243–1255, 2003.

GARCIA, M. G.; MEDEIROS, M. C.; VASCONCELOS, G. F. Real-time inflation forecasting with high-dimensional models: The case of brazil. *International Journal of Forecasting*, Elsevier, v. 33, n. 3, p. 679–693, 2017.

GENTZKOW, M.; KELLY, B.; TADDY, M. Text as data. *Journal of Economic Literature*, v. 57, n. 3, p. 535–74, 2019.

GEWEKE, J. The dynamic factor analysis of economic time series. *Latent variables in socio-economic models*, North-Holland, 1977.

GIACOMINI, R. Testing conditional predictive ability. In: *The Oxford Handbook of Economic Forecasting*. [S.l.: s.n.], 2010.

HUBERT, P.; LABONDANCE, F. Central bank sentiment. *URL: https://www. nbp. pl/badania/seminaria/14xi2018. pdf. Working paper*, 2018.

JONES, J. T.; SINCLAIR, T. M.; STEKLER, H. O. A textual analysis of bank of england growth forecasts. *International Journal of Forecasting*, Elsevier, 2019.

KADIYALA, K. R.; KARLSSON, S. Numerical methods for estimation and inference in bayesian var-models. *Journal of Applied Econometrics*, Wiley Online Library, v. 12, n. 2, p. 99–132, 1997.

KOOP, G. M. Forecasting with medium and large bayesian vars. *Journal of Applied Econometrics*, Wiley Online Library, v. 28, n. 2, p. 177–203, 2013.

LEE, T.-H.; YANG, Y. Bagging binary and quantile predictors for time series. *Journal of econometrics*, Elsevier, v. 135, n. 1-2, p. 465–497, 2006.

LIMA, L. R.; GODEIRO, L.; MOHSIN, M. Time-varying dictionary and the predictive power of fed minutes. *Available at SSRN 3312483*, 2019.

LOUGHRAN, T.; MCDONALD, B. When is a liability not a liability? textual analysis, dictionaries, and 10-ks. *The Journal of Finance*, Wiley Online Library, v. 66, n. 1, p. 35–65, 2011.

LOUGHRAN, T.; MCDONALD, B. Textual analysis in accounting and finance: A survey. *Journal of Accounting Research*, Wiley Online Library, v. 54, n. 4, p. 1187–1230, 2016.

MATHY, G.; STEKLER, H. Was the deflation of the depression anticipated? an inference using real-time data. *Journal of Economic Methodology*, Taylor & Francis, v. 25, n. 2, p. 117–125, 2018.

MEDEIROS, M. C.; MENDES, E. F. 1-regularization of high-dimensional time-series models with non-gaussian and heteroskedastic errors. *Journal of econometrics*, Elsevier, v. 191, n. 1, p. 255–271, 2016.

MEDEIROS, M. C. et al. Forecasting inflation in a data-rich environment: the benefits of machine learning methods. *Journal of Business & Economic Statistics*, Taylor & Francis, p. 1–22, 2019.

MEDHAT, W.; HASSAN, A.; KORASHY, H. Sentiment analysis algorithms and applications: A survey. *Ain Shams engineering journal*, Elsevier, v. 5, n. 4, p. 1093–1113, 2014.

MEINSHAUSEN, N.; YU, B. et al. Lasso-type recovery of sparse representations for high-dimensional data. *The annals of statistics*, Institute of Mathematical Statistics, v. 37, n. 1, p. 246–270, 2009.

MINCER, J. A.; ZARNOWITZ, V. The evaluation of economic forecasts. In: *Economic forecasts and expectations: Analysis of forecasting behavior and performance*. [S.l.]: NBER, 1969. p. 3–46.

MOL, C. D.; GIANNONE, D.; REICHLIN, L. Forecasting using a large number of predictors: Is bayesian shrinkage a valid alternative to principal components? *Journal of Econometrics*, Elsevier, v. 146, n. 2, p. 318–328, 2008.

SARGENT, T. J.; SIMS, C. A. et al. Business cycle modeling without pretending to have too much a priori economic theory. *New methods in business cycle research*, v. 1, p. 145–168, 1977.

SHAPIRO, A. H.; SUDHOF, M.; WILSON, D. J. Measuring news sentiment. *Journal of Econometrics*, Elsevier, 2020.

SHIN, K.-S.; LEE, T. S.; KIM, H.-j. An application of support vector machines in bankruptcy prediction model. *Expert systems with applications*, Elsevier, v. 28, n. 1, p. 127–135, 2005.

SILVA, P. H. N. et al. Mensurando o sentimento de incerteza da política econômica: uma análise a partir da comunicação do banco central do brasil. Universidade Federal da Paraíba, 2020.

SIMS, C. A. Macroeconomics and reality. *Econometrica: journal of the Econometric Society*, JSTOR, p. 1–48, 1980.

STEKLER, H.; SYMINGTON, H. Evaluating qualitative forecasts: The fomc minutes, 2006–2010. *International Journal of Forecasting*, Elsevier, v. 32, n. 2, p. 559–570, 2016.

STOCK, J. H.; WATSON, M. W. Forecasting using principal components from a large number of predictors. *Journal of the American statistical association*, Taylor & Francis, v. 97, n. 460, p. 1167–1179, 2002.

TIBSHIRANI, R. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, Wiley Online Library, v. 58, n. 1, p. 267–288, 1996.

VAPNIK, V. *The nature of statistical learning theory.* [S.l.]: Springer science & business media, 1995.

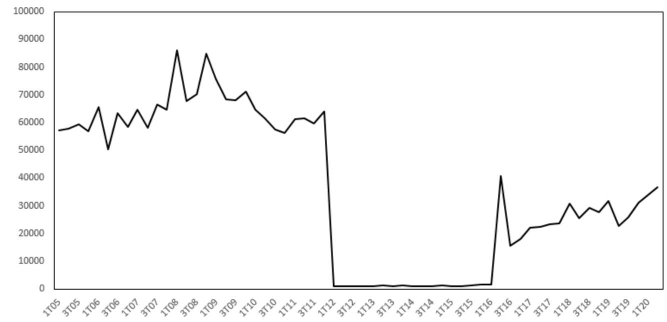
VAPNIK, V. N. An overview of statistical learning theory. *IEEE transactions on neural networks*, IEEE, v. 10, n. 5, p. 988–999, 1999.

ZHAO, P.; YU, B. On model selection consistency of lasso. *Journal of Machine learning research*, v. 7, n. Nov, p. 2541–2563, 2006.

APPENDIX A – Textual Analysis Results

Figure 5 – Number of words - Copom Minutes and Inflation Report

Copom Minutes



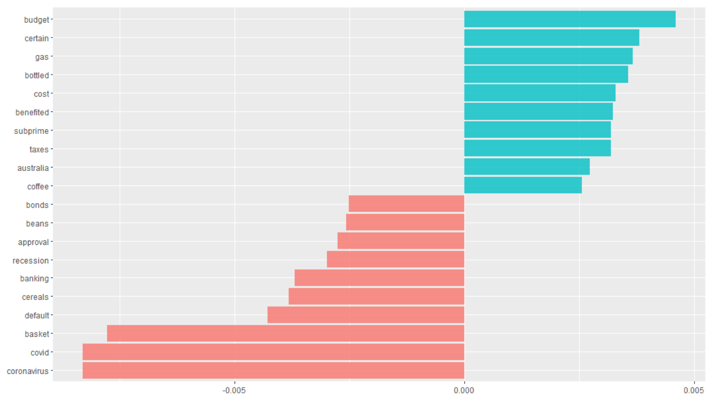
Inflation Report

Figure 6 – Word cloud - Copom Minutes and Inflation Report

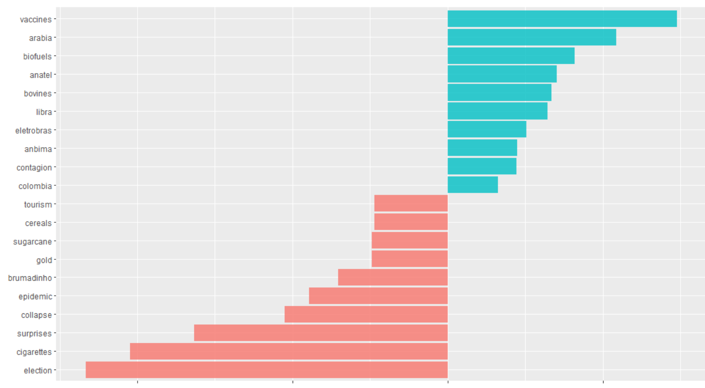
Copom Minutes

Inflation Report

Figure 7 – The 10 most positive and negative coefficients for IPCA - Copom Minutes and Inflation Report

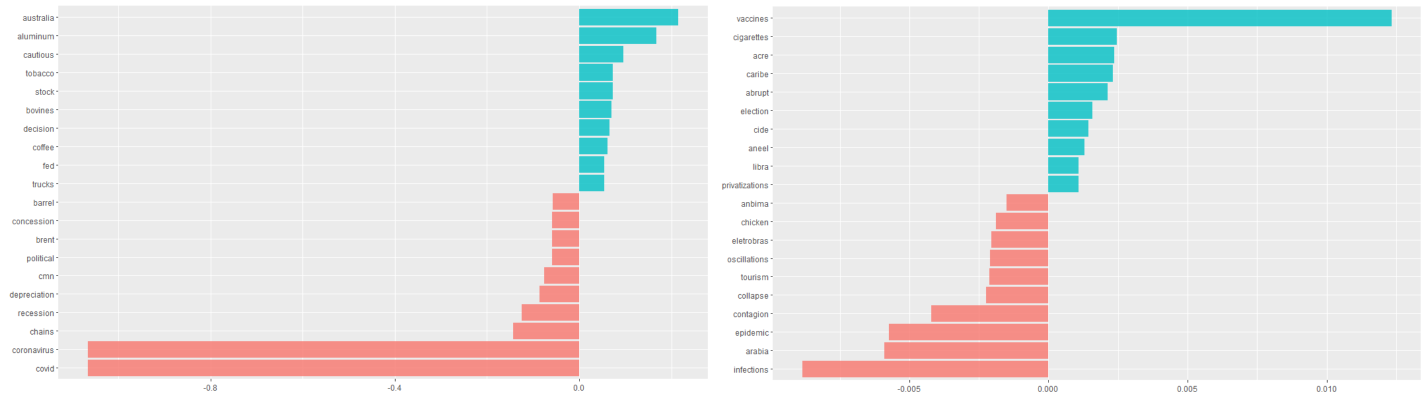


Copom Minutes



Inflation Report

Figure 8 – The 10 most positive and negative coefficients for GDP - Copom Minutes and Inflation Report



Copom Minutes Inflation Report

APPENDIX B – Estimation Results

Table 9 – Correlation between sentiment scores and macroeconomic variables

	$IPCA_t$	$IPCA_{t+1}$	PIB_t	PIB_{t+1}
CM_Index	0.17	-0.08	-0.15	-0.39
IR_Index	0.15	-0.05	0.18	0.13
IPCA_CM_Index	0.45	0.52	-	-
IPCA_IR_Index	0.35	0.57	-	-
GDP_CM_Index	-	-	0.75	0.75
GDP_IR_Index	-	-	0.41	0.71

Table 10 – Diebold-Mariano test for the three best IPCA prediction models

Ata do Copom				Relatório da Inflação			
h = 1	LASSO1	LASSO2	LASSO3	h = 1	LASSO2	LASSO1	LASSO3
LASSO1	-	0.9803	0.3423	LASSO2	-	0.6823	0.3095
LASSO2	0.9803	-	0.3206	LASSO1	0.6823	-	0.588
LASSO3	0.3423	0.3206	-	LASSO3	0.3095	0.588	-
h = 2	LASSO1	SVM2	LASSO2	h = 2	LASSO1	SVM1	SVM2
LASSO1	-	0.2543	0.2296	LASSO1	-	0.7853	0.5341
SVM2	0.2543	-	0.9383	SVM1	0.7853	-	0.8787
LASSO2	0.2296	0.9383	-	SVM2	0.5341	0.8787	-
h = 3	LASSO1	LASSO2	SVM2	h = 3	LASSO1	LASSO2	LASSO3
LASSO1	-	0.1294	0.0623	LASSO1	-	0.9004	0.6923
LASSO2	0.1294	-	0.4732	LASSO2	0.9004	-	0.7299
SVM2	0.0623	0.4732	-	LASSO3	0.6923	0.7299	-
h = 4	LASSO1	LASSO2	SVM2	h = 4	LASSO2	LASSO1	LASSO3
LASSO1	-	0.1113	0.0023	LASSO2	-	0.6811	0.6166
LASSO2	0.1113	-	0.253	LASSO1	0.6811	-	0.9177
SVM2	0.0023	0.253	-	LASSO3	0.6166	0.9177	-

Table 11 – Diebold-Mariano test for the three best GDP prediction models

Ata do Copom				Relatório da Inflação			
h = 1	RF1	SVM1	RF2	h = 1	SVM1	RF1	RF2
RF1	-	0.9793	0.7632	SVM1	-	0.8994	0.7471
SVM1	0.9793	-	0.8182	RF1	0.8994	-	0.8802
RF2	0.7632	0.8182	-	RF2	0.7471	0.8802	-
h = 2	SVM1	RF2	SVM2	h = 2	SVM1	RF3	SVM4
SVM1	-	0.7871	0.0020	SVM1	-	0.0114	0.0000
RF2	0.7871	-	0.0398	RF3	0.0114	-	0.3612
SVM2	0.0020	0.0398	-	SVM4	0.0000	0.3612	-
h = 3	RF1	RF3	SVM1	h = 3	RF2	SVM1	RF4
RF1	-	0.8509	0.8033	RF2	-	0.0555	0.0398
RF3	0.8509	-	0.9596	SVM1	0.0555	-	0.9932
SVM1	0.8033	0.9596	-	RF4	0.0398	0.9932	-
h = 4	LASSO1	RF1	RF3	h = 4	RF1	SVM1	LASSO2
LASSO1	-	0.9625	0.8472	RF1	-	0.8733	0.2154
RF1	0.9625	-	0.9483	SVM1	0.8733	-	0.2994
RF3	0.8472	0.9483	-	LASSO2	0.2154	0.2994	-