

# Narrativas do Banco Central e Previsões Macroeconômicas: Usando Análise Textual de *Machine Learning*

Diego Pitta de Jesus\*

Dr. Cássio da Nóbrega Besarria†

## Resumo

O presente trabalho tem como objetivo principal construir preditores a partir de análise de sentimentos da Ata do Copom produzida pelo Banco Central do Brasil. Este artigo busca verificar se as narrativas nas atas contêm informações úteis que podem ser usadas para melhorar a precisão das previsões de indicadores macroeconômicos para um período à frente. O primeiro passo será construir séries temporais que representem a polaridade (sentimentos) nos textos da ata do Copom por meio de métodos de análise textual. Após a obtenção das séries de polaridade, iremos testar o seu poder preditivo para os indicadores macroeconômicos para um período a frente. Também vamos verificar se esses novos preditores são capazes melhorar o desempenho de modelos *machine learning*.

**Palavras-chave:** Previsão macroeconômica. *Machine learning*. Análise textual. Ata do Copom.

## 1 Introdução

A previsão de variáveis macroeconômicas, em particular indicadores-chave como crescimento do PIB, inflação e taxas de juros, são insumos fundamentais para o planejamento orçamentário do governo, a formulação de políticas do banco central e as decisões dos empresários. O uso de abordagens de séries temporais para previsão macroeconômica ganhou impulso nas décadas de 1970 e 1980, pois as previsões dos modelos univariados ARIMA (Box et al., 2015) e vetores autoregressivos (VAR) (Sims, 1980) mostraram desempenho superior aos modelos macroeconômicos estruturais<sup>1</sup>. Durante essa época, os conjuntos de informações usados para formar previsões geralmente continham apenas um pequeno número de variáveis. Entretanto, tais métodos possuem uma limitação importante, em que esses modelos suportam apenas um pequeno número de preditores. Essa situação mudou no início dos anos 2000, quando os pesquisadores começaram a usar dados macroeconômicos de alta dimensão. Nesse cenário, modelos capazes de lidar com um grande número de preditores começaram a ganhar destaque. Então, a literatura de previsão macroeconômica passou a utilizar com maior frequência os modelos de fatores e os modelos de encolhimento de *machine learning*.

Dois exemplos que podem ser encontrados na literatura são o conjunto de dados dos EUA que contém 149 variáveis medidas com uma frequência mensal apresentada em Stock and Watson (2002) e o conjunto de dados da área do euro contendo 447 variáveis mensuradas com frequência mensal apresentadas Forni et al. (2003). Nos dois estudos, a utilização de um grande número de preditores em uma estrutura de modelagem de fatores dinâmicos (ver Geweke (1977); Sargent, Sims et al. (1977)) apresentou um melhor desempenho nas previsões da produção industrial em relação aos modelos tradicionais de referência. Um fator importante na popularidade dessa abordagem é sua simplicidade, em que os componentes principais fornecem estimativas consistentes dos fatores dinâmicos e podem

---

\*Doutorando em Economia pelo PPGE/UFPB [diegopitta13@hotmail.com](mailto:diegopitta13@hotmail.com)

†Orientador e docente do PPGE/UFPB [cassiodanobrega@yahoo.com.br](mailto:cassiodanobrega@yahoo.com.br)

<sup>1</sup> Para uma discussão mais aprofundada, ver Diebold (1998) e suas referências

subsequentemente ser usados em regressões preditivas auxiliares. Existe uma extensa literatura que mostra que quando o modelo de fatores é usado com um grande número de preditores, produz boas previsões para variáveis macroeconômicas, como PIB e inflação, para várias economias diferentes (Eickmeier and Ziegler, 2008).

Apesar de seu sucesso, o modelo de fatores dinâmicos não é a única estrutura para previsão com um grande número de preditores. Os avanços da estatística e da literatura de aprendizado de máquina também foram explorados no contexto macroeconômico. Por exemplo, Mol, Giannone and Reichlin (2008) consideram a regressão de ridge e de *least absolute shrinkage and selection operator* LASSO (Tibshirani, 1996) para os dados de Stock and Watson (2002) e obtiveram previsões com desempenho semelhante ao obtido em com o modelo de fatores dinâmicos. Bai and Ng (2008) usam regressão LARS (Efron et al., 2004) para selecionar um conjunto de preditores. As previsões foram produzidas usando esses preditores selecionados pelo modelo. Bai and Ng (2008) mostram que, pelo menos em alguns períodos dos dados, os métodos baseados no LARS produzem melhores previsões de inflação, renda, vendas no varejo, produção industrial e emprego total em comparação com o modelo de fatores principais.

Também surgiram na literatura trabalhos que usam métodos que respondem pela incerteza do modelo, como agregação de *bootstrap* ou "bagging" (ver Breiman (1996); Bühlmann, Yu et al. (2002); Lee and Yang (2006)) foram bem-sucedidos na previsão da inflação por Inoue and Kilian (2008). Finalmente, na classe de previsão multivariada, houve um foco em VARs "grandes" estimados usando técnicas bayesianas. Exemplos incluem Kadiyala and Karlsson (1997) e, mais recentemente, Bańbura, Giannone and Reichlin (2010), Carriero, Kapetanios and Marcellino (2011) e Koop (2013) que utilizam priores de encolhimento.

No caso do Brasil, nos últimos anos, surgiu um crescente corpo de literatura sobre previsão macroeconômica com destaque para os modelos de *machine learning*. Medeiros and Mendes (2016) consideraram diferentes modelos de alta dimensão para prever a inflação brasileira. Os autores mostraram que as técnicas baseadas no *least absolute shrinkage and selection operator* (LASSO) apresentam os menores erros de previsão para previsões de horizonte curto. Para horizontes mais longos, o *benchmark* de AR é o melhor modelo para previsão de pontos, mesmo que não haja diferenças significativas entre eles. Modelos fatoriais também produzem boas previsões de longo horizonte em alguns casos. Mais recentemente, Garcia, Medeiros and Vasconcelos (2017) e Medeiros et al. (2019) usaram modelos de alta dimensão para prever a inflação em tempo real do Brasil e mostraram que o desempenho dos modelos de encolhimento é superior em relação às técnicas mais tradicionais.

Barbosa, Ferreira and Silva (2020) analisaram o desempenho de modelos fatoriais de alta dimensão para prever quatro variáveis macroeconômicas brasileiras: duas variáveis reais, taxa de desemprego e o índice de produção industrial, e duas variáveis nominais, IPCA e IPC. Os autores usaram três tipos de técnicas de aprendizado estatístico foram aplicados: métodos de *shrinkage*, combinações de previsões e seleção de previsores. Os fatores foram extraídos de forma supervisionada e não supervisionada. Os resultados indicaram que métodos de aprendizado estatístico melhoram o desempenho preditivo das variáveis econômicas brasileiras.

Araujo and Gaglianone (2020) realizaram um exercício de previsão fora da amostra, através de uma variedade de técnicas de aprendizado de máquina e modelos econométricos tradicionais. Os resultados encontrados pelos autores corroboram conclusões recentes a favor dos procedimentos automatizados não-lineares, indicando que algoritmos de *machine learning* (em particular, *Random Forest*) podem superar os métodos tradicionais de previsão em termos de erro quadrático médio.

Contudo, surgiram alguns trabalhos na literatura internacional de previsão macroeconômica que voltaram o seu foco para a construção de índices de sentimentos<sup>2</sup> a partir das informações textuais contidas nos relatórios publicados pelo Banco Central da Inglaterra (Jones, Sinclair and Stekler,

<sup>2</sup> De acordo com Medhat, Hassan and Korashy (2014) a análise de sentimento é o estudo computacional das opiniões, atitudes e emoções contidas em textos escritos. Em geral, a mineração de opinião ajuda a coletar informações sobre os aspectos positivos e negativos de um tópico específico.

2019);(Clements and Reade, 2020), que são usados como preditores para a taxa de inflação, PIB e outros indicadores macroeconômicos. Esses indicadores de polaridade do Banco Central podem ser usados como preditores diretos para as variáveis macroeconômicas do próximo período. Também podem ser utilizados em modelos mais gerais (por exemplo, modelos VAR, modelos de *machine learning*, modelos de *deep learning* e etc) para melhorar o desempenho de tais modelos de previsão.

De acordo com Jones, Sinclair and Stekler (2019) a avaliação das previsões econômicas concentrou-se tipicamente na qualidade das previsões numéricas. No entanto, esse foco nas previsões quantitativas negligencia a quantidade substancial de texto que freqüentemente as acompanha, principalmente nas publicações dos bancos centrais. O texto é incluído para fornecer contexto e nuances e pode ser avaliado usando análise textual. Além disso, as previsões numéricas nem sempre estão disponíveis ao público, mas o texto dos relatórios, as atas das reuniões e os discursos podem revelar informações sobre a avaliação das condições econômicas atuais e futuras.

A literatura estabeleceu o valor da análise textual, bem como uma metodologia geral para converter texto em *scores* quantitativos que avaliam principalmente as polaridades dos textos. De acordo com Gentzkow, Kelly and Taddy (2019), as informações codificadas no texto são um complemento rico para os tipos de dados mais estruturados tradicionalmente usados na pesquisa empírica. De fato, nos últimos anos, ocorreu um uso intenso de dados textuais em diferentes áreas de pesquisa.

Existe um corpo de literatura em rápida expansão sobre o uso de informações textuais, como as narrativas sutis do tipo que aparecem nos relatórios de inflação e atas dos bancos centrais. Um artigo-chave é o de Stekler and Symington (2016) que investiga as "narrativas" que constituem as atas do *Federal Open Market* (FOMC) entre 2006 e 2010. Seu estudo quantifica as declarações qualitativas das atas do FOMC sobre o atual e o futuro tendências da economia e compara os índices resultantes às previsões do *Greenbook* e da Pesquisa de Previsões Profissionais dos EUA.

Dossani (2019) analisa como o tom das entrevistas coletivas do Banco Central dos Estados Unidos afeta os prêmios de risco no mercado de câmbio. Ele mede o tom como a diferença entre o número de frases *hawkish* e *dovish* feitas durante uma conferência de imprensa. Ele usou quatro contratos futuros de moeda negociados na *Chicago Mercantile Exchange* (CME) e descobriu que a aversão ao risco implícita aumenta quando os Bancos Centrais são *hawkish* e diminui quando os Bancos Centrais estão *dovish*.

Goldfarb, Stekler and David (2005) desenvolveram um procedimento de pontuação que eles usaram para criar um índice de previsões qualitativas feitas durante a Grande Depressão com base em uma análise textual de artigos de jornais contemporâneos. Lundquist and Stekler (2012) aplicaram o procedimento de Goldfarb, Stekler and David (2005) no estudo de avaliações qualitativas feitas por economistas durante a Grande Recessão. Mais recentemente, Stekler and Symington (2016) usaram o procedimento para examinar as atas das reuniões do FOMC, e Mathy and Stekler (2018) ajustaram o procedimento de pontuação para incorporar perspectivas mais negativas. Eles também propuseram uma metodologia de avaliação que envolvia o uso de coeficientes de correlação para comparar as pontuações com uma série de *benchmarks*. Catalfamo et al. (2018) incorporou essas duas inovações em seu exame dos noticiários franceses da economia dos EUA durante a Grande Recessão.

Jones, Sinclair and Stekler (2019) realizaram análises qualitativas via análise de sentimentos dos textos dos relatórios de inflação do Banco Central da Inglaterra no período de 2005-2014. Também construíram índices de polaridade para fora da amostra. Em seguida, compararam as pontuações com os dados de crescimento do produto em tempo real e com as projeções quantitativas correspondentes publicadas pelo Banco Central. Concluíram que a evolução geral da economia do Reino Unido foi representada com precisão no texto do Relatório de Inflação. Além disso, as regressões de eficiência sugeriram que há informações no texto que poderiam melhorar os relatórios quantitativos do Banco da Inglaterra e as previsões para um trimestre à frente.

Clements and Reade (2020) analisaram as narrativas que acompanham as previsões numéricas nos relatórios trimestrais de inflação do Banco da Inglaterra, para o período de 1997–2018. O trabalho se concentrou em saber se as narrativas contêm informações úteis sobre o curso futuro das principais

variáveis macroeconômicas além das previsões pontuais, em termos de se as narrativas podem ser usadas para aprimorar a precisão das previsões numéricas. Também foi considerado se as narrativas são capazes de prever mudanças futuras nas previsões numéricas. Os autores concluíram que uma medida de sentimento derivada das narrativas pode prever os erros nas previsões numéricas de crescimento do produto, mas não da inflação. Não encontraram evidências de que mudanças passadas no sentimento prevejam mudanças subsequentes nas previsões pontuais do crescimento do produto ou da inflação, mas descobriram que os ajustes nas previsões numéricas do crescimento do produto têm um elemento sistemático.

Até o momento, essa literatura para países emergentes como o Brasil é escassa. Portanto, o presente trabalho busca preencher essa lacuna da literatura. Nesse contexto, o trabalho tem como objetivo principal construir índices de sentimentos via análise de sentimentos a partir da Ata do Copom e do Relatório de Inflação (RI) produzidos pelo Banco Central do Brasil (BCB). Este artigo busca verificar se as narrativas na ata do Copom e no RI contêm informações textuais úteis que podem ser usadas para melhorar a precisão das previsões de indicadores macroeconômicos, tais como a taxa de inflação e o crescimento do PIB. Então, para realizar esses objetivos iremos realizar dois exercícios. No primeiro vamos fazer previsões para a taxa de inflação e para o crescimento do PIB via modelos tradicionais e modelos de *machine learning*. Nesse caso, pretendemos saber se a inclusão de índices de sentimentos nesses modelos é capaz de melhorar a acurácia da previsão. Vamos comparar o desempenho das nossas previsões com o desempenho da média do mercado, representada pelo Focus. O segundo exercício vai relacionar diretamente as informações textuais da ata do Copom e do RI com as previsões do Focus. Neste caso, vamos verificar se as informações textuais são capazes de melhorar as previsões do Focus.

Assim, o primeiro passo será construir séries temporais que representem a polaridade (sentimentos) nos textos da ata do Copom e do Relatório de Inflação por métodos de análise textual. Nessa etapa, usamos o tradicional dicionário de [Loughran and McDonald \(2016\)](#) com léxicos fixos, que é amplamente utilizado e difundido nos trabalhos que analisam textos financeiros. Adicionalmente, empregamos a abordagem de [Lima, Godeiro and Mohsin \(2019\)](#) que permite que o conteúdo dicionário seja variante ao longo do tempo.

Após a obtenção das séries de sentimentos textuais, vamos verificar se essas novos indicadores são capazes melhorar o desempenho de modelos de previsão multivariados. Os modelos multivariados utilizados nesse exercício são o modelo de Vetores Auto-Regressivos (VAR), *least absolute shrinkage and selection operator* (LASSO), *Random Forest* (RF) e *Support Vector Machines* (SVM) Como *benchmark* usaremos as previsões fornecidas pelo Boletim Focus. As previsões serão realizadas para cada período à frente fora da amostra ( $\pi_{t+h}$ ) e levará apenas em consideração as informações disponíveis até o momento  $X_t$ . Também, de forma similar a [Jones, Sinclair and Stekler \(2019\)](#), vamos testar a eficiência da previsão do Focus pela versão estendida de equação de [Mincer and Zarnowitz \(1969\)](#). Neste caso, será possível saber se as previsões realizadas pelo Focus podem ser melhoradas com a incorporação de todas as informações e nuances contidas nas narrativas do BCB.

Assim, o presente trabalho possui algumas contribuições. A primeira é incorporar na comparação de desempenho das previsões o uso de um modelo *machine learning* que nenhum trabalho mencionado anteriormente utilizou, neste caso o SVM. A segunda contribuição é usar um método de dicionário variante no tempo, pois todos os trabalhos da literatura citados anteriormente usam dicionários fixos. A terceira é de preencher algumas lacunas na literatura brasileira de previsão macroeconômica por meio de uma abordagem alternativa e inovadora que está na fronteira na literatura mundial. Por fim, a quarta contribuição é mostrar aos agentes acadêmicos e de mercado que os textos dos relatórios produzidos pelas instituições, principalmente os provenientes do Banco Central, contém informações textuais revelantes para o exercício de previsão que não podem ser ignoradas e devem ser incorporadas no arcabouço metodológico dos principais agentes que realizam previsão de variáveis macroeconômicas.

Os resultados obtidos mostram de forma geral que os modelos de previsão que usam as séries

de sentimentos apresentam os menores erros de previsão, e em alguns casos apresentando erros de previsão inferiores ao Focus. As melhores previsões foram obtidas com os modelos que usaram as séries de sentimentos provenientes do dicionário variante no tempo. Esse fato ocorreu porque esse tipo de dicionário é capaz de incorporar novos termos que aparecem nos relatórios. Assim, as séries de sentimentos construídas a partir do dicionário variante conseguiram captar, por exemplo, a pandemia do Covid-19, pois os termos relacionados a pandemia surgiram em 2019 e não constam nos dicionários fixos, então, o sentimento dos relatórios é mais realista com o dicionário variante no tempo.

Por fim, a equação de [Mincer and Zarnowitz \(1969\)](#) para a eficiência das previsões do Focus mostrou que os índices de sentimentos são capazes de prever os erros das previsões do Focus para o PIB, ou seja, na média o mercado para as previsões do PIB não leva em consideração a totalidade das informações e nuances textuais contidas nos relatórios. Assim, as previsões do crescimento do PIB realizadas pelo Focus podem ser melhoradas com os índices de sentimento. Mas, isso não foi constatado para a inflação. Esse fato está em conformidade com os resultados anteriormente encontrados, pois os erros de previsão do Focus para a inflação é bem menor do que para o PIB. Ao mesmo tempo, nossas previsões com sentimentos conseguiram superar o Focus apenas nas previsões realizadas para o PIB.

## 2 Dados

### 2.1 Ata do Copom e Relatório de Inflação

De acordo com [Filho and Rocha \(2010\)](#) A ata do Copom é um dos principais instrumentos de comunicação do Banco Central do Brasil, apresenta projeções econômicas para o cenário nacional e internacional, controle da inflação, decisões a respeito dos juros etc. É através dela que a autoridade monetária explica os procedimentos utilizados para a tomada de decisão de política monetária com o objetivo de tornar a comunicação mais transparente e manter sob controle as expectativas.

Assim como no trabalho de Silva, Besarria e Silva (2019) que também usaram as atas do Copom para um exercício de análise textual, é necessário levar em conta uma mudança referente a periodicidade de publicação, durante o período 2000 a 2005 as reuniões ocorriam de forma mensal assim como as publicações, a partir do ano de 2006 as reuniões passaram a acontecer a cada quarenta e cinco dias, sendo divulgadas oito atas ao ano. Outro ponto que merece ser destacado é o fato de que em 2002 foram publicadas treze atas ao invés de doze. A série de polaridade após ser estimada precisou ser convertida em dados trimestrais com o objetivo de que toda a amostra passasse a ter a mesma frequência.

De acordo com o BCB o Relatório de Inflação (RI) apresenta as diretrizes das políticas adotadas pelo Copom, considerações acerca da evolução recente do cenário econômico e projeções para a inflação. As projeções são apresentadas em cenários com condicionantes para algumas variáveis econômicas. O Copom utiliza um conjunto amplo de modelos e cenários para orientar suas decisões de política monetária. Ao expor alguns desses cenários, o Copom procura dar maior transparência às decisões de política monetária, contribuindo para sua eficácia no controle da inflação, que é seu objetivo principal.

A frequência do RI é trimestral, entretanto, o relatório possui um "gap" entre os anos de 2012 e 2015. Nesses anos o Copom disponibiliza apenas o sumário executivo do RI, que é um resumo do documento. Portanto, optamos por usar o sumário executivo como proxy do RI durante os anos de 2012 e 2015.

Assim, escolhemos para a compor a amostra com janela temporal com início no primeiro trimestre de 2005 até o segundo trimestre de 2020, totalizando 62 observações. Foi possível obter todas as atas do Copom e os Relatórios de Inflação por *web scraping*, em que importamos e trabalhamos com os documentos no formato PDF na versão em língua inglesa.



## 2.2 Variáveis Macroeconômicas

Para o exercício de predição macroeconômica nos modelos de *machine learning* foram utilizadas um grande conjunto de variáveis macroeconômicas e financeiras como preditores. As variáveis de interesse a serem previstas são a taxa inflação e a taxa de crescimento do PIB. Para a taxa de inflação optamos por escolher o índice IPCA que é o índice de inflação oficial adotado pelo Copom para fins de política monetária. A frequência e periodicidade das variáveis são idênticas aos dos relatórios da inflação na seção anterior. A [Tabela 1](#) e a [Tabela 2](#) ilustram todas as variáveis.

Tabela 1: Lista de variáveis macroeconômicas e financeiras

Séries	Categoria	Nome	Fonte	Unidade
1	Inflação	IPCA	IBGE	%
2	Inflação	IPCA - Preços de mercado	IBGE	%
3	Inflação	IPCA - Preços monitorados	IBGE	%
4	Inflação	IPCA - Negociáveis	BCB	%
5	Inflação	IPCA - Não-negociáveis	BCB	%
6	Inflação	IPC-Fipe	Fipe	%
7	Inflação	IPC-Br	FGV	%
8	Inflação	IPA-DI - Atacado	FGV	%
9	Inflação	IGP-DI	FGV	%
10	Inflação	IGP-M	FGV	%
11	Inflação	IGP-10	FGV	%
12	Inflação	INCC	FGV	%
13	Inflação	IPC-Br - Núcleo	FGV	%
14	Inflação	IPCA - Núcleo (EX0 exclusão)	BCB	%
15	Inflação	IPCA - Núcleo (EX1 exclusão)	BCB	%
16	Inflação	IPCA - Núcleo (dupla ponderação)	BCB	%
17	Inflação	IPCA - Núcleo (médias aparadas sem suavização)	BCB	%
18	Inflação	IPCA - Núcleo (Break Even, 1 ano)	Anbima	%
19	Inflação	IPCA - Núcleo (Break Even, 2 ano)	Anbima	%
20	Inflação	IPCA - Núcleo (Break Even, 5 ano)	Anbima	%
21	Taxa de juros	Selic	BCB	%
22	Taxa de juros	TJLP	BCB	%
23	Taxa de juros	Tesouro prefixado (1 ano)	Anbima	%
24	Taxa de juros	Tesouro prefixado (2 ano)	Anbima	%
25	Taxa de juros	Tesouro prefixado (5 ano)	Anbima	%
26	Taxa de juros	Pré-DI Swap (1 ano)	BCB	%
27	Taxa de juros	Pré-DI Swap Focus (1 ano)	BCB	%
28	Taxa de juros	Tesouro IPCA (1 ano)	Anbima	%
29	Taxa de juros	Tesouro IPCA (2 ano)	Anbima	%
30	Taxa de juros	Tesouro IPCA (5 ano)	Anbima	%
31	Moeda	Base monetária	BCB	R\$ mil
32	Moeda	Oferta de moeda	BCB	R\$ mil
33	Moeda	Depósitos (conta corrente)	BCB	R\$ mil
34	Moeda	Depósitos (poupança)	BCB	R\$ mil
35	Moeda	M1	BCB	R\$ mil
36	Moeda	M2	BCB	R\$ mil
37	Moeda	M3	BCB	R\$ mil
38	Moeda	M4	BCB	R\$ mil
39	Setor bancário	Spread de crédito (recursos livres, taxa Selic)	BCB	Pontos base
40	Setor bancário	Índice de inadimplência da carteira de crédito	BCB	%
41	Setor bancário	Proporção dos empréstimos	BCB	Unidade
42	Setor bancário	Proporção das reservas	BCB	Unidade
43	Setor bancário	Crescimento das operações de crédito	BCB	R\$ milhões
44	Mercado de capitais	IPOs (acumulados em 12 meses)	BCB	R\$ milhões
45	Mercado de capitais	Patrimônio líquido dos fundos de ações	BCB	R\$ milhões
46	Mercado de capitais	Patrimônio líquido dos fundos de investimento financeiro	BCB	R\$ milhões
47	Mercado de capitais	Ibovespa	Reuters	Índice
48	Mercado de capitais	MSCI países emergentes (US\$)	Reuters	Índice
49	Mercado de capitais	MSCI países desenvolvimentos (US\$)	Reuters	Índice
50	Prêmio de risco	EMBI	Reuters	Pontos base
51	Prêmio de risco	EMBI (composto - média de 16 países)	Reuters	Pontos base
52	Prêmio de risco	CDS (5 anos)	Reuters	Pontos base
53	Taxa de câmbio	Taxa de câmbio nominal (R\$/US\$)	BCB	Unidades
54	Taxa de câmbio	Taxa de câmbio real efetiva (IPA - 13 moedas)	Funcex	Índice
55	Economia mundial	U.S. dollar index (média geométrica de 6 em relação ao US\$)	Reuters	Índice
56	Economia mundial	U.S. Treasury 2 years (Treasury nominal interest rates)	Reuters	%
57	Economia mundial	U.S. Treasury 10 years (Treasury nominal interest rates)	Reuters	%
58	Economia mundial	U.S. Treasury 5 years TIPS (Treasury Inflation Protected Securities)	Reuters	%
59	Economia mundial	CRB all commodities index	Reuters	Índice
60	Economia mundial	Oil price (WTI, OklahomaUSA)	Reuters	US\$/barril
61	Economia mundial	VIX CBOE volatility index (30 day expected volatility of the S&P500)	Reuters	Índice

Tabela 2: Lista de variáveis macroeconômicas e financeiras - continuação

Séries	Categoria	Nome	Fonte	Unidade
62	Setor externo	Índice de preços de importação	Funcex	Índice
63	Setor externo	Índice de quantidade de importação	Funcex	Índice
64	Setor externo	Índice de preços de exportação	Funcex	Índice
65	Setor externo	Índice de quantidade de exportação	Funcex	Índice
66	Setor externo	Importação (FOB, total)	MDIC/Secex	US\$
67	Setor externo	Exportação (FOB, total)	MDIC/Secex	US\$
68	Setor externo	Exportação (FOB, bens primários)	MDIC/Secex	US\$
69	Setor externo	Reservas internacionais (total)	BCB	US\$ milhões
70	Setor externo	Conta corrente (líquida)	BCB	US\$ milhões
71	Setor externo	Conta corrente (acumulada em 12 meses como proporção do PIB)	BCB	%
72	Setor externo	Investimento estrangeiro direto (acumulado em 12 meses)	BCB	US\$ milhões
73	Setor externo	Investimento estrangeiro em portfólio (acumulado em 12 meses)	BCB	US\$ milhões
74	Atividade econômica	IBC-Br	BCB	Índice
75	Atividade econômica	PIB (acumulado em 12 meses, preços de mercado)	BCB	R\$ milhões
76	Atividade econômica	Índice de confiança do consumidor	Fecomercio	Índice
77	Trabalho	Taxa de desemprego	IBGE	%
78	Trabalho	Índice de empregados registrados (comércio por atacado e varejo)	MTE	Índice
79	Trabalho	Índice de empregados registrados (setor de construção)	MTE	Índice
80	Trabalho	Horas trabalhadas na produção (São Paulo)	Fiesp	Índice
81	Trabalho	Salário real (indústria, São Paulo)	Fiesp	Índice
82	Indústria	Produção industrial (total)	IBGE	Índice
83	Indústria	Produção industrial (extração mineral)	IBGE	Índice
84	Indústria	Produção industrial (indústria manufatureira)	IBGE	Índice
85	Indústria	Produção industrial (bens de capital)	IBGE	Índice
86	Indústria	Produção industrial (bens intermediários)	IBGE	Índice
87	Indústria	Produção industrial (bens de consumo)	IBGE	Índice
88	Indústria	Produção industrial (bens duráveis)	IBGE	Índice
89	Indústria	Produção industrial (bens semi-duráveis e não duráveis)	IBGE	Índice
90	Indústria	Utilização da capacidade instalada (São Paulo)	Fiesp	%
91	Indústria	Utilização da capacidade instalada (indústria manufatureira)	FGV	%
92	Indústria	Produção de aço	BCB	Índice
93	Indústria	Produção de veículos (total)	BCB	Unidades
94	Indústria	Produção de veículos (veículos de passeio e veículos comerciais leves)	BCB	Unidades
95	Indústria	Produção de caminhões	BCB	Unidades
96	Indústria	Produção de ônibus	BCB	Unidades
97	Indústria	Produção de máquinas agrícolas (total)	BCB	Unidades
98	Vendas	Índice de volume de vendas no setor de varejo (total)	BCB	Índice
99	Vendas	Índice de volume de vendas no setor de varejo (combustíveis e lubrificantes)	BCB	Índice
100	Vendas	Índice de volume de vendas no setor de varejo (hiperm. superm. alim. beb. tab.)	BCB	Índice
101	Vendas	Índice de volume de vendas no setor de varejo (têxtil, vestuário e calçados)	BCB	Índice
102	Vendas	Índice de volume de vendas no setor de varejo (móveis e itens brancos)	BCB	Índice
103	Vendas	Índice de volume de vendas no setor de varejo (veículos e motos, peças de reposição)	BCB	Índice
104	Vendas	Índice de volume de vendas no setor de varejo (hipermercados e supermercados)	BCB	Índice
105	Vendas	Vendas de veículos (total)	BCB	Unidades
106	Vendas	Vendas de veículos (domésticos)	BCB	Unidades
107	Energia	Consumo de energia elétrica (comercial)	Eletrobras	GWh
108	Energia	Consumo de energia elétrica (residencial)	Eletrobras	GWh
109	Energia	Consumo de energia elétrica (industrial)	Eletrobras	GWh
110	Energia	Consumo de energia elétrica (outros)	Eletrobras	GWh
111	Energia	Consumo de energia elétrica (total)	Eletrobras	GWh
112	Setor Público	Resultado primário do setor público consolidado (fluxos mensais atuais)	BCB	R\$ milhões
113	Setor Público	Resultado primário do setor público consolidado (fluxos acumulados em 12 meses)	BCB	R\$ milhões
114	Setor Público	Resultado primário do setor público consolidado (fluxos acumulados em 12 meses, % PIB)	BCB	%
115	Setor Público	Dívida pública líquida (total, governo federal e banco central, % PIB)	BCB	%
116	Setor Público	Dívida pública líquida (interna, governo federal e banco central, % PIB)	BCB	%
117	Setor Público	Dívida pública líquida (externa, governo federal e banco central, % PIB)	BCB	%
118	Setor Público	Dívida pública líquida (total, setor público consolidado, saldos em reais)	BCB	R\$ milhões
119	Setor Público	Dívida pública líquida (interna, setor público consolidado, saldos em reais)	BCB	R\$ milhões
120	Setor Público	Dívida pública líquida (externa, setor público consolidado, saldos em reais)	BCB	R\$ milhões

As variáveis possuem frequência trimestral iniciando no primeiro trimestre de 2006 até o quarto trimestre de 2019. Os preditores foram obtidos nos sites do Instituto Brasileiro de Geografia e Estatística (IBGE), Banco Central do Brasil (BCB), Fundação Getúlio Vargas (FGV), Fundação Instituto de Pesquisas Econômicas (Fipe), Associação Brasileira das Entidades dos Mercados Financeiro e de Capitais (Anbima), Reuters, Fundação Centro de Estudos do Comércio Exterior (Funcex), Secretaria de Comércio Exterior (SECEX), Fecomércio, Ministério do Trabalho (MTE), Federação das Indústrias do Estado de São Paulo (Fiesp) e Eletrobrás.

### 3 Metodologia

Esta seção descreve os métodos usados neste artigo para as previsões. Assim, como [Garcia, Medeiros and Vasconcelos \(2017\)](#) consideramos uma abordagem direta de previsão em que a variável a ser prevista está  $h$  períodos à frente,  $y_{t+h}$ , é modelada em função de um conjunto de preditores

medido no tempo  $t$ , como:

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

em que  $T(x_t)$  é um mapeamento de um conjunto de  $q$  preditores,  $u_{t+h}$  é o erro de previsão e  $x_t = (x_{1t}, \dots, x_{qt})' \in \mathbb{X} \subseteq \mathbb{R}^q$  pode incluir preditores fracamente exógenos, valores defasados da variável de interesse e vários fatores calculados a partir de um grande número de covariáveis potenciais. É importante ressaltar que  $x_t$  contém apenas variáveis observadas e disponíveis no momento  $t$ . Observe que a consideração de modelos de previsão direta para cada horizonte evita a necessidade de estimar um modelo para a evolução de  $x_t$ .

Para a maioria dos métodos considerados neste artigo, o mapeamento  $T(\cdot)$  é linear, de modo que:

$$y_{t+h} = \beta' x_t + u_{t+h} \quad (2)$$

em que  $\beta \in \mathbb{R}^q$  é um vetor de parâmetros desconhecidos.

### 3.1 Mensurando os índices de sentimento

Nesta seção será apresentada a metodologia de construção dos índices de sentimentos (S) a partir dos textos das atas do Copom e do RI. Cada  $S_t$  visa capturar algumas das informações da narrativa no relatório no momento  $t$ , para cada ata em nossa amostra. Essa medida transforma milhares de palavras em um único número. Para obter cada série de sentimento  $S_t$  usamos duas abordagens: uma que mensura os sentimentos a partir de dicionários com léxicos fixos e outra que usa modelos de *machine learning* para construir um dicionário variante no tempo.

Antes de executar a análise lexicográfica nos documentos, realizamos uma série de transformações no texto original. O texto é primeiro dividido em uma sequência de *substrings* (*tokens*) cujos caracteres são todos transformados em minúsculas. Removemos *stop words* em inglês e limpamos o texto usando o pacote *tolower* do R.

De acordo com [Shapiro, Sudhof and Wilson \(2020\)](#) existem duas metodologias gerais para quantificar o sentimento no texto. A primeira é conhecida como metodologia lexical. Esta abordagem se baseia em listas predefinidas de palavras, chamadas de léxicos ou dicionários, com cada palavra atribuída uma pontuação para a emoção de interesse. Geralmente, essas pontuações são simplesmente 1, 0 e -1 para positivo, neutro e negativo, mas alguns léxicos têm mais de três categorias. As aplicações típicas desta abordagem medem o conteúdo emocional de um determinado corpus de texto com base na prevalência de palavras negativas vs. positivas no corpus. Esses métodos de correspondência de palavras são chamados de métodos de bag-of-words (BOW) devido as características contextuais de cada palavra, como sua ordem no texto, classe gramatical, coocorrência com outras palavras e outras características contextuais específicas ao texto em que a palavra aparece, são ignorados.

Dentre esse tipo de método destaca-se o dicionário criado por [Loughran and McDonald \(2011\)](#) (LM). Os autores contruíram listas de palavras negativas e positivas que são selecionadas para serem apropriadas ao texto financeiro. Eles mostram que seus dicionários são superiores para classificar textos econômicos e financeiros a outros dicionários, por exemplo, o de [Apel and Grimaldi \(2012\)](#) e o *Harvard Psychosociological Dictionary*, que tende a categorizar incorretamente palavras neutras em um contexto financeiro/econômico (por exemplo, impostos, custos, capital, despesa, responsabilidade, risco, excesso e depreciação). Existem 2.355 palavras negativas e 354 palavras positivas nos dicionários LM. Portanto, para a construção dos índices de sentimentos via abordagem de dicionários fixos usamos o dicionário de LM.

[Shapiro, Sudhof and Wilson \(2020\)](#) afirma que a segunda abordagem, mais incipiente, emprega técnicas de aprendizado de máquina (ML) para construir modelos complexos para prever probabilisticamente o sentimento de um determinado conjunto de texto. Uma das aplicações dos modelos ML



é na construção de dicionários variantes no tempo. [Lima, Godeiro and Mohsin \(2019\)](#) usaram essa abordagem para criar um método de dicionário variante.

De acordo com [Lima, Godeiro and Mohsin \(2019\)](#) a suposição de um dicionário invariável no tempo não parece ser realista em documentos que introduzem novas palavras ao longo do tempo ou se o vocabulário usado em períodos de recessão difere do usado em períodos de expansões econômicas. Os autores ressaltam que mesmo se o vocabulário fosse constante ao longo do tempo, o poder preditivo de algumas palavras pode variar, ou seja, a relevância das palavras se alteram ao longo do tempo, mas a literatura existente não explica esse efeito e, portanto, os preditores resultantes não refletem as informações textuais mais preditivas encontradas nos documentos em um determinado momento. Portanto, aplicamos para a construção dos índices de sentimentos via dicionários variantes no tempo usamos a abordagem desenvolvida por [Lima, Godeiro and Mohsin \(2019\)](#).

Assim, utilizando a metodologia proposta pelos autores para construir o dicionário variante no tempo, primeiramente criamos um vetor de séries temporais,  $X_t$ , em que cada elemento do vetor mostra observações em série temporal da frequência em que cada palavra (ou combinação de palavras) aparece na ata do Copom e no RI até o tempo  $t$ . Portanto, esta etapa transforma as palavras em valores numéricos sem usar um dicionário pré-especificado (fixo). Essa representação numérica é de alta dimensão e esparsa; portanto, a redução da dimensionalidade deve ser empregada na próxima etapa. Na segunda etapa, usamos o *supervised machine learning* para selecionar as séries temporais mais preditivas (palavras)  $X_t^* \subset X_t$ .

O modelo de *elastic net* foi escolhido para realizar a segunda etapa:

$$y_{t+h} = W_t' \beta_h + X_t' \phi_h + \epsilon_{t+h} \quad (3)$$

em que  $h \geq 0$  é o horizonte de previsão,  $\hat{\beta}_h$  e  $\hat{\phi}_h$  são estimadas minimizando a seguinte função objetivo:

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

em que  $W_t$  é um vetor  $k \times 1$  de preditores pré-determinados, como defasagens de  $y_t$  bem como preditores tradicionais de dados estruturados e  $\|\cdot\|_{\ell_1}$  e  $\|\cdot\|_{\ell_2}$  são a norma  $\ell_1$  e  $\ell_2$ , respectivamente. Então, a partir da seleção das palavras com maior poder preditivo, temos para cada período  $t$  um conjunto de palavras que servem como dicionário léxico para a obtenção da série de sentimentos  $S$ . No presente artigo, vamos aplicar esse método na inflação e no PIB.

Por fim, ambas abordagens de dicionário, calculam o índice de sentimento pela diferença entre palavras positivas e negativas, dividida pela soma de palavras positivas e negativas, como foi proposto por [Hubert and Labondance \(2018\)](#):

$$S_t = \frac{\text{PositiveWords}_t - \text{NegativeWords}_t}{\text{PositiveWords}_t + \text{NegativeWords}_t} \quad (5)$$

Portanto, obtemos a medida de sentimentos,  $S$ , que varia entre -1 e 1.

## 3.2 Modelos de previsão

### 3.2.1 ARMA

Um dos modelos estatísticos mais comuns usados para previsão de séries temporais é o modelo de média móvel autoregressiva (ARMA), que pressupõe que observações futuras sejam guiadas principalmente por observações recentes. A informação que geralmente exibe comportamento persistente, é amplamente consistente com essa suposição. O melhor modelo para a taxa de inflação  $y_t$ , em nossa amostra, em que a representação mais simples é o AR (1), descrito a seguir:

$$y_t = \alpha + \beta y_{t-1} + \varepsilon_t \quad (6)$$

em que os parâmetros estimados  $[\hat{\alpha}; \hat{\beta}]'$  podem ser calculados usando uma amostra com  $t = 1, \dots, T$  observações:

$$y_{t+h} = \hat{\beta}^h y_T + \sum_{i=0}^{h-1} \hat{\alpha} \hat{\beta}^i \quad (7)$$

em que  $y_{t+h}$  é a previsão  $h$ -passos à frente.

### 3.3 Modelo de Fatores

A idéia de que variações de tempo em um grande número de variáveis podem ser resumidas por um pequeno número de fatores é empiricamente atraente e é empregada em um grande número de estudos em economia e finanças (Forni et al., 2000); (Stock and Watson, 2002). Considere,  $x_{i,t}$  serem os dados observados para o  $i$ -ésimo unidade de cross-section no tempo  $t$ , para  $i = 1, \dots, N$  e  $t = 1, \dots, T$ , e considere a seguinte representação fatorial dos dados:

$$x_{i,t} = \lambda_i' F_t + e_{i,t} \quad (8)$$

em que  $F_t$  é um vetor de fatores comuns,  $\lambda_i$  é um vetor de cargas fatoriais associados com  $F_t$  e  $e_{i,t}$  é o componente indiossincrático de  $x_{i,t}$ . Note que  $F_t$ ,  $\lambda_i$  e  $e_{i,t}$  são desconhecidos, pois apenas  $x_{i,t}$  é observável. Aqui, estimamos os fatores e respectivas cargas usando a análise de componentes principais (PCA), que é uma técnica bem estabelecida para redução de dimensão em séries temporais. O número de componentes é determinado pelo critério de Bai and Ng (2002). Após a estimativa do PCA dos fatores comuns  $F_t$ , empregamos a abordagem de previsão direta para modelar a taxa de inflação e o crescimento do PIB no tempo  $t + h$ , da seguinte forma:

$$y_{t+h} = \beta_h F_t + \varepsilon_{t+h} \quad (9)$$

Portanto, a previsão de informações da abordagem do modelo de fator direto,  $y_{t+h}$ , usando uma amostra de  $t = 1, \dots, T$  observações, é dado por:

$$y_{t+h} = \widehat{\beta}_h \widehat{F}_T, \quad \text{for } h = 1, \dots, H \quad (10)$$

em que  $\widehat{\beta}_h$  e  $\widehat{F}_T$  são os parâmetros a serem estimados.

#### 3.3.1 LASSO

Neste tipo de método de encolhimento a ideia é reduzir os parâmetros que correspondem a variáveis irrelevantes a zero. Sob algumas condições, é possível estimar modelos com mais variáveis que observações.

Entre os métodos de encolhimento, o *least absolute shrinkage and selection operator* (LASSO), introduzido por Tibshirani (1996) recebeu atenção especial. Foi demonstrado que o LASSO pode lidar com mais variáveis que observações, e o subconjunto correto de variáveis relevantes pode ser selecionado (Efron et al., 2004); (Meinshausen, Yu et al., 2009); (Zhao and Yu, 2006).

O estimador LASSO é definido como:

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

em que  $\lambda$  controla a quantidade de encolhimento e é determinado por técnicas orientadas a dados, como validação cruzada ou uso de critérios de informação. O modelo funciona mesmo quando o número de variáveis aumenta mais rapidamente que o número de observações e quando os erros são não gaussianos e heterocedásticos.

### 3.3.2 Random forest

A metodologia de *random forest* (RF) foi proposta inicialmente por Breiman (2001) como uma maneira de reduzir a variação de árvores de regressão e baseia-se na agregação de *bootstrap* (*bagging*) de árvores de regressão construídas aleatoriamente.

De acordo com Garcia, Medeiros and Vasconcelos (2017) uma árvore de regressão é um modelo não paramétrico baseado no particionamento binário recursivo do espaço covariável  $\mathbb{X}$ , em que a função  $T(\cdot)$  é uma soma de modelos locais (normalmente apenas uma constante), cada um dos quais é determinado em  $K \in \mathbb{N}$  regiões diferentes (partições) de  $\mathbb{X}$ . O modelo geralmente é exibido em um gráfico que possui o formato de uma árvore de decisão binária com nós  $N \in \mathbb{N}$  pai (ou divididos) e nós  $K \in \mathbb{N}$  terminais (também chamados de folhas) e que cresce a partir do nó raiz para os nós do terminal. Geralmente, as partições são definidas por um conjunto de hiperplanos, cada qual ortogonal ao eixo de uma determinada variável preditora, chamada de variável dividida. Portanto, condicional ao conhecimento das sub-regiões, a relação entre  $y_{t+h}$  e  $x_t$  na Equação 1 é aproximado por um modelo constante por partes, em que cada folha (ou nó terminal) representa um regime distinto.

Pode-se representar matematicamente um modelo complexo de árvore de regressão introduzindo a seguinte notação. O nó raiz está na posição 0 e um nó pai na posição  $j$  gera nós filhos a esquerda e a direita nas posições  $2j + 1$  e  $2j + 2$ , respectivamente. Todo nó pai possui uma variável dividida associada  $x_{s_j t} \in \mathbf{x}_t$ , onde  $s_j \in S = \{1, 2, \dots, q\}$ . Além disso, se permitirmos que  $\mathbf{J}$  e  $\mathbf{T}$  sejam os conjuntos de índices dos nós pais e terminal, respectivamente, uma arquitetura em árvore pode ser determinada completamente a partir de  $\mathbf{J}$  e  $\mathbf{T}$ .

O modelo de previsão baseado em árvores de regressão pode ser representado matematicamente como:

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

em que:

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

$$I(x_{s_j, t}; c_j) = \begin{cases} 1 & \text{if } x_{s_j, t} \leq c_j \\ 0 & \text{caso contrário} \end{cases} \quad (14)$$

$$n_{i,j} = \begin{cases} -1 & \text{se o caminho para a folha } i \text{ não inclui o nó pai} \\ 0 & \text{se o caminho para a folha } i \text{ inclui} \\ & \text{o nó filho direito do nó pai } j \\ 1 & \text{se o caminho para a folha } i \text{ inclui} \\ & \text{o nó filho esquerdo do nó pai } j \end{cases}$$

Seja  $\mathbb{J}_i$  o subconjunto de  $\mathbb{J}$  que contém os índices dos nós pais que formam o caminho para a folha  $i$ ; então,  $\boldsymbol{\theta}_i$  é o vetor que contém todos os parâmetros  $c_k$ , de modo que  $k \in \mathbb{J}_i$ ,  $i \in \mathbb{T}$ . Observe que  $\sum_{j \in \mathbb{J}} B_{\mathbb{J}_i}(\mathbf{x}_t; \boldsymbol{\theta}_j) = 1, \forall \mathbf{x}_t \in \mathbb{R}^{q+1}$ .

Uma *random forest* é uma coleção de árvores de regressão, cada uma especificada em uma subamostra de inicialização dos dados originais. Suponha que haja subamostras com inicialização  $B$  e denote a árvore de regressão estimada para cada uma das subamostras por  $H_{\mathbb{J}_b \mathbb{T}_b}(\cdot; \psi_b)$ . A previsão final é definida como:

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

Uma árvore de regressão é estimada para cada uma das subamostras de inicialização, repetindo as etapas a seguir recursivamente para cada nó terminal da árvore até que o número mínimo de observações em cada nó seja alcançado.

1. Selecione aleatoriamente  $m$  de  $q$  covariáveis como possível variáveis de divisão. variáveis.
2. Escolha a melhor variável/ponto de divisão entre os  $m$  candidatos.
3. Divida o nó em dois nós filhos.

As florestas aleatórias podem lidar com um número muito grande de variáveis explicativas, e o modelo previsto é altamente não linear. É importante notar que as amostras de autoinicialização são calculadas usando autoinicializações em bloco, pois estamos lidando com séries temporais.

### 3.3.3 Support Vector Machine (SVM)

Desde que o SVM foi introduzido a partir da teoria de aprendizagem estatística por [Vapnik \(1995\)](#), uma série de estudos foi anunciada sobre sua teoria e aplicações. Comparado com a maioria das outras técnicas de aprendizado, o SVM leva a aumentar o desempenho em reconhecimento de padrões, estimativa de regressão, previsão de séries temporais financeiras, dentre outras aplicações aplicações. A seguinte breve descrição de SVM se concentra inteiramente no problema de reconhecimento de padrões no campo de classificação. A explicação detalhada e as provas de SVM podem estar contidas nos livros de ([Vapnik, 1995](#)) e ([Vapnik, 1999](#)).

De acordo com [Shin, Lee and Kim \(2005\)](#) o SVM produz um classificador binário, os chamados hiperplanos de separação ótimos, por meio do mapeamento extremamente não linear dos vetores de entrada no espaço de recursos de alta dimensão. O SVM constrói um modelo linear para estimar a função de decisão usando limites de classe não lineares com base em vetores de suporte. Se os dados forem separados linearmente, o SVM treina máquinas lineares para um hiperplano ideal que separa os dados sem erro e na distância máxima entre o hiperplano e os pontos de treinamento mais próximos. Os pontos de treinamento mais próximos do hiperplano de separação ideal são chamados de vetores de suporte. Todos os outros exemplos de treinamento são irrelevantes para determinar os limites das classes binárias. Em casos gerais em que os dados não são separados linearmente, o SVM usa máquinas não lineares para encontrar um hiperplano que minimiza o número de erros do conjunto de treinamento.

## 3.4 Acurácia da previsão

Testes formais de capacidade preditiva podem ser feitos usando abordagens como as popularizadas por [Diebold and Mariano \(2002\)](#) (ver, por exemplo, [Clark and McCracken \(2010\)](#) e [Giacomini \(2010\)](#), para análises recentes) para verificar se as diferenças no RMSE refletem diferenças estatisticamente significativas entre as previsões. Assim, como no trabalho de [Clements and Reade \(2020\)](#), empregamos um teste baseado em regressão ([Diebold and Mariano, 2002](#)), onde construímos o termo  $e_t^{DM}$  como:

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

em que  $\hat{y}_{t|t-h}^{EM}$  é a previsão de referência, em nosso caso, a referência são as previsões do Focus. Convencionalmente, a função de perda  $L$  é a perda de erro ao quadrado, isto é,  $L(e) = e^2$ . O desempenho igual da previsão implica  $E(e_t^{DM}) = 0$ , e esse teste pode ser implementado usando o modelo de regressão:

$$e_t^{DM} = \alpha + u_t \quad (17)$$

com hipótese nula de precisão igual,  $H_0 : \alpha = 0$ . A significância de  $\alpha$  implica uma diferença no desempenho da previsão. Se  $\alpha > 0$ , isso implica que a previsão do *benchmark* é 'melhor' do que a previsão realizada, enquanto  $\alpha < 0$  implica o contrário, ou seja, que a previsão realizada é diferente em relação à previsão do *benchmark*.

### 3.5 Eficiência da previsão

Nessa etapa vamos comparar as projeções realizadas pelo mercado, representado pelo Focus e as informações qualitativas contidas no texto da ata do Copom e do Relatório de Inflação, regredindo os erros de previsão do Focus em função dos índices de sentimentos, e vamos repetir o mesmo procedimento as os erros de previsão do Focus para um trimestre à frente. Assim como em [Jones, Sinclair and Stekler \(2019\)](#), para este exercício de comparação usaremos uma versão extendida de uma regressão de [Mincer and Zarnowitz \(1969\)](#). Se as previsões do Focus contêm todos das informações relevantes incluídas no texto, então os erros de previsão do Focus não devem ser previstos pelos índices de sentimentos. A equação pode ser adaptada da seguinte maneira:

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

em que  $e_{t+h}^{Focus}$  é o erro de previsão do Focus que é definido como a diferença entre os valores realizados e os valores previstos pelo Focus,  $S_t$  o índice de sentimento e  $u_t$  é o resíduo. Testando a hipótese nula de que o coeficiente do índice de sentimento é igual a zero nos permite determinar se as previsões do Focus foram ou não melhoradas com a incorporação de informações de o texto.

## 4 Resultados

### 4.1 Índices de Sentimento

No total foram construídos três índices a partir da ata do Copom e três provenientes do RI, totalizando seis indicadores de sentimento. Os índices e suas respectivas definições podem ser visualizados na Tabela 3.

Tabela 3: Definição dos índices de sentimento

Índice de Sentimento	Definição
CM_Index	Índice de sentimento da ata do Copom usando o dicionário fixo
IR_Index	Índice de sentimento do Relatório de Inflação usando o dicionário fixo
IPCA_CM_Index	Índice de sentimento da ata do Copom usando o dicionário variante selecionando as palavras mais preditivas para o IPCA
IPCA_IR_Index	Índice de sentimento do Relatório de Inflação usando o dicionário variante selecionando as palavras mais preditivas para o IPCA
GDP_CM_Index	Índice de sentimento da ata do Copom usando o dicionário variante selecionando as palavras mais preditivas para o PIB
GDP_IR_Index	Índice de sentimento do Relatório de Inflação usando o dicionário variante selecionando as palavras mais preditivas para o PIB

As variáveis CM\_Index e IR\_Index foram construídas utilizando o dicionário de [Loughran and McDonald \(2011\)](#), portanto são índices de sentimentos advindos de um rol de palavras pré-selecionadas que não sofrem alteração com o tempo. Já os indicadores IPCA\_CM\_Index e IPCA\_IR\_Index são criadas por um conjunto de palavras que variam a cada período  $t$ , ou seja, para cada relatório o dicionário consiste das palavras mais preditivas para o IPCA que são selecionadas via *machine learning*. O mesmo ocorre para os índices GDP\_CM\_Index e GDP\_IR\_Index, a diferença reside na variável de responder, neste caso, o PIB. A principal vantagem dos índices de sentimentos provenientes do dicionário variante é que elas são capazes de representar um sentimento mais realista, pois podem



mensurar de forma mais precisa o sentimento do documento quando ocorrem eventos extremos, como por exemplo, a pandemia de Covid-19.

As Figuras 7 e 8 em anexo, ilustram bem esse ponto. É possível ver na Figura 7 os 10 coeficientes mais positivos e os 10 mais negativos para o IPCA na ata do Copom e RI, ou seja, as palavras mais preditivas para o IPCA nesses documentos. Na ata do Copom para o IPCA "budget", "certain" e "gas" as três palavras mais positivas e "coronavirus", "covid" e "basket" as mais negativas. No RI as três palavras mais positivas são "vaccines", "arabia" e "fuels", já as mais negativas são "election", "cigarettes" e "surprises", mas também aparecem termos que representam cenários de desastre ambiental como "brumadinho" e o cenário atual da Covid-19 como "epidemic". Para o PIB também temos que os termos "coronavirus", "covid", "infections" e "vaccines" são os mais preditivos. No caso dos índices de dicionário fixo é impossível essas séries captarem com precisão a pandemia de Covid-19, pois palavras como "covid" e "coronavirus" não estão presentes em tal dicionário.

Quando analisamos o tamanho dos documentos percebemos que eles variam com o tempo, em que o número de palavras é decrescente ao longo dos trimestres, como é evidente a partir da Figura 1. A ata do Copom no primeiro trimestre de 2005 continha cerca de 5.328 palavras e no segundo trimestre de 2020 uma quantidade de 1.797 palavras. O número de palavras na ata do Copom foi drasticamente reduzido a partir do ano de 2014. Já o RI no primeiro trimestre de 2005 possuía 57.091 palavras e 36.664 no fim da amostra. Ainda sobre o RI, podemos ver pela Figura 1 que entre os anos de 2012 e 2015 existe uma redução extrema na quantidade de palavras. Como já foi explicado na seção de base de dados, isso ocorreu porque durante esses anos o BCB não produziu o RI, apenas disponibilizou um sumário executivo sobre o documento.

As palavras com maior frequência na ata do Copom e no RI podem ser visualizadas na Figura 6, que mostra a nuvem de palavras de cada documento. Palavras como "growth", "goods", "rate", "credit", "industrial", "exchange", "expectation" e "sales" são os termos mais encontrados em ambos documentos. As nuvens de palavras também ajudam a mostrar que apesar do RI ser um documento com um número de palavras muito superior em relação à ata do Copom, os termos principais que aparecem nesses documentos são similares.

A Tabela 9 mostra as correlações entre os índices de sentimentos com o IPCA e o crescimento do PIB. É possível perceber que os índices de sentimento de dicionário fixo, CM\_Index e IR\_Index, possuem baixas correlações com as variáveis macroeconômicas. Esse fato não é totalmente anormal, pois trabalhos similares como os de [Catalfamo et al. \(2018\)](#) e [Mathy and Stekler \(2018\)](#) também encontraram baixas correlações com dados dos Estados Unidos.

Quando observamos os índices de sentimento de dicionário variante as correlações aumentam de forma significativa. Também podemos perceber que os índices que usam o crescimento do PIB na sua construção apresentam correlações com o PIB maiores do que a correlação do IPCA índices que usam o IPCA. Além disso, quando olhamos para os índices de dicionários variante, vemos que os provenientes da ata do Copom produzem correlações mais elevadas.

A maior correlação de todas foi do GDP\_CM\_Index com o crescimento do PIB (0.75). Esse valor é inferior ao encontrado nos trabalhos de [Jones, Sinclair and Stekler \(2019\)](#) e [Clements and Reade \(2020\)](#) que fizeram com dados do Banco Central da Inglaterra e usaram um dicionário fixo. Contudo, os autores não usaram o índice de sentimento "bruto", antes de os usar como preditores eles os "melhoraram" para ficarem mais próximos ao PIB e a inflação via dimensionamento e escalonamento. Então, se ele utilizassem os índices de dicionário fixo originais, eles iriam obter correlações muito inferiores.

Outro fato interessante é que os índices de sentimento variante no tempo possuem correlações ainda maiores com o IPCA e o PIB de um trimestre a frente. Isso mostra que o sentimento textual dos relatórios do BCB podem conter informações que já antecipam os movimentos do crescimento do PIB e da inflação.

Figura 1: Índices de sentimento com dicionário fixo e variante no tempo e IPCA

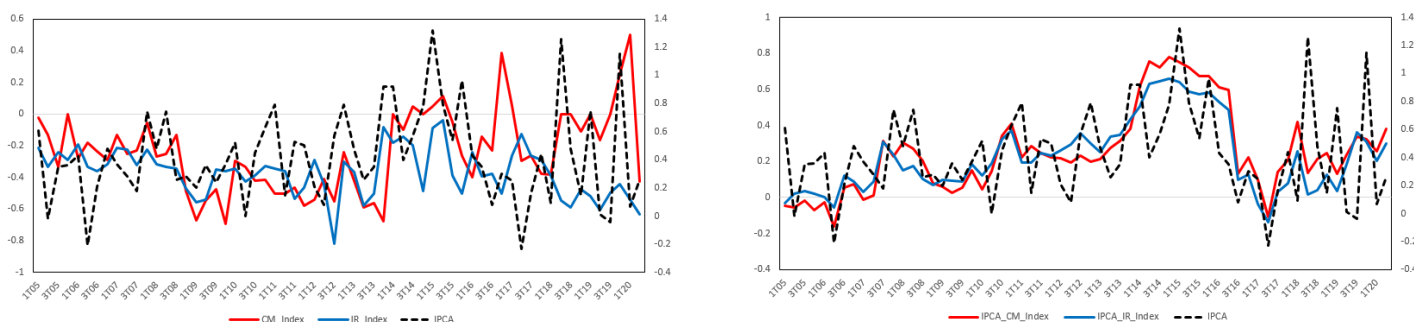


Figura 2: Índices de sentimento com dicionário fixo e variante no tempo e PIB



As Figuras 1 e 2 ilustram graficamente o movimento dos índices de sentimento juntamente com o crescimento do PIB e do IPCA ao longo do tempo. Apesar das escalas diferentes podemos ver que os índices de sentimento de dicionário variante se ajustam melhor as variáveis macroeconômicas. Então, a análise gráfica confirma visualmente o que já foi constatado nas correlações.

Os índices `IPCA_CM_Index` e `IPCA_IR_Index` seguem com maior precisão os picos e vales do IPCA em relação aos indicadores `CM_Index` e `IR_Index` e como já foi mencionado, devido a baixa correlação entre os índices de sentimento de dicionário fixo. A mesma conclusão pode ser adotada para a análise do crescimento do PIB. Entretanto, `CM_Index` e `IR_Index` acompanham mais de perto os movimentos do PIB do que o IPCA. É visível que esses índices conseguem captar a forte queda no produto devido à crise do Subprime e também conseguem acompanhar a queda do PIB por problemas fiscais em 2015, mas falham em captar a abrupta redução do PIB em no primeiro e segundo trimestre de 2020 proveniente da pandemia de Covid-19. Esse fato já era esperado, pois o dicionário financeiro de [Loughran and McDonald \(2011\)](#) possui termos capazes de detectar crises financeiras e fiscais, mas são incapazes de mensurar diretamente crises como a do Coronavírus. `CM_Index` ainda consegue acompanhar a queda do PIB de forma tardia a partir do segundo trimestre de 2020, pois o Coronavírus já tinha deteriorado de forma significativa a economia brasileira, dessa forma os documentos passaram a ter um sentimento pessimista devido ao fato de termos negativos aparecerem posteriormente com mais frequência.

Quando comparamos os índices de sentimento pela ótica do relatório de origem, vemos que de forma geral as séries de sentimentos provenientes da ata do Copom são mais otimistas em relação as séries do RI, inclusive `IR_Index` não possui nenhum valor positivo. Além disso, podemos notar que os índices de origem da ata do Copom possuem fortes picos de otimismo após 2014, já os índices de sentimento do RI se mostram mais pessimistas. Isso deve ocorrer devido a função e estrutura distinta entre os documentos, pois a ata do Copom é um documento com natureza mais comunicativa e objetiva, já o RI é um documento mais analítico. Então, como o BCB usa a ata do Copom para comunicar a decisão de política monetária o documento tende a ser mais otimista ancorar as expectativas do mercado e notamos que a partir de 2014 o tom da ata do Copom passou ter um padrão mais otimista.

Os índices construídos por [Jones, Sinclair and Stekler \(2019\)](#) e [Clements and Reade \(2020\)](#) que usam o RI do Banco Central da Inglaterra também seguem a linha dos nossos índices de origem do RI, ou seja, tendem a ter um tom mais pessimista.

## 4.2 Previsões e Acurácia

Para o nosso exercício de previsão usamos cinco tipo de modelos de previsão: ARMA, VAR, LASSO, Random Forest e SVM. O nosso *benchmark* são as previsões médias realizadas pelo mercado, representadas pelo Focus. Com o objetivo verificar se a inclusão dos índices de sentimento em modelos de previsão é capaz de melhorar a acurácia das previsões para o IPCA e o PIB estimamos os modelos multivariados com diferentes categorias. O que vai definir a categoria são os índices de sentimento. Neste caso, a primeira categoria inclui todas os índices de sentimento (com dicionário fixo e dicionário variante), a segunda usa apenas os índices de dicionário variante, a terceira utiliza apenas os índices de dicionário fixo e por fim a quarta categoria é de modelos sem nenhum índice de sentimento. A Tabela 4 a seguir ilustra as diferentes categorias de modelos.

Tabela 4: Definição das variáveis

Modelo	Definição
LASSO1	Modelo LASSO com todas as variáveis macroeconômicas + índices de sentimento com dicionário variante + índices de sentimento com dicionário fixo
LASSO2	Modelo LASSO com todas as variáveis macroeconômicas + índices de sentimento com dicionário variante
LASSO3	Modelo LASSO com todas as variáveis macroeconômicas + índices de sentimento com dicionário fixo
LASSO4	Modelo LASSO com todas as variáveis macroeconômicas
RF1	Modelo Random Forest com todas as variáveis macroeconômicas + índices de sentimento com dicionário variante + índices de sentimento com dicionário fixo
RF2	Modelo Random Forest com todas as variáveis macroeconômicas + índices de sentimento com dicionário variante
RF3	Modelo Random Forest com todas as variáveis macroeconômicas + índices de sentimento com dicionário fixo
RF4	Modelo Random Forest com todas as variáveis macroeconômicas
SVM1	Modelo Suport Vector Machines com todas as variáveis macroeconômicas + índices de sentimento com dicionário variante + índices de sentimento com dicionário fixo
SVM2	Modelo Suport Vector Machines com todas as variáveis macroeconômicas + índices de sentimento com dicionário variante
SVM3	Modelo Suport Vector Machines com todas as variáveis macroeconômicas + índices de sentimento com dicionário fixo
SVM4	Modelo Suport Vector Machines com todas as variáveis macroeconômicas
FM1	Modelo de Fatores com todas variáveis macroeconômicas + índices de sentimento com dicionário variante + índices de sentimento com dicionário fixo
FM2	Modelo de Fatores com todas variáveis macroeconômicas + índices de sentimento com dicionário variante
FM3	Modelo de Fatores com todas variáveis macroeconômicas + índices de sentimento com dicionário fixo
FM4	Modelo de Fatores com todas variáveis macroeconômicas

As Figuras 3 e 4<sup>3</sup> ilustram as previsões do IPCA e o crescimento do PIB, respectivamente, para fora da amostra e considerando  $h = 1$ , ou seja, um trimestre a frente. A amostra total vai do primeiro trimestre de 2005 até o segundo trimestre de 2020, totalizando 62 observações. 80% da amostra total

<sup>3</sup> Nas Figuras 3 e 4 optamos para fins de melhor visualização exibir as séries a partir do primeiro trimestre de 2015.

foi utilizada para compor a amostra de treinamento dos modelos e os 20% para compor a amostra de validação. Em ambas figuras percebemos que as previsões do Focus ficam muito próximas aos dados realizados e em alguns trimestres os modelos de *machine learning* conseguem superar o Focus.

Figura 3: Previsões do IPCA para  $h = 1$  - ata do Copom e Relatório da Inflação

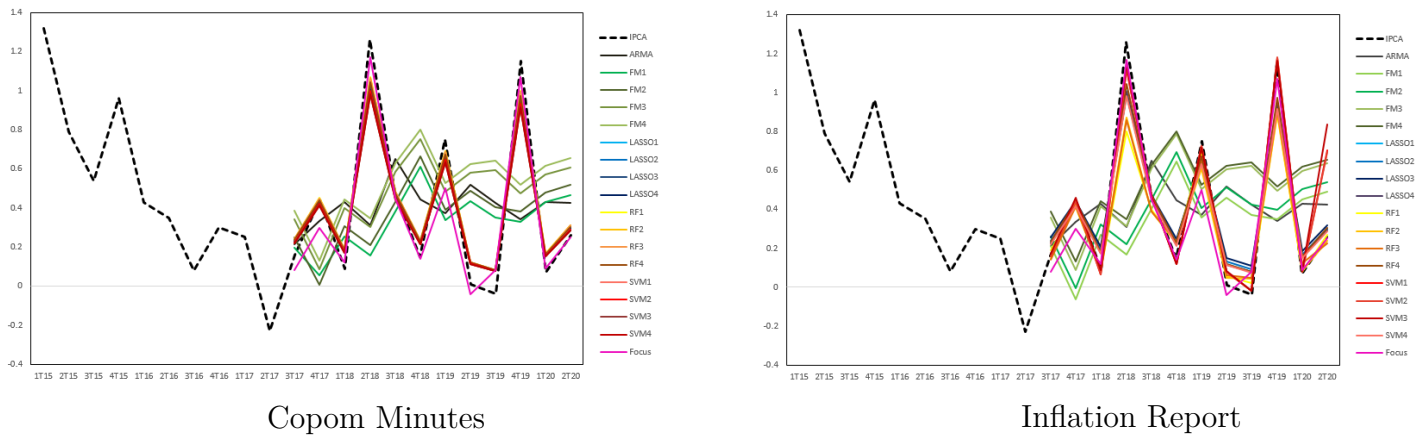
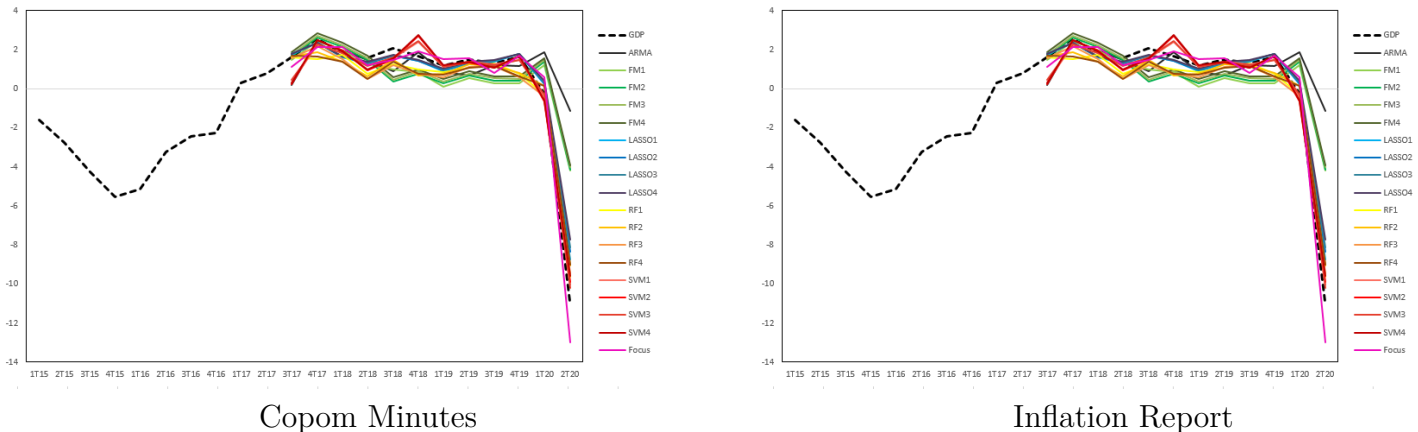


Figura 4: Previsões do PIB para  $h = 1$  - ata do Copom e Relatório da Inflação



As Tabelas 5 e 6 ilustram o erro de previsão (MSE) de cada modelo para até quatro trimestres à frente, juntamente com o teste de Diebold-Mariano. Neste caso, para o teste de DM consideramos o Focus como referência, de modo que a hipótese nula é de previsão igual ao do Focus e a hipótese alternativa é previsão diferente do Focus.

Foi possível constatar alguns fatos. O primeiro é que assim como reportado nos trabalhos de ((Medeiros and Mendes, 2016), (Medeiros et al., 2019), (Araujo and Gaglianone, 2020)) os modelos de *machine learning* apresentaram um desempenho satisfatório em termos de acurácia para um período à frente, pois as previsões ficaram próximas às previsões do Focus e alguns até superando o *benchmark*, entretanto, para mais períodos à frente as previsões já começam a se afastar do Focus. Ou seja, previsões de curto prazo de até um trimestre à frente, conseguimos nos aproximar e até superar o Focus, mas para trimestres mais à frente nossas previsões ficam longe do Focus. O segundo é que os MSEs reportados para o IPCA são inferiores em relação aos MSEs para o PIB. Então, como já era esperado, tanto para nós como para a média do mercado é mais fácil prever a inflação do que o crescimento do produto.

O terceiro é que os modelos que incluíram os índices de sentimento obtiveram um desempenho melhor do que os modelos que não incorporam as séries de polaridade. Os melhores modelos são os

que usam ao mesmo tempo os índices de sentimento de dicionário fixo e dicionário variante (categoria 1) seguidos pelos modelos que consideram apenas os índices de sentimento de dicionário variante. Outro fato é que os modelos que usaram as séries de sentimentos oriundas da ata do Copom obtiveram um menor MSE, tanto para o IPCA como para o crescimento do PIB.

Tabela 5: Mean Squared Error (MSE) e teste de Diebold-Mariano para as previsões do IPCA

Modelo	Copom				Relatório da Inflação			
	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

Tabela 6: Mean Squared Error (MSE) e teste de Diebold-Mariano para as previsões do PIB

Modelo	Copom				Relatório da Inflação			
	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

A Tabela 6 ilustra que para a previsão do IPCA o modelo LASSO se destacou, tanto com os índices de sentimento da ata do Copom como para os do RI. Para a inflação apenas esse modelo conseguiu não rejeitar a hipótese nula do teste de DM de igualdade com o Focus, os demais modelos rejeitaram a hipótese. Já para as previsões do PIB os modelos RF e SVM apresentaram um desempenho melhor, inclusive alguns deles conseguiram superar a acurácia do Focus, tais como o RF1 e RF2 com os índices de sentimento da ata do Copom e SVM1 em ambos documentos. Todos esses modelos usam séries de polaridade, então, para o crescimento do produto só foi possível superar a média do mercado com a inclusão de índices de sentimento nas previsões.

Apesar de Medeiros et al. (2019) e Araujo and Gaglianone (2020) também terem obtido sucesso em superar o Focus com previsões do IPCA para um período à frente sem usar índices de sentimento, vale ressaltar que os autores usaram uma frequência mensal e o presente artigo usou uma frequência trimestral. Além disso, como foi constatado no presente trabalho, a inclusão de informações textuais da ata do Copom e do Relatório de Inflação só foi fundamental para superar o Focus nas previsões do



crescimento do PIB. Para as previsões do IPCA as informações textuais foram capazes de melhorar o desempenho dos nossos modelos, mas não possibilitou ganhar da média do mercado. Esse fato não é estranho, pois como já mencionamos, é mais fácil obter menores erros de previsão com a inflação do que com o PIB, então, com o IPCA existe uma menor margem para a melhoria no desempenho com o uso de índices de sentimento.

Adicionalmente, verificamos se os melhores modelos de previsão ilustrados anteriormente são diferentes estatisticamente entre si. Então, como pode ser visto nas Tabelas 8 e 9, realizamos novamente o teste de DM para os três melhores modelos de previsão para cada horizonte de previsão. De forma geral é possível ver que a hipótese nula de igualdade de acurácia entre os modelos não pode ser rejeitada. Então, não existe diferença estatística no desempenho dos melhores modelos de cada horizonte, relatório e variável macroeconômica.

### 4.3 Regressões de Eficiência

Comparamos as previsões da média do mercado para o IPCA e para o PIB com as informações qualitativas provenientes dos textos da ata do Copom e do RI, regredindo os erros de previsão de curto prazo do Focus nos índices de sentimento e os erros de previsão do Focus de um trimestre à frente com os índices de sentimento, usando uma versão estendida de uma regressão de [Mincer and Zarnowitz \(1969\)](#). Se as previsões do Focus contiverem todas as informações relevantes que estão incluídas no texto da ata do Copom e do RI, então os erros de previsão não devem ser previstos pelas índices de sentimento dos textos.

Tabela 7: Regressão de Mincer–Zarnowitz para a inflação

Erro de previsão em (t)				
	Copom_Index	IR_Index	IPCA_CM_Index	IPCA_IR_Index
Constante	-0.024 (0.016)	-0.037 (0.030)	-0.005 (0.017)	-0.015 (0.018)
Coefficiente do	0.054	0.070	0.024	0.015
Índice de Sentimento	(0.046)	(0.077)	(0.050)	(0.059)
$R^2$	0.023	0.013	0.040	0.011
Erro de previsão em (t+1)				
	Copom_Index	IR_Index	IPCA_CM_Index	IPCA_IR_Index
Constante	-0.013 (0.016)	-0.016 (0.031)	-0.009 (0.017)	-0.013 (0.018)
Coefficiente do	0.005	0.012	0.011	0.006
Índice de Sentimento	(0.047)	(0.080)	(0.051)	(0.060)
$R^2$	0.018	0.038	0.084	0.165

Tabela 8: Regressão de Mincer–Zarnowitz para o PIB

Erro de previsão em (t)				
	CM_Index	IR_Index	GDP_CM_Index	GDP_IR_Index
Constante	-0.243 (0.186)	-0.497 (0.361)	-0.088 (0.181)	-0.388** (0.130)
Coefficiente do	0.624	0.281	1.367*	1.577**
Índice de Sentimento	(0.533)	(0.908)	0.564	(0.553)
$R^2$	0.022	0.016	0.089	0.079
Erro de previsão em (t+1)				
	CM_Index	IR_Index	GDP_CM_Index	GDP_IR_Index
Constante	-0.330 (0.189)	-0.368 (0.369)	-0.079 (0.198)	-0.389** (0.122)
Coefficiente do	0.274	0.076	1.345*	2.604***
Índice de Sentimento	(0.545)	(0.940)	(0.623)	(0.650)
$R^2$	0.043	0.011	0.073	0.214

Os resultados das regressões mostram que todos os coeficientes dos índices de sentimento não rejeitam a hipótese nula de que são zero para os erros de previsão do IPCA, ou seja, as informações

textuais da ata do Copom e do RI não afetam os erros de previsão do Focus para a inflação. Então, não há evidências de que as informações textuais desse documentos possam melhorar as previsões realizadas pelo Focus para a inflação. O mesmo foi constatado para as previsões do IPCA pelo Focus para um trimestre à frente. Esses resultados estão em conformidade ao que foi encontrado por [Clements and Reade \(2020\)](#) que também não encontraram um coeficiente estatisticamente significativo do índice de sentimento na explicação dos erros de previsão da inflação do Banco Central da Inglaterra.

Já para as previsões do crescimento do PIB encontramos evidências de que as informações textuais contêm informações relevantes que podem melhorar as previsões do Focus, pois os erros de previsão em tempo real e para um trimestre à frente do PIB são explicados pelos índices de sentimento. Contudo, esse resultado só é válido para os índices de sentimento de dicionário variante (GDP\_CM\_Index e GDP\_IR\_Index) e os maiores coeficientes e mais significativos foram os do GDP\_IR\_Index. [Jones, Sinclair and Stekler \(2019\)](#) encontraram significância estatística do coeficiente do índice de sentimento apenas para os erros de previsão em tempo real, já os erros de previsão para um trimestre à frente do crescimento do produto não apresentaram relação com o índice de sentimento. Já [Clements and Reade \(2020\)](#) obtiveram resultados similares em termos de significância estatística do coeficiente do índice de sentimento, tanto para o erro de previsão no mesmo período da polaridade como para os erros de previsão em  $t+1$ . Assim como no presente artigo, [Clements and Reade \(2020\)](#) encontraram sinais positivos dos coeficientes, enquanto [Jones, Sinclair and Stekler \(2019\)](#) estimaram sinais negativos.

Como os erros de previsão são definidos como valores realizados menos previstos, isso significa que as previsões do Focus ao longo deste período foram excessivamente pessimistas, e que o tom mais positivo do texto é uma avaliação mais precisa que proporciona uma redução nos erros de previsão.

Portanto, esses resultados mostram que levar em consideração as informações textuais da ata do Copom e do RI são úteis para obter uma imagem mais completa das condições econômicas atuais e futuras.

## 5 Conclusões

A proposta do presente artigo foi de verificar se as informações textuais contidas na ata do Copom e no Relatório de Inflação publicados pelo Banco Central do Brasil são capazes de melhorar as previsões de variáveis macroeconômicas, especificamente, a taxa de inflação e o crescimento do PIB.

A partir dos resultados obtidos, podemos concluir alguns pontos. Constatamos que o uso de uma abordagem de análise textual via machine learning para selecionar as palavras que vão compor o dicionário, produz índices de sentimentos que captam de forma mais realista o sentimento existente nos textos das publicações do BCB, além de produzir índices mais correlacionados com as variáveis macroeconômicas. As correlações entre os índices de sentimento e as variáveis macroeconômicas revelam que os indicadores de sentimento, neste caso os de dicionário variante, que além de apresentarem correlações elevadas em tempo real, também possuem uma elevada correlação com as variáveis macroeconômicas para um trimestre à frente. Isso revela que as informações textuais dos documentos do BCB podem conter informações relevantes sobre o futuro do IPCA e, principalmente, do crescimento do PIB.

Também foi possível verificar que os modelos de machine learning proporcionam um elevado grau de acurácia em relação os modelos tradicionais como ARMA e VAR. Outro ponto interessante é que os modelos que incorporaram índices de sentimentos como preditores obtiveram os menores MSEs e que os melhores modelos foram os que usaram ao mesmo tempo os índices de dicionário variante e o dicionário fixo, seguidos pelos modelos que utilizaram os índices provenientes de dicionário variante e os piores modelos foram os que não consideraram nenhuma variável de sentimento no seu conjunto de preditores. Além disso, o teste de Diebold-Mariano ilustrou que as previsões de curto prazo para um trimestre à frente dos modelos de machine learning ficaram próximas às previsões realizadas pelo Focus, e para o crescimento do PIB, algumas de nossas previsões foram capazes de superar o Focus.

Um resultado revelante que foi obtido foi que os índices de sentimento de dicionário variante oriundos tanto da ata do Copom como do RI conseguem explicar os erros de previsão do Focus em tempo real para o crescimento do PIB. O mesmo foi encontrado para os erros de previsão para um trimestre à frente. Esse fato significa que o texto da ata do Copom e do RI podem reduzir o erro de previsão do Focus e, neste caso, melhorar as previsões do Focus. Essa análise também aponta que o coeficiente dos índices de sentimento foi positivo, idnicando que as previsões do Focus para o PIB foram excessivamente pessimistas durante a janela temporal da amostra. Já para o IPCA os mesmo resultado não foi encontrado pois os coeficientes estimados não são estatisticamente significativos. Esses resultados fazem sentido, pois o Focus apresentou menor MSE nas previsões para o IPCA e maiores para o PIB. Além disso, só conseguimos superar o Focus com o uso de índices de sentimento no crescimento do PIB.

Por fim, podemos concluir que as informações textuais da ata do Copom e do RI conseguiram melhorar as previsões de variáveis macroeconômicas, principalmente previsões para o PIB e que é importante o mercado passe a considerar tais informações textuais em suas previsões de curto prazo. Além disso, tais informações podem ser úteis para mostrar as condições futuras da economia.

## References

- 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.
- BAÑBURA, 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.
- GOLDFARB, R. S.; STEKLER, H. O.; DAVID, J. Methodological issues in forecasting: Insights from the egregious business forecast errors of late 1930. *Journal of Economic Methodology*, Taylor & Francis, v. 12, n. 4, p. 517–542, 2005.
- HUBERT, P.; LABONDANCE, F. Central bank sentiment. URL: <https://www.nbp.pl/badania/seminaria/14xi2018.pdf>. *Working paper*, 2018.
- INOUE, A.; KILIAN, L. How useful is bagging in forecasting economic time series? a case study of us consumer price inflation. *Journal of the American Statistical Association*, Taylor & Francis, v. 103, n. 482, p. 511–522, 2008.
- 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.
- LUNDQUIST, K.; STEKLER, H. O. Interpreting the performance of business economists during the great recession. *Business Economics*, Springer, v. 47, n. 2, p. 148–154, 2012.
- 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.
- SIMS, C. A. Macroeconomics and reality. *Econometrica: journal of the Econometric Society*, JSTOR, p. 1–48, 1980.

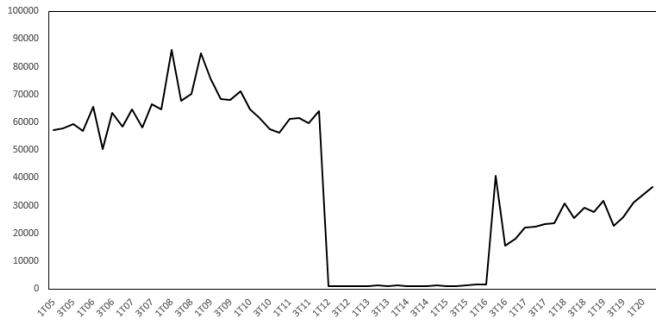


- STEKLER, H.; SYMINGTON, H. Evaluating qualitative forecasts: The fomc minutes, 2006–2010. *International Journal of Forecasting*, Elsevier, v. 32, n. 2, p. 559–570, 2016.
- STOCK, J. H.; WATSON, M. W. Forecasting using principal components from a large number of predictors. *Journal of the American 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.

Figura 5: Quantidade de palavras na ata do Copom e no Relatório da Inflação



## Copom Minutes



# Inflation Report

Figura 6: Nuvem de palavras - ata do Copom e Relatório da Inflação

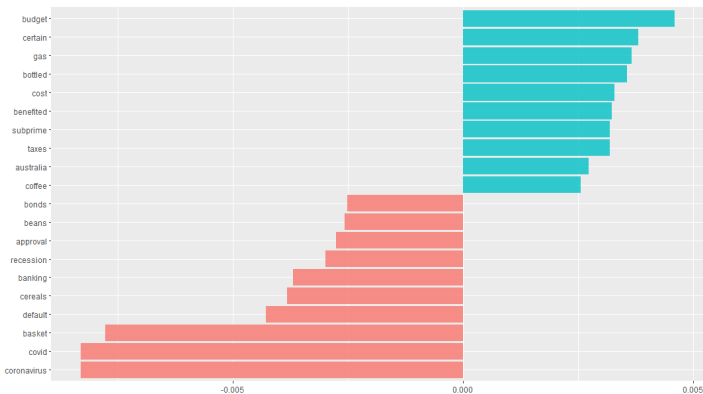


## Copom Minutes

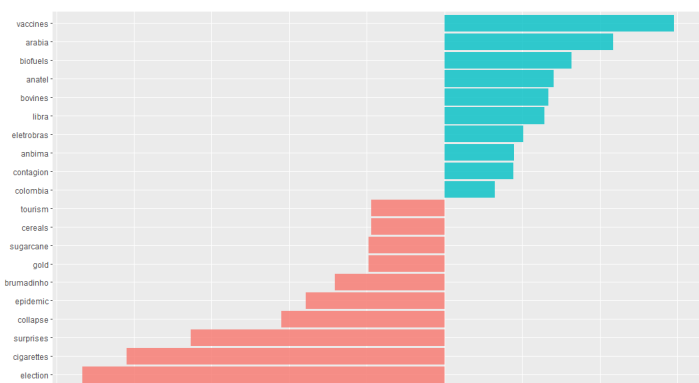


# Inflation Report

Figura 7: Os 10 coeficientes mais positivos e negativos da ata do Copom e do Relatório da Inflação para o IPCA



## Copom Minutes



# Inflation Report

Figura 8: Os 10 coeficientes mais positivos e negativos da ata do Copom e do Relatório da Inflação para o PIB

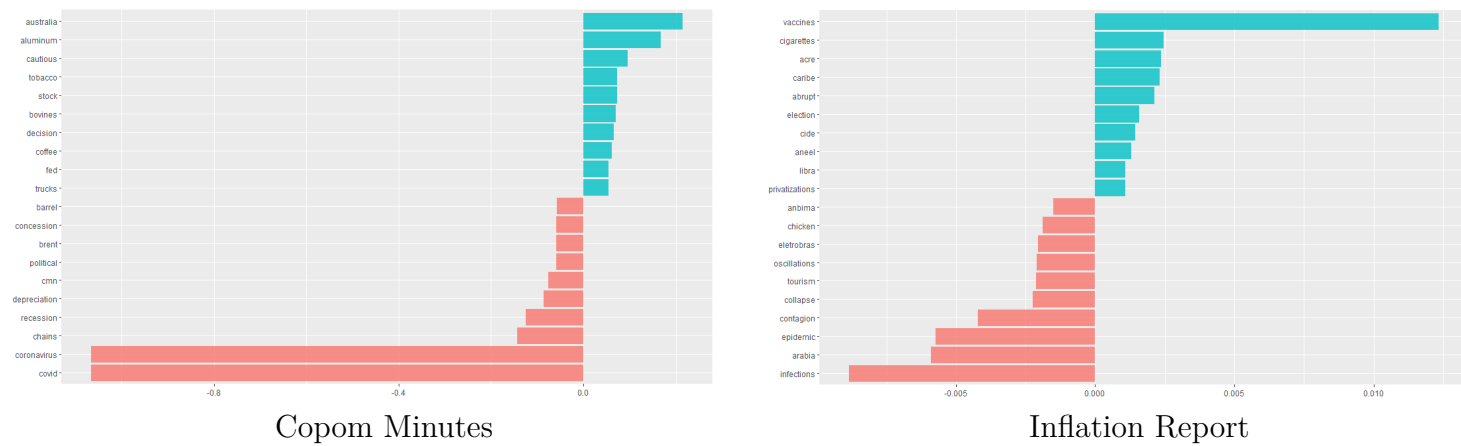


Tabela 9: Correlação entre os índices de sentimento e as variáveis macroeconômicas

	$IPCA_t$	$IPCA_{t+1}$	$PIB_t$	$PIB_{t+1}$
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

Tabela 10: Teste de Diebold-Mariano para os três melhores modelos de previsão do IPCA

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	-

Tabela 11: Teste de Diebold-Mariano para os três melhores modelos de previsão do PIB

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	-