



UNIVERSIDADE FEDERAL DO CEARÁ
PROGRAMA DE PÓS-GRADUAÇÃO EM ECONOMIA – CAEN
MESTRADO ACADÊMICO EM ECONOMIA

MARESSA SOARES DE SOUSA

**PREVISÃO DE CRESCIMENTO ECONÔMICO A PARTIR DA ANÁLISE TEXTUAL
DAS ATAS DO COPOM (2013-2023)**

FORTALEZA

2024

MARESSA SOARES DE SOUSA

PREVISÃO DE CRESCIMENTO ECONÔMICO A PARTIR DA ANÁLISE TEXTUAL DAS
ATAS DO COPOM (2013-2023)

Dissertação apresentada ao Programa de Pós-Graduação em Economia da Universidade Federal do Ceará, como requisito parcial à obtenção do título de Mestrado em Economia. Área de concentração: Crescimento Econômico.

Orientador: Prof. Dr. Sérgio Aquino de Souza.

FORTALEZA

2024

Dados Internacionais de Catalogação na Publicação
Universidade Federal do Ceará
Sistema de Bibliotecas
Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

S697p Sousa, Maressa Soares de.
Previsão de crescimento econômico a partir da análise textual das atas do COPOM (2013-2023) /
Maressa Soares de Sousa. – 2024.
51 f.

Dissertação (mestrado) – Universidade Federal do Ceará, Faculdade de Economia, Administração,
Atuária e Contabilidade, Programa de Pós-Graduação em Economia, Fortaleza, 2024.
Orientação: Prof. Dr. Sérgio Aquino de Souza.

1. Crescimento Econômico. 2. Análise Textual. 3. Índice de Sentimento. I. Título.

CDD 330

MARESSA SOARES DE SOUSA

PREVISÃO DE CRESCIMENTO ECONÔMICO A PARTIR DA ANÁLISE TEXTUAL DAS
ATAS DO COPOM (2013-2023)

Tese ou Dissertação apresentada ao Programa de Pós-Graduação em Economia da Universidade Federal do Ceará, como requisito parcial à obtenção do título de Mestrado em Economia. Área de concentração: Crescimento Econômico.

Aprovada em: 29/05/2024.

BANCA EXAMINADORA

Prof. Dr. Sérgio Aquino de Souza (Orientador)
Universidade Federal do Ceará (UFC)

Prof. Dr. Fabrício Carneiro Linhares
Universidade Federal do Ceará (UFC)

Prof. Dr. Ricardo Brito Soares
Universidade Federal do Ceará (UFC)

Ao meu esposo, Edson Soares.

AGRADECIMENTOS

Ao Prof. Dr. Sérgio Aquino de Souza, por sua orientação exemplar, paciência e dedicação incansável ao longo desta pesquisa.

Aos professores participantes da banca examinadora Prof. Dr. Fabrício Carneiro Linhares e Prof. Dr. Ricardo Brito Soares pelo tempo dedicado, pelas contribuições valiosas e sugestões enriquecedoras.

Aos professores Paulo Neto, Maurício Benegas, Márcio Veras, Fred, Leandro Rocco, Emerson, Flávio Ataliba, Zilânia, Ivan Castelar e Sebastião Carneiro pelas lições de vida, ensinamentos profissionais e aulas inspiradoras.

Aos colegas da turma de mestrado Wendell, Victor, Yuri Timbó, Kaique, Letícia, Brisa, Pâmela, José Carlos e Guilherme, por compartilharem experiências e companhias durante os momentos mais desafiadores nos corredores acadêmicos.

A minha mãe, pai e irmãs (Carol e Priscila) pelo apoio constante e por acreditarem em meu potencial.

A minha psicóloga, Janice, por suas palavras acertadas nos momentos de aflição e incerteza.

A minha querida amiga, Amanda, por sua presença constante, apoio e compreensão em todos os momentos.

Ao meu esposo, Edson, cuja fé em mim transcende a minha própria. Seu amor, demonstrado através de ações concretas, como o incentivo à organização do meu tempo e o estímulo ao meu crescimento contínuo, foi fundamental. Seu apoio nas vitórias e desafios me motivou a seguir adiante em cada etapa deste caminho.

“Os economistas têm começado a aprender [...] como as ‘histórias’ e a narração influenciam o comportamento econômico. [...] Não podemos prever o futuro, mas podemos ser mais bem preparados para os contos que podem vir a influenciar dele” (SHILLER, R. J, 2019, p. 25).

RESUMO

Este estudo investiga a influência dos sentimentos expressos nas atas do Comitê de Política Monetária (COPOM) do Banco Central do Brasil (BACEN) sobre o crescimento econômico, utilizando análise textual no contexto das tecnologias de informação emergentes. A pesquisa abrange um período de 11 anos, de 2013 a 2023, analisando 88 documentos para determinar como as percepções e expectativas registradas podem prever variações econômicas futuras. A metodologia empregada foi bifásica: inicialmente, construiu-se um índice de predição baseado no Dicionário Loughran-McDonald, adaptado ao contexto brasileiro; em seguida, aplicaram-se modelos de regressão linear simples e autorregressivos de ordem 2 (AR(2)). Essa abordagem foi crucial em um cenário global de incertezas, onde métodos tradicionais frequentemente não capturam todas as nuances que afetam as tendências econômicas. Os resultados mostraram uma correlação significativa entre os sentimentos expressos nas atas e as mudanças subsequentes no crescimento econômico, confirmando a utilidade do método do dicionário como uma ferramenta capaz de melhorar essa previsão. Esses *insights* são valiosos para orientar políticas econômicas e estratégias de investimento, reforçando a relevância da integração de novas técnicas analíticas para aprimorar a precisão e a confiabilidade dos modelos econômicos.

Palavras-chave: crescimento econômico; análise textual; índice de sentimento.

ABSTRACT

This study investigates the influence of sentiments expressed in the minutes of the Monetary Policy Committee (COPOM) of the Central Bank of Brazil (BACEN) on economic growth, using textual analysis in the context of emerging information technologies. The research covers a period of 11 years, from 2013 to 2023, analyzing 88 documents to determine how recorded perceptions and expectations can predict future economic variations. The methodology employed was biphasic: initially, a prediction index was constructed based on the Loughran-McDonald Dictionary, adapted to the Brazilian context; subsequently, simple linear regression and autoregressive models of order 2 (AR(2)) were applied. This approach was crucial in a global context of uncertainties, where traditional methods often fail to capture all the nuances that affect economic trends. The results showed a significant correlation between the sentiments expressed in the minutes and subsequent changes in economic growth, confirming the utility of the dictionary method as a tool to improve this prediction. These insights are valuable for guiding economic policies and investment strategies, reinforcing the relevance of integrating new analytical techniques to enhance the precision and reliability of economic models.

Keywords: economic growth; textual analysis; sentiment index.

LISTA DE FIGURAS

Figura 1 – Fases da etapa de construção do indicador	30
--	----

LISTA DE GRÁFICOS

Gráfico 1 – Taxa de Crescimento (%) Trimestre / mesmo trimestre do ano anterior	37
Gráfico 2 – Taxa de Crescimento (%) Trimestre / trimestre imediatamente anterior (com ajuste sazonal)	38

LISTA DE TABELAS

Tabela 1 – Classificação das palavras da Ata nº 172 (Jan./2013).....	31
Tabela 2 – Índice de sentimento dos anos de 2013 e 2014.....	33
Tabela 3 – Resultados para os anos de 2013 e 2014	40
Tabela 4 – Estatísticas dos Modelos.....	41
Tabela 5 – Resultado das análises de regressão.....	42

LISTA DE ABREVIATURAS E SIGLAS

AR(2)	Autorregressivo de ordem 2
API	Interface de Programação de Aplicações
BACEN	Banco Central do Brasil
<i>BOW</i>	<i>Bag-of-words</i>
CMN	Conselho Monetário Nacional
COPOM	Comitê de Política Monetária
<i>DSGE</i>	<i>Dynamic Stochastic General Equilibrium</i>
EUA	Estados Unidos das Américas
EQM	Erro Quadrático Médio
ILM	Índice de Sentimento de <i>Loughran-McDonald</i>
IBGE	Instituto Brasileiro de Geografia e Estatística
<i>JSON</i>	<i>JavaScript Object Notation</i>
<i>ML</i>	<i>Machine Learning</i>
PIB	Produto Interno Bruto
PLN	Processamento de Linguagem Natural
<i>URL</i>	<i>Uniform Resource Locator</i>

LISTA DE SÍMBOLOS

% Porcentagem

SUMÁRIO

1	INTRODUÇÃO	14
2	REFERENCIAL TEÓRICO	16
2.1	Política Monetária	17
<i>2.1.1</i>	<i>Modelos de Expectativas</i>	<i>18</i>
<i>2.1.2</i>	<i>Modelos de Metas de Inflação</i>	<i>20</i>
2.2	Análise Textual	23
<i>2.2.1</i>	<i>Método do Dicionário</i>	<i>24</i>
<i>2.2.1.1</i>	<i>Índice de Sentimentos</i>	<i>25</i>
<i>2.2.1.2</i>	<i>Dicionário Loughran-McDonald</i>	<i>26</i>
2.3	Crescimento Econômico	28
3	METODOLOGIA	29
3.1	Construção do Indicador de sentimento para Predição	29
3.2	Análise de Correlação (regressão simples) e Modelo Autorregressivo (AR-2)	34
4	RESULTADOS E DISCUSSÃO	36
5	CONCLUSÃO	43
	REFERÊNCIAS	44
	APÊNDICE A – TABELA COMPLETA DOS DADOS UTILIZADOS NOS MODELOS DE REGRESSÃO	49
	APÊNDICE B – TABELA COMPLETA COM TODAS AS REGRESSÕES AR(2)	51

1 INTRODUÇÃO

A utilização de modelos econométricos para análise de dados quantitativos é uma prática estabelecida e refinada ao longo do tempo na comunidade acadêmica. Por meio da organização de números em gráficos e tabelas, essa abordagem permite uma visualização clara das tendências e valores, acessível tanto a pesquisadores especializados quanto a leigos. No entanto, nem sempre os dados estão disponíveis em formato numérico para essa análise direta. Em contrapartida, há uma vasta quantidade de informações expressas em forma de texto, abordando uma variedade de temas, que requerem interpretação por parte de profissionais antes que possam ser analisadas e compartilhadas.

Portanto, a análise textual é essencial para compreender e interpretar informações disponíveis. Com a vasta quantidade de dados, é crucial usar uma ferramenta computacional que possa realizar essa tarefa, compilando dados e fornecendo rapidamente as informações extraídas. Isso pode ser feito através de algoritmos complexos que convertem texto em valores numéricos, permitindo o treinamento de modelos que preveem resultados ou classificam dados, especialmente em situações onde as relações entre palavras e resultados são complexas e dinâmicas.

No entanto, podem ser utilizados métodos que não envolvem inferência estatística, como o método do dicionário. Este método opera com dicionários predefinidos de palavras associadas a sentimentos ou categorias específicas, eliminando a necessidade de dados de treinamento. Essa abordagem analisa o texto pela presença dessas palavras especificadas, interpretando seu impacto conforme a classificação do dicionário (GENTZKOW *et al.*, 2019).

O método do dicionário se destaca por sua simplicidade e transparência, oferecendo consistência e clareza na classificação de dados. O modelo é simples porque utiliza listas de palavras positivas e negativas para classificar o sentimento dos documentos. Ele é transparente, permitindo aos usuários ver exatamente quais palavras no texto influenciam a classificação e como cada uma contribui para a pontuação final. A consistência é garantida, pois o modelo gera resultados semelhantes quando aplicado repetidamente aos mesmos documentos, proporcionando confiança nas previsões. Além disso, a clareza na classificação de dados é evidente, pois o sentimento expresso em cada documento é apresentado de forma que permite a compreensão rápida do sentimento predominante. Embora não se ajuste automaticamente a novos contextos ou capture nuances fora de seu escopo predefinido, sua previsibilidade e facilidade de interpretação são inestimáveis. Assim, mesmo com o avanço de abordagens mais modernas, os métodos de dicionário continuam sendo valorizados, sendo ferramentas essenciais em análises onde

a estabilidade e a clareza são mais importantes que a adaptabilidade (LOUGHRAN & MCDONALD, 2011).

Pesquisas nessa área tornaram-se indispensáveis para identificar padrões e estabelecer conexões entre as informações, permitindo que os algoritmos aprendam e se aprimorem com base nos dados coletados. Mesmo quando apresentados inicialmente em formato textual, esses dados podem ser tabulados e categorizados, facilitando assim a análise e a geração de previsões futuras com base em dados históricos. Assim sendo, os textos digitais constituem um recurso valioso de informações sobre atividades econômicas e sociais. Ferramentas estatísticas modernas capacitam os pesquisadores a extrair e codificar esses dados de maneira quantitativa, adequada para análises descritivas ou causais. Tanto a disponibilidade de dados textuais quanto o progresso dos métodos estão se expandindo rapidamente, e é esperado que a relevância do texto na pesquisa empírica continue a crescer.

Por isso, a ata do Comitê de Política Monetária (COPOM), órgão do Banco Central do Brasil (BACEN) encarregado de definir a cada 45 dias a taxa básica de juros da economia – a Selic, desempenha um papel crucial na análise do sentimento textual, possibilitando a captura de informações apresentadas de forma indireta e subjetiva pelo BACEN, mesmo que não tenham sido explicitamente mencionadas ao longo do texto. A ata do COPOM figura como um dos principais instrumentos de comunicação do BACEN, fornecendo projeções econômicas para o cenário nacional e internacional, políticas de controle da inflação e decisões relacionadas às taxas de juros. Por meio dela, a autoridade monetária explica os procedimentos adotados na tomada de decisão de política monetária, com o intuito de promover uma comunicação mais transparente e manter as expectativas sob controle (FILHO & ROCHA, 2010).

Correa *et al.* (2017) desenvolveram um índice para avaliar o sentimento de estabilidade financeira com base nos relatórios de vários Bancos Centrais. Os autores identificaram sinais de que o sentimento bancário tende a se deteriorar antes de uma crise bancária. Da mesma forma, Machado e Silva (2017) encontraram evidências de que o ‘tom’ de incerteza nos relatórios de desempenho trimestrais das indústrias influencia a dinâmica do mercado financeiro.

No Brasil, o avanço da temática ainda está em estágio inicial, com alguns trabalhos pioneiros merecendo destaque, tais como Filho (2014), Ferreira *et al.* (2017) e Barboza e Zilberman (2018). Por exemplo, Godeiro e Lima (2017) desenvolveram um Índice de Incerteza Macroeconômica agregada, baseado na volatilidade. Já Montes e Nicolay (2015) conduziram um estudo empírico sobre a influência da comunicação do BACEN nas expectativas de inflação por meio das atas do COPOM, evidenciando que os

especialistas reagem ao ‘tom’ empregado nessas atas, confirmando a eficácia desse instrumento de comunicação do BACEN.

Portanto, este estudo visa analisar as atas do COPOM ao longo de um período de 11 anos, de 2013 a 2023, abrangendo 88 documentos. Utilizou-se uma metodologia de análise textual que incluiu a categorização dos sentimentos expressos nos textos das atas, classificando-os como positivos ou negativos. O objetivo principal foi prever uma variável macroeconômica crucial para o desempenho econômico do Brasil no âmbito nacional e internacional, como a taxa de crescimento econômico, para períodos futuros. O método foi dividido em duas fases principais: (i) a construção de um indicador de predição; e (ii) a aplicação de um modelo de regressão simples junto ao modelo autorregressivo de ordem 2 – AR(2), a fim de examinar a correlação entre as variáveis.

Para alcançar os objetivos delineados, este estudo se organiza da seguinte maneira: além desta introdução e da metodologia previamente exposta, inclui-se um referencial teórico que aborda elementos importantes da Política Monetária, como os modelos de expectativas e os de metas de inflação. Prossegue com a análise textual, dando ênfase ao método do Dicionário no seu uso como ferramenta para melhorar a previsão de crescimento econômico. Além disso, o tema do crescimento econômico é abordado, sendo essencial para a compreensão dos resultados desta pesquisa. Após a metodologia e a descrição da base de dados, uma seção é dedicada à discussão dos resultados, utilizando gráficos e tabelas para elucidar os dados. O estudo conclui com uma seção de conclusões, onde é introduzido um indicador projetado para estimar a taxa de crescimento econômico em períodos futuros.

2 REFERENCIAL TEÓRICO

A evolução do pensamento econômico e o advento das tecnologias digitais têm provocado mudanças significativas nas abordagens utilizadas para entender tanto a macroeconomia quanto o comportamento do consumidor. Em 1976, Robert Lucas revolucionou a macroeconomia com uma crítica influente às metodologias tradicionais de estimativa dos parâmetros de política econômica. Lucas argumentava que as expectativas dos agentes econômicos são racionais, ou seja, eles usam todas as informações disponíveis para formar previsões sobre o futuro, em vez de se basearem apenas em dados históricos. Isso sugeria que as expectativas são dinâmicas e adaptativas, moldadas por novas informações, e não meramente reativas a padrões passados.

No contexto das políticas econômicas, a incerteza sobre o futuro econômico frequentemente resulta na postergação de decisões estratégicas cruciais. Segundo Dixit e

Pindyck (1994), em ambientes caracterizados por incertezas e custos irreversíveis, muitas vezes emerge como mais sensato para os empresários adotar uma postura de “esperar para ver”. Essa abordagem cautelosa é adotada para evitar compromissos com estratégias que possam se revelar desvantajosas diante de mudanças nas condições de mercado ou na política econômica.

Paralelamente, a expansão das redes sociais e a internet têm transformado a maneira como as empresas interagem com os consumidores e analisam seus comportamentos. A mineração textual, como descreve Algaba (2020), tornou-se uma ferramenta valiosa para as empresas que buscam compreender as percepções e as emoções dos consumidores. Esta técnica envolve a análise de grandes volumes de dados textuais para extrair *insights* sobre as atitudes dos consumidores em relação a produtos, serviços e marcas. A análise de sentimentos, uma faceta da mineração textual, oferece uma janela para as reações emocionais dos consumidores, fornecendo informações cruciais que podem ser utilizadas para refinar estratégias de marketing e desenvolvimento de produtos.

Essa convergência de teoria econômica avançada e tecnologia digital está remodelando as estratégias empresariais e políticas econômicas, destacando a importância de adaptar-se continuamente a um ambiente em rápida evolução.

Este referencial teórico aborda as tecnologias utilizadas na pesquisa científica, enfatizando a inovação na decodificação de documentos. O principal objetivo é demonstrar como a análise de sentimentos pode detectar problemas e facilitar decisões rápidas e efetivas, reduzindo impactos adversos e ampliando benefícios. A discussão inclui modelos de expectativas e de metas de inflação, introduzindo a política monetária como contexto para explorar a utilização do índice de sentimentos das atas do COPOM para prever o crescimento econômico. O método de análise textual que terá destaque neste capítulo é o método do dicionário com ênfase especial ao dicionário de Loughran-McDonald. Adicionalmente, o texto examina aspectos do crescimento econômico, buscando criar uma ponte entre esses conceitos e a metodologia de análise textual empregada.

2.1 Política Monetária

Na literatura econômica, a incerteza desempenha um papel significativo, afetando a economia por meio de diversos canais de transmissão que exacerbam seus impactos adversos. Silva (2020) destaca que o temor de um possível *default* financeiro pode levar empresários a adiar decisões críticas de produção e investimento, resultando em demissões e cortes de empregos. Em resposta a esse ambiente incerto, as famílias tendem a reduzir o consumo e aumentar suas poupanças, adotando uma abordagem mais cautelosa

em relação às suas finanças pessoais. A confiança abalada dos agentes econômicos os motiva a investir em ativos considerados mais seguros, buscando minimizar os riscos associados à instabilidade e volatilidade crescentes do mercado. Neste contexto, as instituições financeiras passam a oferecer crédito com maior cautela e seletividade, refletindo um ambiente de aversão ao risco. Conseqüentemente, as finanças públicas e o mercado cambial também são afetados, experimentando uma volatilidade persistente que pode culminar em uma recessão econômica prolongada e profunda.

Keynes (1936), em sua análise sobre a tomada de decisões em contextos de incerteza, argumenta que os indivíduos reagem aos sinais emitidos pelo ambiente econômico. Ele sugere que, ao enfrentarem incertezas, os indivíduos projetam suas expectativas de longo prazo com base na interpretação desses sinais, incorporando essas percepções em suas decisões de investimento. Esta abordagem, conhecida como a teoria das expectativas, propõe que as decisões econômicas são influenciadas não apenas pelas condições atuais, mas também pelas expectativas futuras, refletindo uma compreensão profunda de como a psicologia e a percepção de risco afetam as escolhas econômicas.

Portanto, a compreensão dos efeitos da incerteza na economia exige uma análise que abarque tanto os impactos imediatos nos comportamentos empresarial e consumidor quanto as expectativas de longo prazo que moldam as decisões de investimento. As contribuições de Keynes e os estudos recentes, como os de Silva (2020), fornecem uma base crucial para entender como as políticas monetárias podem ser desenhadas para mitigar os efeitos negativos da incerteza e promover uma recuperação econômica estável.

2.1.1 Modelos de Expectativas

O conceito de expectativas tem desempenhado um papel crucial no desenvolvimento da teoria econômica, particularmente na análise de como os indivíduos antecipam condições futuras e tomam decisões com base nessas previsões. Desde as teorias clássicas, que presumiam um mercado sempre em equilíbrio e com previsões perfeitas, até os complexos modelos de expectativas racionais que modelam como os agentes econômicos utilizam todas as informações disponíveis para formar previsões, a evolução das teorias de expectativas reflete mudanças profundas na maneira como os economistas entendem o comportamento humano e a política econômica. Essa trajetória histórica mostra não apenas um aprimoramento nas técnicas de modelagem econômica, mas também uma mudança na percepção de como as expectativas influenciam a dinâmica macroeconômica e a eficácia da intervenção estatal (LUCAS & SARGENT, 1981).

As expectativas clássicas formam a base do pensamento econômico que dominou até o início do século XX, fundamentadas na teoria clássica que pressupõe que os agentes econômicos são perfeitamente informados e sempre racionais em suas decisões. Segundo esta visão, os agentes conseguem prever o futuro de maneira correta e ajustar seus comportamentos em conformidade, levando a economia sempre a um estado de equilíbrio pleno emprego. Neste modelo, as variáveis macroeconômicas são vistas como previsíveis e as políticas governamentais são consideradas neutras a longo prazo, ou seja, elas não têm impacto real sobre a economia porque os agentes antecipam suas ações e se adaptam a elas (SMITH, 1776).

O modelo de expectativas adaptativas é um conceito crucial na economia comportamental e macroeconômica, atuando como uma ligação entre as teorias clássicas e as contemporâneas sobre expectativas racionais. Esse modelo propõe que os agentes econômicos baseiem suas expectativas de variáveis futuras, como inflação, taxas de juros e crescimento econômico, nas experiências passadas, ajustando-as conforme recebem novas informações. Desenvolvido para superar as limitações das expectativas clássicas, que assumiam previsões perfeitas e instantâneas, o modelo adaptativo considera que os agentes aprendem com os erros anteriores e ajustam suas previsões incrementalmente. Essa abordagem também sugere que políticas econômicas podem ter impactos significativos no curto prazo devido ao tempo necessário para os ajustes nas expectativas. Amplamente utilizado para explicar fenômenos como a persistência inflacionária e a curva de Phillips, o modelo destaca como as expectativas baseadas em experiências passadas moldam as decisões econômicas atuais e futuras (CAGAN, 1956).

O modelo de expectativas racionais é um pilar na teoria econômica moderna, introduzido nos anos 1970 por economistas como John F. Muth e posteriormente popularizado por Robert Lucas e Thomas Sargent. Este modelo assume que os agentes econômicos utilizam todas as informações disponíveis para formar previsões acuradas sobre o futuro, sem cometer erros sistemáticos. De acordo com essa teoria, os agentes antecipam corretamente as consequências de políticas econômicas e ajustam seu comportamento para otimizar seus resultados, limitando assim a eficácia de políticas econômicas previsíveis.

As expectativas racionais revolucionaram a maneira de entender a política macroeconômica, especialmente através da “Crítica de Lucas”, que argumenta que as mudanças na política econômica alterarão as expectativas e, por consequência, os resultados econômicos, tornando modelos baseados em dados históricos ineficazes se não incorporarem as mudanças nas expectativas. Essa visão levou ao desenvolvimento de modelos econômicos que são consistentes com as expectativas racionais, influenciando tanto a teoria quanto a prática da política econômica (SARGENT, 1986).

Os novos keynesianos, emergindo nos anos 1980 com economistas como Gregory Mankiw e David Romer, também aceitam a hipótese de expectativas racionais, mas diferem dos novos clássicos ao incorporar fricções de mercado, como rigidez nos preços e nos salários. Essas fricções permitem que as políticas econômicas tenham efeitos reais, mesmo quando são antecipadas, pois os ajustes de preços não são instantâneos devido a custos de menu e contratos fixos. Os novos keynesianos, portanto, defendem uma política econômica ativa para estabilizar a economia diante de choques e flutuações (MANKIW, 1985).

Portanto, os modelos de expectativas, desde os clássicos até os novos keynesianos, destacam a importância de como as expectativas dos agentes econômicos são formadas e ajustadas, o que é crucial para entender e prever comportamentos econômicos. Fazendo uma relação com o objetivo desta pesquisa, que visa realizar previsões de crescimento econômico utilizando as atas do COPOM por meio da análise textual, a incorporação de expectativas é fundamental. Utilizando a análise textual para extrair índices de sentimentos das atas, pode-se quantificar as expectativas futuras dos formuladores de política monetária e do mercado. Essa abordagem se alinha com os modelos de expectativas racionais e adaptativas, já que considera que os agentes atualizam suas previsões baseados nas informações divulgadas pelos bancos centrais. A eficácia deste método de previsão, portanto, ressalta a importância de entender não apenas as condições econômicas explícitas, mas também as expectativas implícitas e sua dinâmica, o que pode fornecer *insights* para políticas econômicas mais efetivas e para a previsão de tendências econômicas.

2.1.2 Modelos de Metas de Inflação

Os modelos de metas de inflação são uma estratégia de política monetária adotada por diversos bancos centrais ao redor do mundo para controlar a inflação, mantendo-a dentro de uma faixa pré-definida, geralmente estabelecida por metas explícitas de inflação. Este sistema visa proporcionar maior previsibilidade e transparência na condução da política monetária e, por consequência, estabilizar tanto as expectativas de inflação quanto a economia como um todo (BERNANKE *et al.* 1999).

Os modelos de metas de inflação utilizados pelos bancos centrais incorporam várias características chave que fortalecem a gestão da política monetária. Primeiramente, a transparência e credibilidade são essenciais, pois ao definir metas claras de inflação, os bancos centrais não apenas se comprometem com a clareza nas suas intenções, mas também ancoram as expectativas dos agentes econômicos, reforçando a confiança nas políticas implementadas. Em segundo lugar, a responsabilização e a comunicação eficaz são

promovidas através das metas explícitas, que permitem ao público e ao mercado financeiro avaliar de forma mais precisa as ações do banco central, melhorando assim a interação entre o banco e os agentes econômicos (SVENSSON, 2010).

Além disso, a independência do banco central é crucial, visto que a definição de metas de inflação permite que ele opere sem interferências políticas, focando no controle da inflação a longo prazo. Por último, o ajuste das taxas de juros serve como o principal mecanismo para alcançar as metas estabelecidas, influenciando a atividade econômica e ajustando a demanda agregada para manter a inflação dentro dos limites desejados. Essas características fazem dos modelos de metas de inflação uma ferramenta efetiva na estabilização da economia e no controle inflacionário (MISHKIN, 1999).

O Brasil adotou formalmente o regime de metas de inflação em 1999, após uma série de crises econômicas que destacaram a necessidade de uma política monetária mais estável e previsível. Este regime é considerado um pilar central da política monetária brasileira e tem sido crucial na estabilização da economia. Este modelo foi adotado para ancorar as expectativas inflacionárias após a transição do Brasil para um regime de câmbio flutuante. A meta de inflação é definida pelo Conselho Monetário Nacional (CMN), com o Banco Central responsável por alcançá-la principalmente através da manipulação da taxa básica de juros, a Selic (MISHKIN & SCHMIDT-HEBBEL, 2007).

O regime brasileiro caracteriza-se pela definição anual de uma meta central para a inflação, com intervalos de tolerância. Caso a inflação ultrapasse os limites deste intervalo, o presidente do Banco Central deve escrever uma carta aberta ao Ministro da Fazenda explicando as razões do descumprimento, as medidas para que a inflação retorne ao intervalo e o tempo esperado para que isso ocorra (BACEN, 2024).

Desde que foi instituído, o regime de metas de inflação no Brasil estabelece anualmente um objetivo central para a inflação, complementado por uma margem de tolerância que confere uma certa margem de manobra. O Conselho Monetário Nacional (CMN) é responsável por definir essa meta para três anos à frente, assegurando previsibilidade e transparência na política monetária. Por exemplo, a meta para 2023 foi fixada em 3,25%, com uma tolerância de 1,5% para mais ou para menos, enquanto para 2024 e 2025, a meta foi estabelecida em 3%. Caso a inflação ultrapasse esse intervalo de tolerância ao final do ano, o presidente do BACEN deve explicar publicamente o motivo do descumprimento através de uma carta aberta ao Ministro da Fazenda, presidente do CMN. Nesta carta, devem constar as causas detalhadas do desvio, as medidas adotadas para reconduzir a inflação aos limites estipulados e o prazo esperado para que estas medidas surtam efeito (BACEN, 2024).

Além disso, o regime enfatiza a modelagem de expectativas futuras, tanto do

mercado quanto internas ao Banco Central, para a tomada de decisões de política monetária. Essa modelagem se vale de modelos econométricos que consideram variáveis como taxas de câmbio, juros internacionais e projeções de inflação futura, permitindo que o Banco Central antecipe tendências e ajuste proativamente suas intervenções para manter a inflação dentro dos limites desejados.

Os modelos de metas de inflação são fundamentais para a estabilidade econômica, pois oferecem um mecanismo claro e transparente para o controle da inflação, ajudando a ancorar as expectativas dos agentes econômicos e a manter a inflação dentro de limites previsíveis. Esta previsibilidade é crucial para o planejamento a longo prazo tanto por parte dos consumidores quanto das empresas, impactando positivamente o investimento e o consumo. Além disso, ao adotar um regime de metas de inflação, os bancos centrais podem transmitir sinais claros sobre suas intenções futuras, o que é vital para a eficácia da política monetária. A integração entre metas de inflação e análise prospectiva fortalece a capacidade dos formuladores de políticas e analistas de antecipar desafios e ajustar estratégias para fomentar um ambiente econômico estável e propício ao crescimento.

As Atas do COPOM são documentos cruciais dentro do regime de metas de inflação adotado pelo país. Elas são publicadas após cada reunião do COPOM e fornecem uma visão detalhada das deliberações e decisões tomadas pelo comitê em relação à política monetária. Estas atas são fundamentais para a transparência e para a gestão das expectativas de mercado, pois elucidam os raciocínios por trás das decisões sobre a taxa de juros e outras medidas de controle monetário.

A inserção dessas atas no modelo de metas de inflação é vital, pois elas servem como um meio de comunicação direto entre o Banco Central e o público, incluindo investidores, analistas e formuladores de políticas. Ao detalhar os fatores considerados nas decisões de política monetária, as atas ajudam a moldar as expectativas do mercado quanto à futura condução da política monetária e, conseqüentemente, à trajetória esperada da inflação. Isso está alinhado com o objetivo central do regime de metas de inflação, que é ancorar as expectativas de inflação ao redor de uma meta definida pelo CMN, contribuindo assim para a estabilidade de preços e para o crescimento econômico sustentável. Ao manter e reforçar essa comunicação clara e consistente, as atas do COPOM desempenham um papel integral na eficácia do regime de metas de inflação no Brasil.

2.2 Análise Textual

Análise textual engloba uma série de técnicas e ferramentas computacionais que transformam textos em dados quantificáveis, permitindo extrair e interpretar significados complexos de grandes volumes de texto, facilitando a identificação de padrões e conexões que muitas não são perceptíveis sem ajuda de tecnologia. A capacidade de analisar detalhadamente e extrair *insights* de textos torna a análise textual uma ferramenta crucial na era digital, oferecendo perspectivas únicas que vão além das capacidades humanas tradicionais de análise (LIU, 2015).

O Processamento de Linguagem Natural (PLN) facilita a interação entre humanos e máquinas, capacitando computadores a compreender e processar a linguagem humana através da análise de textos. Este campo permite que os sistemas realizem análises sintática, morfológica, semântica e léxica, transformando eficazmente dados qualitativos em valores numéricos. Um dos métodos utilizados para essa transformação é a técnica de *bag-of-words*, onde as palavras são convertidas em vetores, permitindo a manipulação e análise de grandes conjuntos de dados textuais (SILVA, 2020).

A análise textual, enquanto ferramenta para compreensão de grandes volumes de dados escritos, enfrenta desafios significativos que podem comprometer sua eficácia. Um dos principais obstáculos é a ambiguidade linguística, pois a linguagem humana contém homônimos, ironias e outras nuances que dificultam a interpretação correta pelos algoritmos. Além disso, a dependência contextual das palavras, onde o significado pode alterar conforme o ambiente textual, exige o uso de modelos avançados capazes de discernir esses contextos complexos (MIKOLOV *et al.*, 2013). Adicionalmente, a diversidade de dados apresenta outro desafio, visto que a performance dos algoritmos pode variar entre diferentes idiomas e conjuntos de dados, necessitando ajustes ou treinamentos específicos para cada situação.

Apesar dos desafios inerentes à análise textual, suas vantagens são consideravelmente significativas. Combinado a análise de sentimentos, que visa compreender e classificar as emoções expressas em textos, essa ferramenta se destaca pela sua capacidade de escalabilidade. Isso permite o processamento rápido e eficiente de grandes volumes de dados textuais, oferecendo *insights* valiosos para diversas aplicações.

Embora exista o método de Machine Learning (ML), que vem evoluindo significativamente desde a definição inicial de Mitchell em 1997 e é notável por sua capacidade de melhorar automaticamente seu desempenho através da experiência, no contexto de Big Data, ML facilita a criação de modelos que processam grandes volumes de dados, essenciais para a análise de sentimentos em textos digitais. Principais modelos

incluem Árvores de Regressão, que particionam dados em subgrupos para modelar relações não lineares; Deep Learning, que utiliza redes neurais profundas para tarefas complexas; e Naive Bayes, um classificador baseado em probabilidade reconhecido por sua simplicidade e eficácia. Embora poderosos, esses modelos exigem muitos dados históricos e podem ser complexos demais para interpretar textos técnicos, além de atuarem como “caixas-pretas” difíceis de entender, o que pode ser uma desvantagem em análises econômicas. A adaptação desses modelos a novos dados ou mudanças econômicas pode ser lenta, necessitando de intervenções manuais significativas, comprometendo a agilidade nas análises de política econômica. Neste tópico da revisão bibliográfica, será examinado o método do Dicionário, detalhando suas técnicas e principais características, destacando-se por sua importância na abordagem do tema em questão.

2.2.1 Método do Dicionário

O método do dicionário é uma técnica de análise textual amplamente utilizada em diversas áreas de estudo, incluindo a pesquisa econômica. Essa abordagem se baseia no uso de dicionários predefinidos que contêm palavras-chave associadas a sentimentos, conceitos ou categorias específicas. O principal objetivo do método é identificar e quantificar a presença dessas palavras-chave em textos para extrair significados ou avaliar o contexto em que estão inseridos (GENTZKOW *et al.*, 2019).

A importância do método do dicionário reside na sua simplicidade e transparência. Ao contrário dos complexos modelos de *Machine Learning (ML)*, o método do dicionário permite uma interpretação direta e compreensível dos resultados, o que é crucial em campos como a economia, onde os tomadores de decisão precisam de análises claras e fundamentadas para formular políticas. Além disso, essa técnica não requer um conjunto de dados de treinamento volumoso, tornando-a acessível e prática para situações onde os dados são escassos ou de difícil coleta.

No contexto de previsão de crescimento econômico através das atas do COPOM, por exemplo, o método do dicionário pode ser particularmente útil para detectar mudanças nas perspectivas econômicas expressas nos textos. Ao identificar termos relacionados a riscos, expectativas e outros indicadores econômicos, os analistas podem obter *insights* valiosos sobre a direção futura das políticas monetárias e suas implicações para a economia. Assim, o método do dicionário oferece uma ferramenta robusta e eficiente para complementar as análises econômicas, contribuindo significativamente para estudos que exigem precisão e clareza na interpretação dos dados textuais.

2.2.1.1 Índice de Sentimentos

A técnica de análise de sentimento, uma metodologia sofisticada dentro do campo de Análise Textual e Processamento de Linguagem Natural (PLN), tem ganhado destaque nos últimos anos. Esta técnica visa identificar e categorizar as opiniões expressas em textos para determinar a polaridade dos sentimentos do usuário sobre um tópico específico, classificando-os como positivos, negativos ou neutros. Esta classificação é essencial para entender como as opiniões e os sentimentos influenciam as decisões de compra, as perspectivas políticas e as interações sociais. Empresas, por exemplo, utilizam a análise de sentimento para monitorar a recepção de produtos ou campanhas, enquanto cientistas políticos a empregam para sondar a opinião pública sobre temas governamentais (PANG & LEE, 2008).

A análise de sentimentos enfrenta ainda desafios como ambiguidades linguísticas e variações culturais, que podem distorcer a interpretação dos sentimentos manifestados. Assim, a técnica demanda uma evolução contínua das ferramentas e métodos para manter sua precisão e relevância no tratamento de textos complexos.

Ampliando o uso da análise de sentimentos para o campo econômico, estudos recentes se concentraram na criação de índices de sentimentos a partir de textos de relatórios econômicos, como os do Banco Central da Inglaterra. Tais índices, que são utilizados como preditores de variáveis macroeconômicas importantes, como taxa de inflação e Produto Interno Bruto (PIB), ilustram como a análise textual pode ser integrada a modelos econômicos mais complexos para aprimorar previsões. Os trabalhos de Jones, Sinclair e Stekler (2019); Clements e Reade (2020) exemplificam bem isso ao mostrar que a análise de sentimentos em textos dos relatórios de inflação pode refletir a conjuntura econômica e, potencialmente, enriquecer as previsões econômicas, contribuindo para uma compreensão mais aprofundada das dinâmicas macroeconômicas.

Nos últimos anos, a transformação de textos em variáveis numéricas tem permitido seu uso como preditores em diversos modelos de pesquisa. Por exemplo, Garcia (2013) observou que o sentimento extraído de textos do *New York Times* pode prever de forma eficaz os retornos das ações, especialmente durante recessões. De forma similar, Engelberg e Parsons (2011) identificaram que a cobertura de mídia local influencia significativamente o volume de negociações de ações locais, sugerindo que a mídia pode desempenhar um papel crucial na previsão dos retornos de investimentos.

A metodologia aplicada por Lima *et al.* (2019) propõe ajustar dinamicamente o conteúdo dos dicionários com base no poder preditivo das palavras, utilizando uma abordagem de três etapas. Primeiro, gera-se um vetor de séries temporais que transforma

palavras em valores numéricos sem um dicionário fixo e requer redução de dimensionalidade devido à sua complexidade. Em seguida, emprega-se aprendizado de máquina supervisionado para selecionar as palavras mais preditivas que formam a base de um dicionário dinâmico usado para construir preditores. Por fim, esses preditores são usados para realizar previsões fora da amostra, com o processo sendo repetido recursivamente, permitindo que o dicionário evolua com o tempo e melhore a precisão das previsões. Esta metodologia se destaca pela adaptabilidade e capacidade de refletir mudanças dinâmicas no uso das palavras, fornecendo uma ferramenta poderosa para análises preditivas. O aprendizado supervisionado é o caso clássico de modelagem de y usando entradas (preditores) x , no qual escolhe-se um algoritmo de aprendizado que se ajustará ao alvo usando as características de entrada fornecidas.

2.2.1.2 Dicionário Loughran-McDonald

Na análise de sentimentos, uma técnica destacada é a utilização de *lexicons*, estruturas semelhantes a dicionários que compreendem coleções de palavras e expressões associadas a sentimentos específicos. Conforme Turney (2002) descreve, nesta abordagem, um texto é analisado e classificado com base na correspondência das palavras presentes no documento com aquelas listadas no *lexicon*. Isso facilita a determinação da polaridade geral do texto — positiva, negativa ou neutra — através da avaliação das conotações dos termos identificados.

Este método amplamente utilizado nas ciências sociais não recorre a inferência estatística e se baseia em um conjunto pré-definido de termos que categorizam aspectos do texto para identificar sentimentos. Esta abordagem é complementada por métodos como a regressão de texto, que estima a distribuição dos resultados condicionais e a modelagem criativa de causalidade entre variáveis, conforme explicado por Grimmer e Stewart (2013). Esses métodos são intuitivos e interpretáveis, suportados por softwares capazes de processar grandes volumes de dados.

Shapiro, Sudhof e Wilson (2020) discutem duas metodologias principais para a quantificação de sentimentos em textos. A primeira, a metodologia lexical, opera com léxicos ou dicionários que atribuem pontuações a palavras baseadas nas emoções que evocam. Esta técnica, conhecida como “*bag-of-words*” (*BOW*), foca na frequência de palavras com polaridades específicas dentro do texto, desconsiderando o contexto mais amplo em que as palavras ocorrem. Tal abordagem simplifica a análise, mas pode não capturar toda a complexidade e nuances dos sentimentos expressos, especialmente em textos onde aspectos sutis da linguagem são cruciais.

Assim, enquanto a metodologia lexical oferece uma ferramenta para análises preliminares de sentimentos, os desafios associados à alta dimensionalidade dos dados e a necessidade de abordagens mais sofisticadas, como os modelos lineares penalizados sugeridos por Hastie *et al.* (2020), destacam a importância de continuar evoluindo as técnicas de análise para aprimorar a precisão e a relevância das interpretações. Em conjunto, essas metodologias fornecem um conjunto robusto de ferramentas para explorar e entender as dinâmicas emocionais contidas nos textos.

O Dicionário Loughran-McDonald, desenvolvido por Tim Loughran e Bill McDonald, distingue-se de léxicos convencionais ao se concentrar especificamente no jargão de relatórios financeiros e empresariais. Este dicionário classifica as palavras em categorias relativas a sentimentos financeiros, como positividade, negatividade, incerteza e litígio, o que o torna particularmente valioso para a análise de documentos como relatórios anuais e comunicados de imprensa de empresas, além de transcrições de conferências de resultados (LOUGHRAN & MCDONALD, 2011).

Analistas empregam este dicionário para medir o tom dos relatórios anuais de companhias, detectando a presença de palavras que indicam sentimentos negativos ou positivos. O dicionário também é amplamente utilizado por pesquisadores para analisar transcrições de chamadas de resultados e antecipar as reações do mercado, por exemplo, a identificação de um aumento nos termos associados a “incerteza” pode indicar uma possível reação negativa do mercado. Além disso, o dicionário é frequentemente aplicado em estudos acadêmicos para explorar como o sentimento nos textos se correlaciona com variáveis econômicas, tais como retornos de ações, volatilidade do mercado ou indicadores macroeconômicos.

Portanto, optou-se pelo método do dicionário para esta pesquisa devido à sua abordagem transparente e direta na análise de texto, essencial para compreender claramente como termos e conceitos específicos são empregados no contexto das políticas monetárias. Esta metodologia possibilita uma interpretação precisa e objetiva dos dados, simplificando a comunicação dos resultados para *stakeholders* e decisores políticos. Utilizando dicionários especializados, como o Dicionário de Loughran-McDonald, que contém termos econômicos e financeiros pertinentes, somos capazes de detectar eficientemente nuances e ênfases nas comunicações do COPOM. Isso é crucial para identificar alterações nas perspectivas econômicas que podem não ser evidentes por meio de métodos quantitativos tradicionais. Dessa forma, o método do dicionário proporciona uma combinação perfeita de simplicidade, transparência e eficácia, alinhando-se aos objetivos de nossa pesquisa ao prever tendências e mudanças no crescimento econômico com base nas atas do COPOM.

2.3 Crescimento Econômico

O crescimento econômico representa um pilar central na economia, refletindo a capacidade de uma nação de ampliar a produção de bens e serviços ao longo do tempo. Influenciado por uma ampla gama de fatores — desde políticas governamentais e condições de mercado até inovações tecnológicas e mudanças demográficas —, esse crescimento é crucial para elevar o padrão de vida e garantir a estabilidade e o desenvolvimento socioeconômico de um país. Compreender como esses fatores interagem para fomentar o crescimento é vital para a formulação de políticas públicas eficazes.

Investimentos em capital físico e humano são tradicionalmente reconhecidos como motores essenciais do crescimento econômico. Conforme Solow (1956) demonstrou, a produtividade de uma economia pode ser substancialmente aumentada através da acumulação de capital e do avanço tecnológico, que juntos impulsionam o crescimento do produto per capita. Contudo, a eficácia desses investimentos pode variar significativamente em função do ambiente econômico e político vigente. Por exemplo, Bernanke (1983) ressaltou como a incerteza durante a Grande Depressão exacerbou a crise econômica, ao desencorajar os empresários de fazerem novos investimentos.

Além disso, a incorporação de técnicas modernas, como a mineração de texto e a análise de sentimentos, tem trazido novas dimensões à compreensão do crescimento econômico. A utilização de modelos de análise textual para analisar as atas do COPOM, por exemplo, pode revelar nuances sobre a evolução do sentimento econômico e prever futuras tendências econômicas. Essas técnicas oferecem uma análise mais detalhada e imediata das expectativas que influenciam as decisões econômicas, um aspecto crucial em um mundo onde as informações são vastas e se modificam rapidamente.

Portanto, enquanto a abordagem tradicional para estudar o crescimento econômico permanece fundamental, a integração de novas metodologias analíticas é essencial para capturar a complexidade das expectativas e dos sentimentos dos agentes econômicos. A combinação de análise textual com estudos econômicos não apenas enriquece nossa compreensão sobre o crescimento econômico, mas também abre caminhos promissores para a elaboração de políticas mais informadas e efetivas, proporcionando ferramentas robustas para economistas e formuladores de políticas enfrentarem o desafio contínuo de promover um crescimento econômico sustentável.

3 METODOLOGIA

Neste capítulo, são expostos os instrumentos metodológicos empregados no desenvolvimento deste estudo, abrangendo a coleta de dados e o processo de cálculo e estimação do sentimento textual. A metodologia foi organizada em duas etapas principais: na primeira, são descritas as fases de construção do indicador de predição; na segunda, é detalhado o uso desse indicador na análise de correlação por meio do modelo de regressão simples e do AR(2).

Todas as informações necessárias para a elaboração do indicador foram extraídas das atas do COPOM, disponíveis no portal do BACEN (2024). Foram utilizadas as atas em inglês, abrangendo um período de 11 anos, de janeiro de 2013 a dezembro de 2023. Durante esse período, as reuniões ocorrem a cada 45 dias, resultando na divulgação de 8 atas por ano. Portanto, considerando todo esse período, foram analisadas 88 atas nesta pesquisa.

Conforme apontado por Filho e Rocha (2010) a ata do COPOM figura como um dos principais meios de comunicação do BACEN, fornecendo projeções econômicas tanto para o cenário nacional quanto internacional, além de abordar questões como o controle da inflação e as decisões relacionadas às taxas de juros. Através dela, a autoridade monetária detalha os procedimentos adotados na tomada de decisão de política monetária, visando a transparência na comunicação e a manutenção das expectativas sob controle.

3.1 Construção do Indicador de sentimento para Predição

O objetivo desta etapa do estudo é desenvolver um indicador que avalie o sentimento de incerteza relacionado à política econômica presente nas atas das reuniões do COPOM, seguido pela avaliação de sua capacidade de influenciar as variáveis macroeconômicas, sobretudo o crescimento econômico.

Para alcançar esse objetivo, foi necessário seguir uma sequência de procedimentos, conforme ilustrado na Figura 1, que abrange desde a coleta de dados até a criação do Índice de Sentimento Loughran-McDonald (ILM).

Figura 1 – Fases da etapa de construção do indicador.



Fonte: Elaboração do autor.

Os dados foram coletados acessando as atas do COPOM em inglês por meio da Interface de Programação de Aplicações (API) utilizando o *Rstudio*. Após confirmar o sucesso da solicitação, os dados em formato *JSON* foram transformados em um *dataframe*, estabelecendo-se o período da análise de 2013 a 2023. Durante este processo, foi necessário converter os formatos dos dados para facilitar a extração do texto das atas. Isso envolveu a remoção de elementos não essenciais para a análise, como *URLs*, *hashtags*, caracteres de controle, espaços extras, e outros caracteres especiais.

Após a limpeza inicial, procedeu-se à criação do '*Corpus*', uma coleção de textos usada como base para a análise exploratória. Antes de realizar qualquer contagem de palavras, preprocessamos o texto bruto em várias etapas. O objetivo é reduzir o vocabulário para um conjunto de palavras que sejam principalmente significativas. Essa fase incluiu a remoção de espaços desnecessários, pontuações, a conversão de todas as palavras para letras minúsculas, a eliminação de números, caracteres especiais adicionais, e palavras comuns (*stopwords*). Em seguida, foi desenvolvida uma matriz de frequência dos termos usando o pacote específico Dicionário Loughran-McDonald). A matriz de frequência gerada pelo *Rstudio*, após a execução do código, contém 62.194 observações de 3 variáveis: o

termo, o documento e a frequência do termo no documento. Na figura abaixo, temos uma amostra de 10 termos mais frequentes da Ata de número 172 do mês de Janeiro de 2013.

A Tabela 1 apresenta uma amostra que lista as 30 palavras extraídas da Ata 1 do COPOM de 2013, todas com relevância econômica. A tabela inclui a frequência de cada palavra no documento e indica se a conotação associada a cada termo é positiva ou negativa. Esta análise nos permite visualizar tanto a prevalência de certos termos quanto a natureza dos sentimentos que eles transmitem, o que pode oferecer *insights* sobre o tom e as preocupações econômicas expressas na ata.

Tabela 1 – Classificação das palavras da Ata nº 172 (Jan./2013).

Nº	Palavra	Documento	Nº de Ocorrências	Sentimento
1	adverse	1	2	negative
2	better	1	1	positive
3	caution	1	3	negative
4	challenges	1	1	negative
5	challenging	1	1	negative
6	concern	1	1	negative
7	despite	1	2	positive
8	downward	1	1	negative
9	easing	1	1	negative
10	gains	1	2	positive
11	greater	1	3	positive
12	harmful	1	1	negative
13	improvement	1	1	positive
14	lack	1	1	negative
15	lagged	1	1	negative
16	leading	1	1	positive
17	persist	1	1	negative
18	persistence	1	2	negative
19	positively	1	1	positive
20	progress	1	1	positive
21	prolonged	1	1	negative
22	recalled	1	1	negative
23	slowdown	1	2	negative
24	slower	1	1	negative
25	smoothing	1	1	positive
26	stability	1	1	positive
27	stabilization	1	1	positive
28	strengthening	1	1	positive
29	strong	1	1	positive
30	stronger	1	1	positive

Fonte: Gerada pelo *Rstudio* após a execução do código. Elaboração do autor.

A análise das palavras presentes nas atas do COPOM revela uma divisão entre termos com conotações positivas e negativas, refletindo o tom e as nuances das discussões econômicas. Palavras negativas como "adverse", "caution", "challenges", entre outras, indicam dificuldades, riscos ou declínios, sugerindo períodos de preocupação ou condições

adversas. Em contraste, palavras positivas como "better", "gains", e "improvement" apontam para progresso, fortalecimento ou melhorias, o que pode refletir uma perspectiva otimista sobre os desenvolvimentos econômicos abordados. Além disso, termos como "despite" mostram usos contextuais que podem sugerir superação de adversidades. A predominância de palavras de uma ou outra natureza pode oferecer uma visão geral sobre o clima predominante nas discussões do COPOM. Essa análise de sentimentos é crucial para investidores, políticos e analistas que buscam compreender as tendências da política monetária e suas implicações econômicas, ajudando a antecipar decisões e ajustar estratégias com base no contexto apresentado nas atas.

Na etapa final deste processo, procedemos com a criação do Índice de Sentimento Loughran-McDonald (ILM) para cada ata. A aplicação do dicionário ocorre da seguinte forma, cada palavra no texto processado é verificada contra o dicionário de sentimentos. Se a palavra estiver no dicionário, sua pontuação de sentimento correspondente é considerada; caso contrário, pode ser ignorada ou tratada como neutra. O índice de sentimentos de um documento é geralmente calculado como a soma das pontuações de todas as palavras sentimentais no texto. Em alguns casos, a pontuação total pode ser normalizada pelo número de palavras no texto para obter uma média, o que pode fornecer uma comparação mais justa entre textos de tamanhos diferentes. O cálculo do índice de sentimentos consiste na equação (1) que proporciona uma visão rápida do equilíbrio geral entre expressões positivas e negativas em um texto. Essa métrica pode ser especialmente útil em análises de documentos onde o sentimento geral (positivo ou negativo) é mais relevante do que a intensidade ou a nuance específica do sentimento expresso.

$$ILM = \frac{n^{\circ}(+) - n^{\circ}(-)}{N} \quad (1)$$

Onde:

- ILM é o índice de sentimento;
- $n^{\circ}(+)$ é o número de ocorrências de palavras positivas;
- $n^{\circ}(-)$ é o número de ocorrências de palavras negativas;
- N é o total de palavras positivas e negativas no documento.

Para calcular o índice de sentimentos em um texto, primeiramente identificam-se e contam-se as ocorrências de palavras com conotações positivas e negativas usando um dicionário de sentimentos pré-definido. As palavras são categorizadas como positivas ou negativas, e suas frequências são computadas. Em seguida, subtrai-se o número de palavras negativas das positivas para obter uma medida bruta do saldo de sentimentos, indicando se

o texto tem uma tonalidade mais positiva ou negativa. Para adequar a análise a textos de diferentes tamanhos, a diferença é normalizada pelo número total de palavras no documento, resultando em uma proporção que reflete o predomínio dos sentimentos (HUTTO & GILBERT, 2014).

Na Tabela 2 apresenta os índices de sentimentos para as atas publicadas nos anos de 2013 e 2014. Observa-se que algumas atas apresentam sentimentos negativos, enquanto outras exibem sentimentos positivos. Essa fase do método foi projetada especificamente para desenvolver um índice que determina a polaridade dos sentimentos em cada documento.

Tabela 2 – Índice de sentimento dos anos de 2013 e 2014.

Nº	ILM	Nº da Ata	Publicação
1	-0,18447	172	Jan. / 2013
2	-0,01961	173	Mar. / 2013
3	-0,06667	174	Abr. / 2013
4	-0,15966	175	Mai. / 2013
5	-0,11828	176	Jul. / 2013
6	-0,11340	177	Ago. / 2013
7	0,05155	178	Out. / 2013
8	-0,15464	179	Nov. / 2013
9	0,03448	180	Jan. / 2014
10	0,11111	181	Fev. / 2014
11	0,13333	182	Abr. / 2014
12	0,10000	183	Mai. / 2014
13	0,11111	184	Jul. / 2014
14	0,16364	185	Set. / 2014
15	0,15789	186	Out. / 2014
16	0,22034	187	Dez. /2014

Fonte: Gerada pelo *Rstudio* após a execução do código. Elaboração do autor.

Cada fase deste processo foi essencial para identificar as palavras mais utilizadas nas atas do COPOM e avaliar sua polaridade. Essas informações são fundamentais para a próxima etapa da pesquisa, que inclui a aplicação de regressão simples e do modelo AR(2), para realizar a análise de correlação. Neste contexto, selecionou-se uma variável macroeconômica considerada crucial para avaliar o desempenho econômico de um país. A variável escolhida foi a taxa de crescimento econômico.

Para obter os dados necessários, foi essencial acessar o portal do Instituto Brasileiro de Geografia e Estatística – IBGE (2024) e consultar o banco de dados que contém os valores da taxa de crescimento econômico, publicados trimestralmente, referentes ao período de estudo de 2013 a 2023.

3.2 Análise de Correlação (regressão simples) e modelo autorregressivo (AR-2)

Nesta etapa da metodologia, utilizou-se a regressão linear simples para análise de correlação e o modelo AR(2), derivado das atas do COPOM, e a variável macroeconômica escolhida, tal seja, a taxa de crescimento econômico.

A regressão linear simples é um método para modelar a relação entre a taxa de crescimento econômico (y) e o índice de sentimento (ILM) através de uma equação linear (HYNDMAN & ATHANASOPOULOS, 2018). A equação (2) é a fórmula geral para a regressão linear simples.

$$y_t = \beta_0 + \beta_1 ILM_{t-h} + \epsilon_t \quad (2)$$

Onde:

- y_t é a taxa de crescimento econômico no tempo t ;
- ILM_{t-h} é o índice de sentimento no tempo t com defasagem h ;
- β_0 é o intercepto da linha de regressão;
- β_1 é o coeficiente da inclinação;
- ϵ_t é o termo de erro, representando a variação em y não explicada pelo modelo linear.

O modelo AR(2) é um tipo de modelo de série temporal que usa as duas observações defasadas da taxa de crescimento econômico para prever o valor atual (LIMA *et al*, 2019). A equação (3) para o modelo AR(2) está representada abaixo.

$$y_t = c + \beta ILM_{t-h} + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \epsilon_t \quad (3)$$

Onde:

- y_t é o valor atual da taxa de crescimento econômico da série no tempo t ;
- y_{t-1} é o valor da série no tempo $t - 1$ (a primeira defasagem);
- y_{t-2} é o valor da série no tempo $t - 2$ (a segunda defasagem);
- c é uma constante (intercepto);
- ILM_{t-h} é o índice de sentimento no tempo t com defasagem h ;
- ϕ_1 e ϕ_2 são os coeficientes para as defasagens primeiro e segundo, respectivamente;
- ϵ_t é o termo de erro no tempo t , assumindo como ruído branco.

Ambos os modelos são fundamentais em diferentes contextos: a regressão linear simples é útil para relações diretas entre variáveis, enquanto o modelo AR(2) é adequado para análise de dados temporais onde os valores passados influenciam os futuros. Nesta etapa serão testados vários lag (0, 1 e 2) e realizar o teste de significância, pois cada coeficiente testado dirá se há uma relação estatisticamente significativa entre o lag correspondente da taxa de crescimento econômico e o seu valor atual y_t .

Lima *et al.* (2019) destacam que a literatura propõe uma ampla gama de modelos de previsão de crescimento econômico, incluindo desde simples modelos autorregressivos lineares (AR) até métodos mais complexos como os baseados no modelo de equilíbrio geral dinâmico estocástico (DSGE). Optaram pelo modelo AR(2) como o principal referencial a ser superado pelo método proposto por eles, inspirados pelos estudos de Chauvet e Potter (2013). Estes autores analisaram o desempenho de diversos modelos econométricos para prever o crescimento do produto nos Estados Unidos das Américas - EUA de 1992 a 2010, concluindo que o modelo AR(2) se sobressai durante períodos de expansão e apresenta bom desempenho em recessões. Esses resultados corroboram com as evidências de Nelson (1972), que já indicavam a eficácia dos modelos autorregressivos simples em previsões de crescimento econômico. Essas constatações fundamentaram a escolha do modelo de regressão adotado nesta fase da metodologia.

Da mesma forma que no estudo realizado por Lima *et al.* (2019), o foco principal desta pesquisa não é a criação de um novo modelo definitivo para previsão do crescimento econômico. Em vez disso, o objetivo é integrar e empregar de maneira inovadora técnicas de análise de texto e aprendizado de máquina supervisionado. Esta metodologia facilita uma análise mais profunda das atas dos Bancos Centrais, permitindo a identificação de informações altamente preditivas. Tal abordagem fornece uma ferramenta para economistas e formuladores de políticas, auxiliando na interpretação de documentos textuais complexos e melhorando a tomada de decisões econômicas baseadas em evidências mais precisas.

Utilizando o ambiente *RStudio*, os dados do índice de sentimentos e as taxas de crescimento econômico foram analisados através de código específico para executar tanto a regressão linear simples quanto o modelo AR(2). Assim, foi possível determinar qual modelo é mais eficaz para prever o crescimento econômico e avaliar a significância da análise textual na previsão de tendências futuras.

4 RESULTADOS E DISCUSSÃO

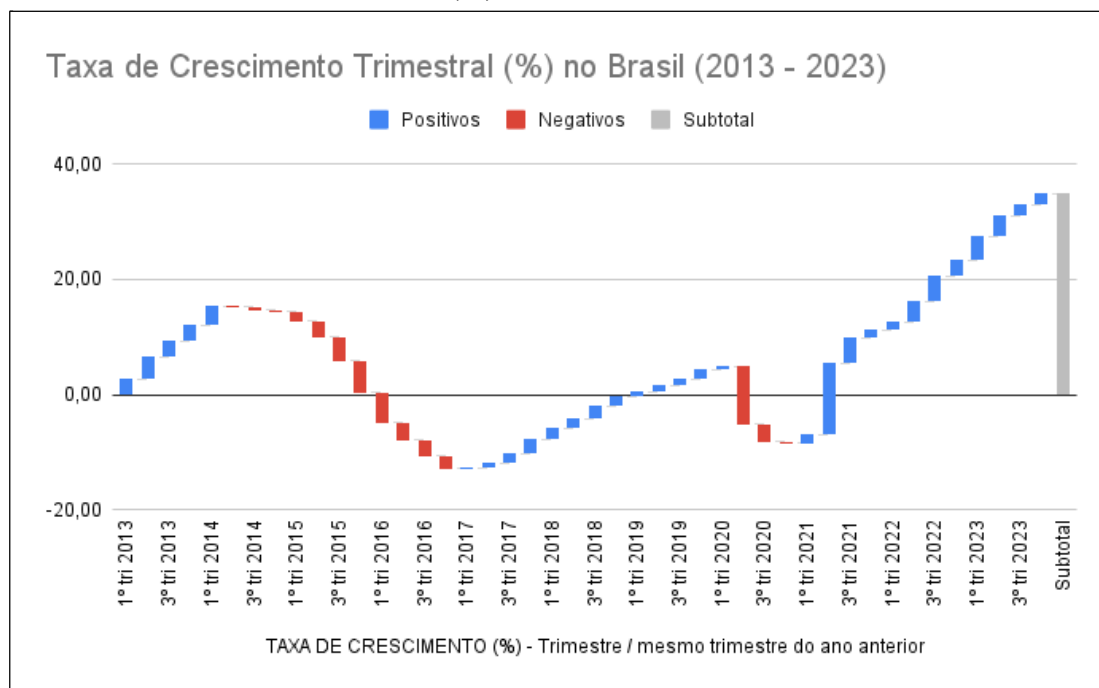
Nesta seção, apresentaremos os resultados das duas etapas metodológicas, enfocando a interação entre o índice de predição desenvolvido na primeira fase e a variável macroeconômica selecionada. Destacaremos o nível de significância dos modelos de regressão utilizados na segunda fase, visando determinar a eficácia desses modelos na previsão do crescimento econômico futuro. Além disso, analisaremos o índice de sentimento derivado das atas do COPOM, obtido por meio da análise textual utilizando o método de dicionário.

O Índice de Sentimento Loughran-McDonald (ILM) derivado das atas do COPOM foi analisado ao longo de 88 atas correspondentes ao período estudado, envolvendo três variáveis: o próprio ILM, que indica a polaridade do sentimento, com 27 resultados positivos e 61 negativos; o número de cada ata; e a data da ata, considerando que durante cada ano do período analisado foram publicadas oito atas, uma para cada reunião do COPOM, que ocorre a cada 45 dias.

A variável macroeconômica escolhida para ser incorporada ao modelo de correlação foi a taxa de crescimento econômico, retirada das séries históricas disponíveis no banco de dados do portal do IBGE. Devido à existência de diferentes tipos de taxas de crescimento, optou-se por selecionar duas variantes para a aplicação dos modelos de regressão: a taxa de crescimento ano a ano (trimestre comparado ao mesmo trimestre do ano anterior) e a taxa de crescimento trimestral ajustada sazonalmente (trimestre comparado ao trimestre imediatamente anterior). A taxa de crescimento ano a ano é ideal para análises de evolução econômica em uma base anual, pois minimiza o impacto das variações sazonais. Em contrapartida, a taxa de crescimento trimestral ajustada sazonalmente é crucial para o monitoramento ágil e ajuste de políticas econômicas, fornecendo dados atualizados e sensíveis às mudanças nas condições econômicas. Ambas as taxas foram empregadas nos modelos devido à sua relevância e utilidade complementar, frequentemente combinadas para oferecer uma visão detalhada da saúde econômica de uma região ou país.

Nos Gráficos 1 e 2 estão apresentadas, respectivamente, a taxa de crescimento trimestral / mesmo trimestre do ano anterior e a taxa de crescimento trimestral / trimestre imediatamente anterior (com ajuste sazonal).

Gráfico 1 – Taxa de Crescimento (%) Trimestre / mesmo trimestre do ano anterior.

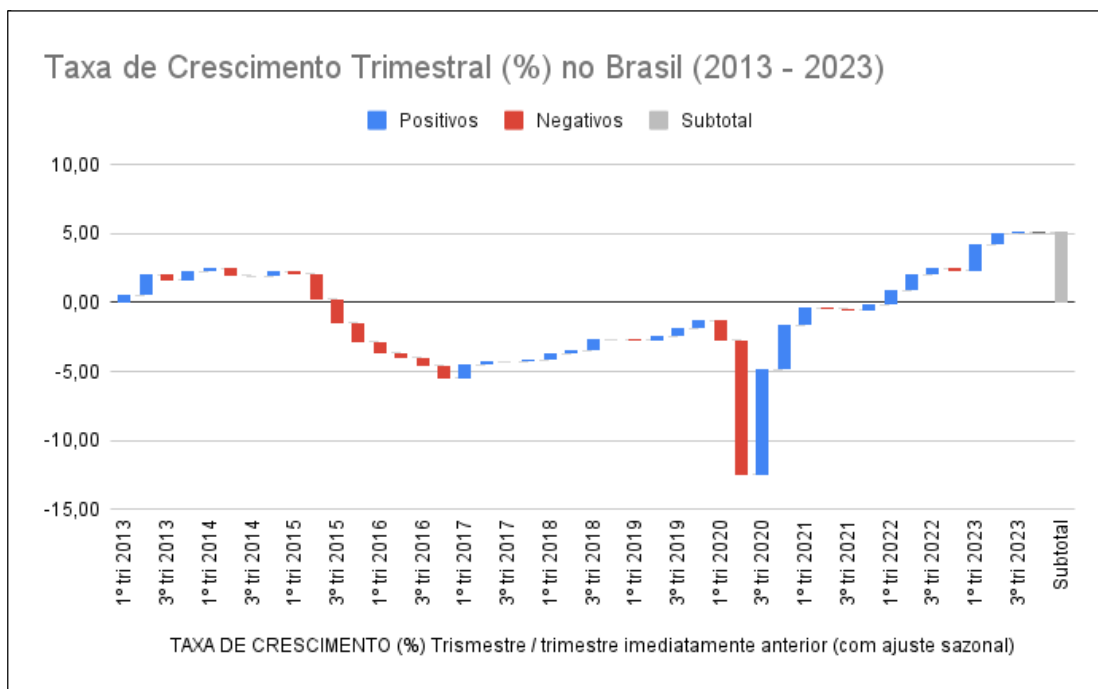


Fonte: IBGE (2024). Elaboração do autor.

No gráfico acima, observa-se uma retração econômica em 2013, seguida por um período prolongado de dificuldades de 2014 até o final de 2016, indicando uma recessão. Entre 2017 e meados de 2019, a economia mostra sinais de recuperação com crescimento positivo. No entanto, uma breve reversão ocorre no final de 2019 até início de 2020, refletindo potencialmente o impacto de eventos externos ou políticas internas. A partir de 2020, observa-se um aumento significativo no crescimento, embora com uma pequena retração em 2021 devido às consequências da pandemia de COVID-19. Após esse período, de 2021 a 2023, a taxa de crescimento aumenta consistentemente, alcançando os níveis mais altos do período examinado, sinalizando uma fase de robusto crescimento econômico.

A oscilação entre períodos de crescimento positivo e negativo observados na economia brasileira sugere a presença de ciclos econômicos que são influenciados por diversos fatores, como políticas governamentais, dinâmicas do mercado global, crises econômicas e eventos disruptivos como pandemias. A fase de recuperação e o crescimento sustentado que se seguem a crises ou recessões podem indicar uma estabilização e fortalecimento da economia. Entretanto, mudanças bruscas, particularmente evidentes nos anos de 2020 e 2021, ressaltam a vulnerabilidade da economia brasileira a choques, sejam eles originados internamente ou através de influências externas.

Gráfico 2 – Taxa de Crescimento (%) Trimestre / trimestre imediatamente anterior (com ajuste sazonal).



Fonte: IBGE (2024). Elaboração do autor.

Ao aplicar o ajuste sazonal e comparar os resultados trimestre a trimestre, o segundo gráfico proporciona uma visão mais imediata das variações econômicas, tornando-o particularmente sensível às mudanças de curto prazo. Esse gráfico revela uma volatilidade acentuada, destacando-se o profundo declínio no início de 2020, um reflexo direto do impacto da pandemia, seguido por uma recuperação rápida e marcante. Embora os efeitos da pandemia sejam claros em ambos os gráficos, a forma como as taxas de crescimento são calculadas influencia a representação desses impactos. No primeiro gráfico, o impacto da pandemia é apresentado de maneira mais suave e estendida ao longo do tempo, enquanto no gráfico com ajuste sazonal, as variações, tanto a queda quanto a recuperação, aparecem de forma mais intensa e imediata.

Portanto, o primeiro gráfico indica uma recuperação econômica que ocorre de forma gradual e consistente, caracterizada por flutuações menos drásticas. Em contraste, o segundo gráfico revela uma recuperação mais volátil, marcada por picos e quedas acentuados, o que demonstra a resposta imediata do mercado a eventos específicos, tais como a implementação de políticas de estímulo ou alterações nos períodos de *lockdown* durante a pandemia.

Este gráfico proporciona uma visão clara das variações na economia brasileira, sendo uma ferramenta valiosa para formuladores de políticas, investidores e analistas

avaliarem a condição econômica do país e realizarem projeções fundamentadas para o futuro. Devido a isso, a taxa de crescimento ajustada sazonalmente é particularmente pertinente para este estudo, pois o índice de sentimentos é derivado de cada ata, que reflete mudanças no curto prazo. Este enfoque permite modelar as variáveis de forma que auxiliem na tomada de decisões imediatas.

A Tabela 3 apresenta os resultados para os anos iniciais da análise, 2013 e 2014. Nesta tabela, é possível verificar os números das atas, os meses das reuniões do COPOM correspondentes a esse período, além do Índice de Sentimento para cada ata. Observa-se que durante quase todo o ano de 2013, os índices de sentimentos foram predominantemente negativos, com exceção de um único resultado positivo; já em 2014, predominam os sentimentos positivos. A coluna 5 mostra a média dos índices de duas reuniões por trimestre, ajustada para corresponder à mesma dimensionalidade das taxas de crescimento apresentadas nas colunas 6 e 7.

A disposição das variáveis na tabela proporciona a fundamentação para as análises que foram conduzidas após a modelagem de regressão. Os modelos foram aplicados considerando tanto a média do Índice de Sentimento de cada trimestre quanto os índices isolados da primeira e última reunião do trimestre. Além disso, cada uma dessas configurações foi testada com as duas diferentes taxas de crescimento mencionadas anteriormente. É importante destacar que a tabela completa, contendo os dados de todo o período analisado, está disponível no Apêndice A.

Tabela 3 – Resultados para os anos de 2013 e 2014.

Ata	Mês	Ano	ILM	ILM (média)	Taxa de Crescimento (%) Trimestre/ mesmo trimestre do ano anterior	Taxa de Crescimento (%) Trimestre/ trimestre imediatamente anterior (com ajuste sazonal)
172	Janeiro	2013	-0,18	-0,10	2,70	0,60
173	Março	2013	-0,02			
174	Abril	2013	-0,07	-0,11	4,00	1,50
175	Maio	2013	-0,16			
176	Julho	2013	-0,12	-0,11	2,80	-0,50
177	Agosto	2013	-0,11			
178	Outubro	2013	0,05	-0,05	2,50	0,70
179	Novembro	2013	-0,15			
180	Janeiro	2014	0,03	0,07	3,50	0,20
181	Fevereiro	2014	0,11			
182	Abril	2014	0,13	0,11	-0,40	-0,60
183	Maio	2014	0,1			
184	Julho	2014	0,11	0,13	-0,60	0,10
185	Setembro	2014	0,16			
186	Outubro	2014	0,16	0,19	-0,20	0,30
187	Dezembro	2014	0,22			

Fonte: IBGE (2024). Elaboração do autor.

Com esses dados em mãos, exploramos todas as possíveis interações entre as variáveis dependentes e independentes. Os resultados dessas análises estão detalhados na Tabela 4, que apresenta as estatísticas dos modelos utilizados. Importa destacar que na análise da autocorrelação dos resíduos o ILM é defasado em dois períodos. Os resultados indicaram que apenas a defasagem de dois períodos (lag2) mostrou-se significativa, com um p-valor abaixo do limiar de 0,05, sugerindo a presença de autocorrelação neste nível. Lags adicionais foram testados, mas não apresentaram significância estatística, indicando que não contribuem significativamente para o modelo e que não se estende além de dois períodos (ILM(-2)). Este resultado destaca a importância de considerar um ajuste no modelo que incorpore a autocorrelação até dois lags para garantir a adequação e precisão da análise.

Tabela 4 – Estatísticas dos Modelos.

Modelo	Medida	Taxa trimestre / trimestre imediatamente anterior			Taxa trimestre / mesmo trimestre do ano anterior		
		Index (primeiro)	Index (média)	Index (último)	Index (primeiro)	Index (média)	Index (último)
SIMPLES	R-quadrado ajustado	6,52%	14,17%	18,66%	13,48%	11,08%	6,09%
	P-valor (ILM(-2))	0,06	0,01	0,00	0,01	0,02	0,06
AR(2)	R-quadrado ajustado	11,57%	19,93%	22,88%	41,90%	38,50%	34,26%
	P-valor (ILM(-2))	0,02	0,02	0,00	0,01	0,03	0,15

Fonte: Elaboração do autor.

A tabela acima mostra estatísticas dos dois modelos de regressão utilizados neste estudo, a regressão linear simples e o AR(2), aplicados a diferentes métricas de crescimento econômico. Esses modelos foram usados para avaliar a relação entre índices de sentimento e taxas de crescimento econômico.

O R-quadrado ajustado no modelo de regressão linear simples varia de 6,52% para o primeiro índice trimestral a 18,66% para o último, sugerindo que este modelo explica apenas uma fração modesta da variabilidade nas taxas de crescimento econômico. E, o R-quadrado ajustado para o modelo AR(2) varia significativamente, alcançando de 11,57% no primeiro índice a 41,90% no primeiro índice comparando o trimestre com o mesmo período do ano anterior, o que demonstra uma maior eficácia deste modelo em captar as dinâmicas das taxas de crescimento.

O p-valor no contexto do modelo estatístico é comumente utilizado para testar a significância estatística dos coeficientes, sendo crucial para avaliar a robustez das relações estabelecidas no modelo. Valores de p-valor abaixo de 0,05 geralmente indicam que a relação entre as variáveis, como o índice ILM(-2) e o crescimento econômico, é estatisticamente significativa, rejeitando a hipótese nula de que o coeficiente é zero. No caso do modelo simples, observa-se que muitos p-valores estão acima de 0,05, implicando que não podemos rejeitar a hipótese nula de que alguns coeficientes são insignificantes, o que pode comprometer a validade das conclusões sobre a eficácia do índice ILM em prever o crescimento econômico.

Na comparação entre os modelos, o AR(2) demonstra ser mais robusto e explicativo que o modelo de regressão linear simples, evidenciado por valores mais altos de R-quadrado ajustado e p-valores geralmente mais baixos. A eficácia dos índices de sentimento como preditores das taxas de crescimento econômico varia; enquanto alguns índices exibem uma forte relação e significância estatística, outros são menos impactantes.

Essas variações nas estatísticas ressaltam como a escolha do tipo de taxa de crescimento e do modelo estatístico pode influenciar significativamente os resultados e interpretações em estudos econômicos futuros.

A partir desses resultados, conclui-se que o modelo AR(2) é o mais eficaz. O índice que apresentou a maior explicatividade, com um R-quadrado ajustado de 41,90%, foi o da taxa de crescimento comparando o trimestre com o mesmo trimestre do ano anterior. Este índice foi especialmente significativo quando aplicado ao primeiro índice de sentimento divulgado no trimestre, destacando-se com um p-valor extremamente baixo, indicando alta significância estatística do ILM defasado em dois períodos. Por isso, vale o destaque para a apresentação dos coeficientes do modelo conforme disposto na Tabela 5 e apresentado no Apêndice B os resultados de todas as regressões AR(2) para todos os parâmetros apresentados anteriormente na metodologia.

Tabela 5 – Resultado das análises de regressão.

Variável	Estimativa	Erro Padrão	Valor-t	Pr(> t)
Intercepto	-0,40117	0,48905	-0,82	0,41715
y(-1)	0,52752	0,14986	3,52	0,00114
y(-2)	0,03899	0,14844	0,263	0,79422
ILM(-2)	-5,38996	1,97884	-2,724	0,0097

Fonte: Análise própria baseada em dados coletados.

A análise de regressão linear múltipla apresentada revela influências significativas de certas variáveis anteriores sobre a variável dependente. O modelo inclui variáveis como y(-1), que mostrou um aumento significativo na variável dependente com cada unidade incrementada, e ILM(-2), que teve um efeito significativamente negativo. Embora o coeficiente para y(-2) não fosse significativo, o modelo como um todo foi estatisticamente significativo, com um R-quadrado ajustado de 41.9%, indicando que cerca de 42% da variação da variável dependente é explicada pelas variáveis independentes.

5 CONCLUSÃO

Este estudo se revela uma ferramenta valiosa para identificar quais modelos e variáveis são mais eficazes na previsão do crescimento econômico, baseando-se nos sentimentos expressos nas atas do COPOM. As análises realizadas podem orientar decisões políticas e estratégias de investimento, dependendo da precisão e confiabilidade dos modelos empregados. Este conhecimento é crucial para adaptar abordagens econômicas que respondam adequadamente às tendências observadas e aos indicadores de sentimentos.

A pesquisa constatou que o modelo AR(2), ao analisar a relação entre os índices de sentimentos e as taxas de crescimento econômico do Brasil, oferece *insights* profundos sobre como esses sentimentos impactam a economia. Com os índices de sentimentos, foi possível prever variações econômicas com notável precisão, especialmente em comparação com o modelo simples. O estudo reforçou que os índices de sentimentos são indicadores importantes e que técnicas de análise textual, aliadas a modelos estatísticos avançados, podem contribuir significativamente para a compreensão e previsão de tendências econômicas. Além disso, destacou-se a sensibilidade da economia brasileira a influências internas e externas, sublinhando a importância de incorporar essas variáveis em futuras análises para uma formulação de política econômica mais informada e proativa.

Para futuras pesquisas, recomenda-se expandir a análise para incluir períodos adicionais e ciclos econômicos variados, integrar variáveis globais como preços de commodities e taxas de câmbio, e aplicar métodos de aprendizado de máquina mais avançados para aprimorar as previsões dos modelos. O uso de modelos multivariados, que considerem várias variáveis dependentes, poderá oferecer uma visão mais abrangente dos impactos econômicos. Sugere-se também realizar estudos comparativos com outras economias em desenvolvimento e colaborar com formuladores de políticas para implementar praticamente os resultados, buscando estratégias mais eficazes para mitigar riscos e ajustar políticas econômicas. Além disso, recomenda-se a aplicação estruturada de testes de causalidade de Granger para investigar as relações dinâmicas entre o índice de sentimento ILM e o crescimento econômico de maneira rigorosa; e, ainda, em vez de considerar um dicionário com conteúdo fixo, recomenda-se construir um dicionário cujo conteúdo possa mudar ao longo do tempo, utilizando Machine Learning (ML) para identificar as palavras mais preditivas de uma determinada ata e usá-las para derivar novos preditores; esses novos preditores, por sua vez, melhorariam as previsões.

Para trabalhos futuros, recomenda-se, ainda, a expansão do modelo de previsão incorporando janelas dinâmicas e horizontes temporais múltiplos, visando adaptar as

previsões às mudanças comportamentais dos dados ao longo do tempo. Essa abordagem deve utilizar métricas de avaliação tradicionais como o Erro Quadrático Médio (EQM) para analisar a precisão das previsões em diferentes contextos e períodos. Além disso, seria proveitoso explorar o impacto de variáveis adicionais e técnicas de aprendizado de máquina mais avançadas para enriquecer o modelo e aumentar sua aplicabilidade e precisão. Implementar e testar essas extensões em um ambiente de dados diversificado ajudará a validar a robustez dos modelos propostos

REFERÊNCIAS

- ALGABA, A. et al. **Econometrics meets sentiment: An overview of methodology and applications**. Journal of Economic Surveys, Wiley Online Library, 2020.
- APEL, M.; GRIMALDI, M. **The information content of central bank minutes**. 2012.
- BANCO CENTRAL DO BRASIL – BACEN. **Atas COPOM**. Disponível em: <<https://www.bcb.gov.br/publicacoes/atascopom>>. Acesso em: 10 maio 2024.
- BARBOZA, R. d. M.; ZILBERMAN, E. **Os efeitos da incerteza sobre a atividade econômica no Brasil**. Revista Brasileira de Economia, SciELO Brasil, v. 72, n. 2, p. 144–160, 2018.
- BERNANKE, B. **Nonmonetary Effects of tge Financial Crisis in the Propagations of the Great Depression**. American Economic Review. 1983.
- BERNANKE, B. S.; LAUBACH, T; MISHKIN, F. S.; POSEN, A. S. **Inflation Targeting: Lessons from the International Experience**. Princeton University Press. 1999.
- BHOLAT, D. et al. **Text mining for central banks**. Available at SSRN 2624811, 2015.
- BISHOP, C. M. **Pattern Recognition and Machine Learning**. Springer. 2006.
- BREIMAN, L.; FRIEDMAN, J.; STONE, C. **Classification and Regression Trees**. Wadsworth International Group. 1984.
- BRUNO, G. **Text mining and sentiment extraction in central bank documents**. In: IEEE. 2016 IEEE International Conference on Big Data (Big Data). [S.l.], 2016. p. 1700–1708.
- BRUNO, G. **Central bank communications: Information extraction and semantic analysis**. In: The R User Conference, useR! 2017 July 4-7 2017 Brussels, Belgium. [S.l.: s.n.], 2017. p. 253.
- CAGAN, P. **The Monetary Dynamics of Hyperinflation**. In M. Friedman (Ed.), Studies ind the Quantity Theory of Money. (pp. 25-117). Chicago: University of Chicago Press. 1956.
- CHAUVET, M.; POTTER, S. **Forecasting output**. In Handbook of Economic Forecasting, volume 2, pages 141–194. Elsevier, 2013.
- CLEMENTS, A; READE, J. **The Predictive Power of Central Bank Sentiment Indices**. 2020.
- CORREA, R. et al. **Sentiment in central banks’ financial stability reports**. Available at SSRN 3091943, 2017.
- DIXIT, A.; PINDYCK, R. S. **Investment under Uncertainty**. Princeton Univerty Press. 1994.
- ENGELBERG, J. E.; PARSONS, C. A. **The causal impact of media in financial markets**.

The Journal of Finance, 66(1):67–97, 2011.

FERREIRA, P. C. *et al.* **Medindo a incerteza econômica no Brasil.** Economia Aplicada, 2017.

FILHO, A. E. C.; ROCHA, F. **Comunicação e política monetária no Brasil.** Revista Brasileira de Economia, SciELO Brasil, v. 63, n. 4, p. 405–422, 2009.

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.

FILHO, A. E. C. **Incerteza e atividade econômica no Brasil.** Economia Aplicada, SciELO Brasil, v. 18, n. 3, p. 421–453, 2014.

GARCIA, D. **Sentiment during Recessions.** The Journal of Finance, 68(3), 1267-1300. 2013.

GENTZKOW, M.; KELLY, B.; TADDY, M. **Text as data.** Journal of Economic Literature, 57(3), 535-574. 2019.

GODEIRO, L. L.; LIMA, L. R. R. de O. **Medindo incerteza macroeconômica para o Brasil.** Economia Aplicada, Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto-USP . . . , v. 21, n. 2, p. 311, 2017.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep Learning.** MIT Press. 2016.

GRIMMER, J.; STEWART, B. M. Text as data: **The promise and pitfalls off automatic content analysis methods for political texts.** Political Analysis, 21(3), 267-297. 2013.

GRIGNANI, F.; FONTANA, R. **Using machine learning and Bayesian networks to objectively analyze central bank statements and market sentiment.** Tese (Doutorado) — Politecnico di Torino, 2018.

HANSEN, S.; MCMAHON, M. **Shocking language: Understanding the macroeconomic effects of central bank communication.** Journal of International Economics, Elsevier, v. 99, p. S114–S133, 2016.

HASTIE, T.; TIBSHIRANI, R.; WAINWRIGHT, M. **High-Dimensional Data Analysis with Low-Dimensional Models: Principles, Computation e Applications.** Cambridge University Press. 2020.

HYBDNAB, R., ATHANASOPOULOS, G. **Forecasting: Principles and Practice.** Otexts. 2018.

HUTTO, C. J.; GILBERT, E. **VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text.** Proceedings of the International AAAI Conference on Web and Social Media (ICWSM). 2014.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA – IBGE. **Biblioteca.** Disponível em: <<https://biblioteca.ibge.gov.br/index.php/biblioteca-catalogo>>. Acesso em: 10 maio 2024.

- JONES, M.; SINCLAIR, T.; STEKLER, H. **Using Textual Analysis to Identify Economic Sentiments: A Study on Central Bank Communications.** 2019.
- KEYNES, J. M. **The General Theory of Employment, Interest and Money.** 1936.
- LIMA, L. R; GODEIRO, L. L; MOHSIN, M. **Time-Varying Dictionary and the Predictive Power of FED Minutes.** 2019.
- LIU, B. **Sentiment Analysis: Mining Opinions, Sentiments and Emotions.** Cambridge University Press. 2015.
- LOUGHRAN, T.; MCDONALD, B. **When is a Liability Not a Liability? Textual Analysis. Dictionaries e 10-Ks.** The Journal of Finance, 66(1), 35-65. 2011.
- LUCAS, R. E. Jr. **Econometric Policy Evaluation: A Critique.** Carnegie-Rochester Conference Series on Public Policy, 1(1), 19-46. 1976.
- LUCAS, R. E., Jr.; SARGENT, T., J. **Rational Expectations and Econometric Practice.** University of Minnesota Press. 1981.
- MACHADO, M. A. V.; SILVA, M. D. d. O. P. d. et al. **Análise do sentimento textual dos relatórios de desempenho trimestral das indústrias brasileiras.** Sociedade, Contabilidade e Gestão, v. 12, n. 1, 2017.
- MANKIW, N. G. **Sticky Prices in the United States.** Journal of Political Economy, 93(6), 1187-1211. 1985.
- MIKOLOV, T., CHEN, K., CORRADO, G., DEAN, J. **Efficient Estimation of Word Representations in Vector Space.** Proceedings of Workshop at ICLR. 2013.
- MISHKIN, F. S. **International Experiences with Different Monetary Policy Regimes.** Journal of Monetary Economics, 43, n. 3, 579-605. 1999.
- MISHKIN, F. S.; SCHMIDT-HEBBEL, K. **Does Inflation Targeting Make a Difference?** Working Paper 12876. 2007.
- MITCHELL, T. M. **Machine Learning.** McGraw-Hill. 1997.
- MONTES, G. C.; NICOLAY, R. T. d. F. **Central bank's perception on inflation and inflation expectations of experts: empirical evidence from brazil.** Journal of Economic Studies, Emerald Group Publishing Limited, v. 42, n. 6, p. 1142–1158, 2015.
- NELSON, C. R. **The prediction performance of the frb-mit-penn model of the us economy.** The American Economic Review, 62(5):902–917, 1972.
- PANG, B.; LEE, L. **Opinion Mining and Sentiment Analysis.** Foundations and Trend in Information Retrieval, 2(1-2), 1-135. 2008.
- RYBINSKI, K. I. **A machine learning framework for automated analysis of formal and informal central bank communication: the case of the national bank of poland.** 2018.

SARGENT, T. J. **Rational Expectations and Inflation**. Harper & Row Publishers. 1986.

SMITH, A. **An Inquiry into the Nature and Causes of the Wealth of Nations**. London: W. Strahan and R. Cadell. 1776.

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

SHILLER, R. J. **Narrative Economics: How Stories Go Viral and Drive Major Economic Events**, 2019, p. 25.

SILVA, A. **Impactos da Incerteza na Economia: Canais de Transmissão e Efeitos Macroeconômicos**. 2020.

SOLOW, R. M. A Contribution to the Theory of Economic Growth. *Quarterly Journal of Economics*. 1956.

SVENSSON, L. E. O. **Inflation Targeting**. In *Handbook of Monetary Economics*, editado por Benjamin M. Friedman e Michael Woodford, 3:1237-1302. Elsevier, 2010.

TURNEY, P. D. **Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews**. In *Proceedings of the 40th annual meeting on association for computational linguistics*, pages 417–424. Association for Computational Linguistics, 2002.

**APÊNDICE A – TABELA COMPLETA DOS DADOS UTILIZADOS NOS
MODELOS DE REGRESSÃO**

Ata	Mês	Ano	ILM	ILM (média)	Taxa de Crescimento (%) Trimestre/ mesmo trimestre do ano anterior	Taxa de Crescimento (%) Trimestre/ trimestre imediatamente anterior (com ajuste sazonal)
172	Janeiro	2013	-0,18	-0,10	2,70	0,60
173	Março	2013	-0,02			
174	Abril	2013	-0,07	-0,11	4,00	1,50
175	Maio	2013	-0,16			
176	Julho	2013	-0,12	-0,11	2,80	-0,50
177	Agosto	2013	-0,11			
178	Outubro	2013	0,05	-0,05	2,50	0,70
179	Novembro	2013	-0,15			
180	Janeiro	2014	0,03	0,07	3,50	0,20
181	Fevereiro	2014	0,11			
182	Abril	2014	0,13	0,11	-0,40	-0,60
183	Maio	2014	0,1			
184	Julho	2014	0,11	0,13	-0,60	0,10
185	Setembro	2014	0,16			
186	Outubro	2014	0,16	0,19	-0,20	0,30
187	Dezembro	2014	0,22			
188	Janeiro	2015	0,25	0,22	-1,60	-0,20
189	Março	2015	0,14			
190	Abril	2015	0,12	0,12	-2,70	-1,90
191	Junho	2015	0,05			
192	Julho	2015	0,07	0,06	-4,30	-1,70
193	Setembro	2015	-0,03			
194	Outubro	2015	-0,1	-0,07	-5,50	-1,40
195	Novembro	2015	-0,13			
196	Janeiro	2016	-0,1	-0,11	-5,10	-0,80
197	Março	2016	0,02			
198	Abril	2016	0,11	0,06	-3,20	-0,30
199	Junho	2016	-0,04			
200	Julho	2016	0,09	0,02	-2,50	-0,60
201	Agosto	2016	0,1			
202	Outubro	2016	0,1	-0,11	-2,30	-0,90
203	Novembro	2016	-0,33			
204	Janeiro	2017	0,04	0,11	0,30	1,00
205	Fevereiro	2017	0,18			
206	Abril	2017	0,09	0,02	0,80	0,20
207	Maio	2017	-0,04			
208	Junho	2017	-0,02	-0,07	1,60	0,10
209	Setembro	2017	-0,13			
Ata	Mês	Ano	ILM	ILM (média)	Taxa de Crescimento (%)	Taxa de Crescimento (%) Trimestre/

					Trimestre/ mesmo trimestre do ano anterior	trimestre imediatamente anterior (com ajuste sazonal)
210	Outubro	2017	-0,12	-0,17	2,60	0,10
211	Dezembro	2017	-0,22			
212	Fevereiro	2018	-0,38	-0,41	1,90	0,40
213	Março	2018	-0,45			
214	Maio	2018	-0,13	-0,25	1,60	0,20
215	Junho	2018	-0,38			
216	Agosto	2018	-0,36	-0,30	2,10	0,80
217	Setembro	2018	-0,25			
218	Outubro	2018	-0,02	-0,07	1,60	0,10
219	Dezembro	2018	-0,13			
220	Fevereiro	2019	-0,23	-0,30	0,90	-0,20
221	Março	2019	-0,38			
222	Maio	2019	-0,35	-0,32	1,20	0,40
223	Junho	2019	-0,3			
224	Julho	2019	-0,03	-0,03	1,10	0,60
225	Setembro	2019	-0,05			
226	Outubro	2019	-0,14	-0,05	1,70	0,50
227	Dezembro	2019	0,04			
228	Fevereiro	2020	-0,25	-0,352	0,40	-1,50
229	Março	2020	-0,45			
230	Maio	2020	-0,6	-0,55	-10,10	-9,70
231	Junho	2020	-0,5			
232	Agosto	2020	-0,29	-0,34	-3,00	7,70
233	Setembro	2020	-0,41			
234	Outubro	2020	-0,31	-0,06	-0,30	3,20
235	Dezembro	2020	0,19			
236	Janeiro	2021	0,08	-0,08	1,70	1,20
237	Março	2021	-0,24			
238	Maio	2021	0,25	0,26	12,40	-0,10
239	Junho	2021	0,27			
240	Agosto	2021	-0,1	-0,14	4,20	-0,10
241	Setembro	2021	-0,19			
242	Outubro	2021	-0,29	-0,33	1,50	0,50
243	Dezembro	2021	-0,38			
244	Fevereiro	2022	-0,43	-0,42	1,50	1,00
245	Março	2022	-0,43			
246	Maio	2022	-0,39			
247	Junho	2022	-0,37	-0,37	3,50	1,20
248	Agosto	2022	-0,51			
249	Setembro	2022	-0,4	-0,46	4,30	0,40
Ata	Mês	Ano	ILM	ILM (média)	Taxa de Crescimento (%) Trimestre/ mesmo trimestre do	Taxa de Crescimento (%) Trimestre/ trimestre imediatamente anterior (com

				ano anterior	ajuste sazonal)	
250	Outubro	2022	-0,5	-0,43	2,70	-0,20
251	Dezembro	2022	-0,36			
252	Fevereiro	2023	-0,42	-0,36	4,20	1,90
253	Março	2023	-0,29			
254	Maio	2023	-0,21	-0,23	3,50	0,90
255	Junho	2023	-0,27			
256	Agosto	2023	-0,2	-0,27	2,00	0,10
257	Setembro	2023	-0,35			
258	Outubro	2023	-0,28	-0,24	2,10	0,00
259	Dezembro	2023	-0,22			

APÊNDICE B – TABELA COMPLETA COM TODAS AS REGRESSÕES AR(2)

	(1)	(2)	(3)	(4)	(5)	(6)
y(-1)	-0,26720	-0,27310	0,54460*	0,52752*	-0,23660	0,52376*
	(0,95100)	(0,06887)	(0,00172)	(0,00114)	(0,09943)	(0,00175)
y(-2)	-0,19810	-0,19100	0,02150	0,03899	-0,18290	0,03783
	(0,18570)	(0,18006)	(0,89279)	(0,79422)	(0,19072)	(0,80628)
ILM(-2)	-3,79790*	-4,90540*	-2,95880	-5,38996*	-4,66000*	-4,73346*
	(0,01900)	(0,00239)	(0,15128)	(0,00970)	(0,00111)	(0,03274)

Apêndice B: Os testes de (1) a (3) abordam a taxa de crescimento comparando o trimestre atual com o trimestre imediatamente anterior. Os testes de (4) a (6) focam na taxa de crescimento comparando o trimestre atual com o mesmo trimestre do ano anterior. São apresentados os valores das estimativas e os respectivos p-valores para cada modelo envolvendo as variáveis $y(-1)$, $y(-2)$ e $ILM(-2)$. Estimativas marcadas com () indicam significância estatística ao nível de 5%.*