



UNIVERSIDADE FEDERAL DA PARAÍBA

Centro de Ciências Sociais Aplicadas

Departamento de Economia

Programa de Pós-Graduação em Economia

BRENO DA SILVA ARAÚJO PEREIRA

**Ensaio em macroeconomia: Previsão das finanças
públicas e Incerteza**

João Pessoa - PB

2023

BRENO DA SILVA ARAÚJO PEREIRA

**Ensaio em macroeconomia: Previsão das finanças
públicas e Incerteza**

Tese apresentada ao Programa de Pós-Graduação em Economia da Universidade Federal da Paraíba, em cumprimento às exigências do Curso de Doutorado em Economia Aplicada.

Orientador: Dr. José Luís da Silva Netto Júnior

Coorientador: Dr. Cássio da Nóbrega Besarria

João Pessoa - PB

2023

Catálogo na publicação
Seção de Catalogação e Classificação

P436e Pereira, Breno da Silva Araujo.

Ensaio em macroeconomia : previsão das finanças públicas e Incerteza / Breno da Silva Araujo Pereira. - João Pessoa, 2023.

75 f. : il.

Orientação: José Luis da Silva Netto Junior.

Coorientação: Cássio da Nóbrega Besarria.

Tese (Doutorado) - UFPB/CCSA.

1. Finanças públicas. 2. Machine learning. 3. Economia - Brasil - Índice de incerteza. 4. Atividade econômica - Redução - Brasil. 5. Inflação - Brasil. 6. Desemprego - Brasil. I. Silva Netto Junior, José Luis da. II. Besarria, Cássio da Nóbrega. III. Título.

UFPB/BC

CDU 336.1(043)



Universidade Federal da Paraíba
Centro de Ciências Sociais Aplicadas
Programa de Pós-Graduação em Economia

Campus Universitário I – Cidade Universitária – CEP 58.059-900 – João Pessoa – Paraíba
Tel: (83) 3216-7482 – <https://sigaa.ufpb.br/sigaa/public/programa/portal.jsf?id=1875> – E-mail: ppge.ccsa@gmail.com

Ata da reunião da Banca Examinadora designada para avaliar o trabalho de tese do doutorando **Breno da Silva Araújo Pereira**, submetida para obtenção do grau de doutor em Economia, área de concentração em **Economia Aplicada**.

Aos vinte e sete dias, do mês de fevereiro, do ano dois mil e vinte e três, às dez horas, no Programa de Pós-Graduação em Economia, do Centro de Ciências Sociais Aplicadas, da Universidade Federal da Paraíba, reuniram-se, em cerimônia pública, os membros da Banca Examinadora, constituída pelos professores doutores **José Luís da Silva Netto Júnior** (Orientador), da Universidade Federal da Paraíba; **Cássio da Nóbrega Besarria** (Coorientador), da Universidade Federal da Paraíba; **Edilean Kleber da Silva Bejarano Aragón** (Examinador Interno), da Universidade Federal da Paraíba; **Alejandro Cayetano Garcia Cintado** (Examinador Interno), da Universidade Federal da Paraíba; **Adriano Firmino Valdevino de Araújo** (Examinador Externo), da Universidade Federal da Paraíba; **Igor Ézio Maciel Silva** (Examinador Externo), da Universidade Federal do Rio Grande do Norte, a fim de examinarem o candidato ao grau de doutor em Economia, área de concentração em **Economia Aplicada**, **Breno da Silva Araújo Pereira**. Além dos examinadores e do examinando, compareceram também, representantes do Corpo Docente e do Corpo Discente. Iniciando a sessão, o professor **José Luís da Silva Netto Júnior**, na qualidade de presidente da Banca Examinadora, comunicou aos presentes a finalidade da reunião e os procedimentos de encaminhamento desta. A seguir, concedeu à palavra ao candidato, para que fizesse oralmente a exposição do trabalho, apresentado sob o título: **“Ensaio em macroeconomia: Previsão das finanças públicas e Incerteza”**. Concluída a exposição, o senhor presidente solicitou que fosse feita a arguição por cada um dos examinadores. A seguir foi concedida a palavra ao candidato, para que respondesse e esclarecesse às questões levantadas. Terminadas as arguições, a Banca Examinadora passou a proceder à avaliação e ao julgamento do candidato. Em seguida, o senhor presidente comunicou aos presentes que a Banca Examinadora, por unanimidade, **aprovou** a tese apresentada e defendida com o conceito **APROVADO**, concedendo assim, o grau de **Doutor em Economia**, área de concentração em **Economia Aplicada**, ao doutorando **Breno da Silva Araújo Pereira**. E, para constar, eu, **Ricardo Madeira Cataldi**, secretário *ad hoc* do Programa de Pós-Graduação em Economia, lavrei a presente ata, que assino junto com os membros da Banca Examinadora. João Pessoa, 27 de fevereiro de 2023.

Prof. Dr. José Luis da Silva Netto Junior
Orientador – UFPB

Documento assinado digitalmente
JOSE LUIS DA SILVA NETTO JUNIOR
Data: 28/02/2023 16:54:39-0300
Verifique em <https://verificador.iti.br>

Prof. Dr. Cássio da Nóbrega Besarria
Coorientador – UFPB

Documento assinado digitalmente
CASSIO DA NOBREGA BESARRIA
Data: 26/03/2023 11:17:11-0300
Verifique em <https://validar.iti.gov.br>

Prof. Dr. Edilean Kleber da Silva Bejarano Aragón
Examinador Interno – UFPB

EDILEAN KLEBER DA SILVA BEJARANO
ARAGON:00784361410
410
Assinado de forma digital por EDILEAN KLEBER DA SILVA BEJARANO
ARAGON:00784361410
Dados: 2023.03.27 10:20:12 -03'00'

Prof. Dr. Alejandro Cayetano Garcia Cintado
Examinador Interno – UFPB

GARCIA CINTADO ALEJANDRO CAYETANO - 28733550H
Firmado digitalmente por GARCIA CINTADO ALEJANDRO CAYETANO - 28733550H
Fecha: 2023.03.27 15:42:29 -02'00'

Prof. Dr. Adriano Firmino Valdevino de Araújo
Examinador Externo – UFPB

Documento assinado digitalmente
ADRIANO FIRMINO VALDEVINO DE ARAUJO
Data: 28/03/2023 08:38:06-0300
Verifique em <https://validar.iti.gov.br>

Prof. Dr. Igor Ézio Maciel Silva
Examinador Externo – UFRN

Documento assinado digitalmente
IGOR EZIO MACIEL SILVA
Data: 28/03/2023 15:12:04-0300
Verifique em <https://validar.iti.gov.br>

Ricardo Madeira Cataldi
Secretário Ad Hoc – PPGE/ UFPB

Documento assinado digitalmente
RICARDO MADEIRA CATALDI
Data: 11/05/2023 09:18:26-0300
Verifique em <https://validar.iti.gov.br>

AGRADECIMENTOS

Primeiramente, agradeço à meus pais, Adelmo e Jeane, e meu irmão, Bruno, pelo apoio e por sempre acreditarem em meu potencial. Também, à Carol, pelo companheirismo de toda essa jornada.

Aos meus orientadores, José Luis da Silva Netto Junior e Cássio da Nóbrega Besarria pela paciência, confiança e suporte para a conclusão deste trabalho. Aos demais docentes do PPGE/UFPB que tive a oportunidade de conhecer durante o mestrado e o doutorado. Cada um contribuiu para o meu crescimento e amadurecimento dentro da academia.

Aos participantes da banca examinadora, Prof. Edilean, Prof. Alejandro, Prof. Adriano e Prof. Igor pelas contribuições sugeridas, que agregaram e aprimoraram as pesquisas que compõem essa tese.

Agradeço aos amigos que fiz durante os anos em que morei na Paraíba e que extrapolam qualquer relação acadêmica. Em especial, Daniel, Robson, Gerrio, Helson, Rômulo, Camilla, Eduarda, Ana, Rennan e Diego.

Aos meus amigos de infância, que mesmo distante se fizeram presentes. Em especial, Matheus, Xande, Allan, Artur, Beltrão e Victor, presentes há mais de 20 anos na minha vida.

À FAPESQ-PB pelo apoio financeiro, via concessão de bolsa de estudos, durante o doutorado.

RESUMO

Capítulo 1 - Previsão das finanças públicas brasileiras usando modelos univariados de Machine Learning

Este artigo analisa o desempenho de modelos univariados para prever cinco variáveis macroeconômicas brasileiras, relacionadas com as finanças públicas: Arrecadação Federal, Receita Líquida, Despesa Total, Resultado Primário e Resultado Nominal. Além do modelo Arima, considerado como método *benchmark* dessa pesquisa, mais 6 modelos são aplicados - Prophet, XGBoost, Arima-XGboost, Prophet-XGBoost e duas combinações das previsões - com o intuito de se identificar métodos que possam aprimorar as previsões das variáveis citadas. Os resultados indicam que para as três primeiras variáveis, o método Prophet-XGBoost e as duas combinações de previsões são modelos concorrentes ao método Arima, apresentando resultados mais precisos em diversos casos. Para as demais variáveis, o método Arima é mais preciso, principalmente nos maiores horizontes temporais.

Palavras-chave: Previsão; Machine Learning; Finanças Públicas.

Capítulo 2 - Incerteza e atividade econômica: um índice de incerteza baseado nas pesquisas do Google

Este trabalho tem como objetivo desenvolver um novo índice de incerteza para o Brasil. A medida utilizada é baseada em uma estratégia que utiliza a base de dados do Google *Trends* de pesquisas na Internet, base em tempo real e de acesso livre. O índice GTU-BR abrange os períodos de janeiro de 2006 à dezembro de 2019 e é positivamente correlacionado com outros índices de incerteza disponíveis para o Brasil, além de identificar bem diversos momentos de aumento da incerteza no país. As investigações de um modelo de vetores autorregressivos apontam o efeito contracionista de um choque de incerteza na economia brasileira, aumentando o desemprego, a inflação e reduzindo a atividade econômica.

Palavras-chave: Incerteza; Atividade Econômica; Inflação; Desemprego.

ABSTRACT

Chapter 1 - Forecasting Brazilian public finances using univariate Machine Learning models

This article analyzes the performance of univariate models to predict five Brazilian macroeconomic variables related to public finances: Federal Revenue, Net Revenue, Total Expenditure, Primary Result and Nominal Result. In addition to the Arima model, considered as the *benchmark* method of this research, 6 more models are applied - Prophet, XGBoost, Arima-XGboost, Prophet-XGBoost and two combinations of the forecasts - in order to identify methods that can improve the predictions of the aforementioned variables. The results indicate that for the first three variables, the Prophet-XGBoost method and the two combinations of forecasts are competing models to the Arima method, presenting more accurate results in several cases. For the other variables, the Arima method is more accurate, especially in longer time horizons.

Keywords: Forecasting; Machine Learning; Public Finances.

Chapter 2 -Uncertainty and Economic Activity: An Uncertainty Index Based on Google Searches

This work aims to develop a new uncertainty index for Brazil. The measurement used is based on a strategy that uses the Google Trends database of Internet searches, a real-time and freely accessible database. The GTU-BR index is positively correlated with other uncertainty indices available for Brazil, in addition to identifying several moments of increased uncertainty in the country. Investigations of an autoregressive vector model point to the contractionary effect of an uncertainty shock in the Brazilian economy, increasing unemployment, inflation and reducing economic activity.

Keywords: Uncertainty; Economic activity; Inflation; Unemployment.

LISTA DE ILUSTRAÇÕES

Figura 1.1 – Receita Líquida e Despesa Total (% do PIB)	17
Figura 1.2 – Resultado Primário do Governo Central - acima da linha (% do PIB) .	18
Figura 1.3 – Resultado Nominal do Governo Central (% do PIB)	18
Figura 1.4 – Validação cruzada k-passos-a-frente	26
Figura 2.1 – Índice de Incerteza - GTU-BR	46
Figura 2.2 – Função impulso-resposta do modelo principal	53
Figura A.1 – Arrecadação das Receitas Federais - em R\$ milhões	68
Figura A.2 – Receita Líquida do Governo Central - em R\$ milhões	68
Figura A.3 – Despesa Total do Governo Central - em R\$ milhões	69
Figura A.4 – Resultado Primário do Governo Central - em R\$ milhões	69
Figura A.5 – Resultado Nominal do Governo Central - em R\$ milhões	69
Figura C.1 – Trajetória temporal das variáveis	71
Figura D.1 – Função impulso-resposta com índice de incerteza IIE-BR	73
Figura D.2 – Função impulso-resposta com índice de incerteza EPU	73
Figura D.3 – Função impulso-resposta com índice de incerteza como variável dummy	74
Figura D.4 – Função impulso-resposta para PIM-PF	74
Figura D.5 – Função impulso-resposta para Varejo	74
Figura D.6 – Função impulso-resposta para diferentes durações de restrição	75

LISTA DE TABELAS

Tabela 1.1 – Receita do Governo Central (% do PIB)	15
Tabela 1.2 – Despesa Total do Governo Central (% do PIB)	16
Tabela 1.3 – Comparação das previsões	30
Tabela 2.1 – Correlação entre os índices de incerteza	47
Tabela 2.2 – Descrição dos dados utilizados no modelo principal	50
Tabela 2.3 – Restrições de sinais impostas nas respostas estruturais	52
Tabela B.1 – Palavras-chave do índice GTU-BR	70
Tabela C.1 – Correlação entre as variáveis utilizadas no trabalho	72
Tabela C.2 – Estatística descritiva das variáveis	72
Tabela C.3 – Testes de raiz unitária*	72

SUMÁRIO

1	Previsão das finanças públicas brasileiras usando modelos univariados de Machine Learning	10
1.1	Introdução	10
1.2	Revisão da Literatura	12
1.3	Indicadores Fiscais do Governo Central	15
1.4	Metodologia	20
1.4.1	Modelos ARIMA	21
1.4.2	Prophet	22
1.4.3	Extreme Gradient Boosting - XGBoost	23
1.4.4	Modelos Híbridos	24
1.4.5	Avaliação dos Modelos de Previsão	25
1.5	Dados	27
1.6	Resultados	29
1.7	Considerações finais	32
2	Incerteza e atividade econômica: um índice de incerteza baseado nas pesquisas do Google	34
2.1	Introdução	34
2.2	Revisão da Literatura	38
2.2.1	Medidas de Incerteza	38
2.2.2	Choques de Incerteza	39
2.2.2.1	Atividade econômica	40
2.2.2.2	Inflação	41
2.2.2.3	Juros	42
2.2.2.4	Mercado de ações	42
2.3	Índice Google de Incerteza Econômica - GTU-BR	44
2.3.1	Correlação com índices de incerteza existentes	47
2.4	Modelo VAR com Restrições de Sinais	48
2.4.1	Descrição dos dados	49
2.5	Resultados	52
2.5.1	Resultados do modelo principal	53
2.5.2	Teste de robustez	54
2.5.2.1	Diferentes medidas para a incerteza	54
2.5.2.2	Diferentes medidas para a atividade econômica	55
2.5.2.3	Choque de incerteza com diferentes durações	56
2.6	Considerações finais	57

Referências	59
Apêndices	67
APÊNDICE A – Série temporal das variáveis previstas	68
APÊNDICE B – Palavras-chave utilizadas na criação do índice GTU-BR	70
APÊNDICE C – Trajetória e testes econométricos aplicados as séries temporais utilizadas no modelo VAR	71
APÊNDICE D – Gráficos do teste de robustez	73

1 PREVISÃO DAS FINANÇAS PÚBLICAS BRASILEIRAS USANDO MODELOS UNIVARIADOS DE MACHINE LEARNING

1.1 INTRODUÇÃO

As séries temporais são estudadas para diversos fins, como a previsão do futuro com base no conhecimento do passado, a compreensão do fenômeno subjacente às medidas ou simplesmente uma descrição precisa das características de uma determinada série. De forma mais exclusiva, a previsão de valores futuros de uma série de tempo observada desempenha um papel importante em quase todos os campos da ciência e engenharia, como economia, finanças, inteligência de negócios, meteorologia e telecomunicações.

Previsões de variáveis econômicas contribuem para que agentes econômicos possam tomar decisões com menor grau de incerteza. Do lado das famílias, essas previsões influenciam em quanto economizar para épocas de maior incerteza. Da mesma forma, a escolha das empresas sobre investimento depende de suas previsões de fluxos de caixa futuros, de forma que, mesmo um pequeno aumento na precisão das previsões pode gerar ganhos notáveis em atividades como planejamento de gestão e definição de estratégias (MAKRIDAKIS, 1996); (SYNTETOS; BOYLAN; DISNEY, 2009).

Do lado do governo, previsões com maior acurácia da renda, por exemplo, permitem a construção de políticas públicas mais adequadas para o presente, mas também para o futuro. Além disso, as decisões de um banco central, cujo principal papel é definir as taxas de juros - como instrumento para controle da inflação - tendo como incerto os valores futuros para o crescimento do produto, desemprego e da própria inflação, dependendo de suas previsões para a tomada de decisão mais correta (ELLIOTT; TIMMERMANN, 2008).

Assim, há cada vez mais esforços para se identificar tendências, padrões e prever trajetórias dos indicadores econômicos como taxa de juros, arrecadação e despesas. Além disso, diversos autores, como Hollauer e Issler (2006), Bohn e Bueno (2015) e Berge (2015), se dedicam a identificar indicadores antecedentes, que são um conjunto de variáveis que tendem a se mover com certa antecedência em relação ao ciclo econômico, sendo uma possível ferramenta para a previsão mais apurada de variáveis fiscais e econômicas.

Os benefícios de uma previsão mais acurada da arrecadação federal brasileira, por exemplo, vão de auxiliar a Receita Federal do Brasil (RFB) no acompanhamento e controle das receitas sob sua administração, bem como, de antecipar o montante de receitas disponíveis a cargo do administrador tributário, uma vez que estas funcionam como

limitador na fixação do montante de gastos públicos, o que contribui para o necessário equilíbrio nas contas (CAMPOS, 2009).

Assim, a previsão de variáveis relacionadas ao orçamento público são requisitos essenciais na gestão fiscal, garantindo que eventuais desvios em relação à previsão – ou ao conjunto de previsões – sejam corrigidos ao longo do exercício fiscal, compensando os riscos orçamentários por meio de realocação e/ou redução de despesas, com o objetivo de que a meta fiscal seja cumprida.

Estas previsões podem ser efetuadas de formas diferentes, desde modelos mais simples até modelos mais complexos. O modelo SARIMA se destaca por sua fácil compreensão e implementação, sendo considerado o *benchmark* do trabalho. Além deste método, serão utilizados métodos não lineares como o Prophet e métodos de *Machine Learning* baseado em árvores de decisão que utiliza uma estrutura de *Gradient Boosting*¹.

Assim, o objetivo principal dessa pesquisa é de observar como novos métodos univariados de previsões se comportam na previsão de variáveis fiscais brasileiras. As variáveis fiscais estudadas serão: 1) Arrecadação Federal; 2) Receita Líquida do Governo Geral; 3) Despesa Total do Governo Central; 4) Resultado Primário do Governo Central acima da linha; e 5) Resultado Nominal do Governo Central, disponibilizados pela Receita Federal do Brasil (RFB) e pela Secretaria do Tesouro Nacional (STN).

Com isso, esse trabalho contribui para a literatura de métodos de previsões de variáveis relacionadas às finanças públicas brasileiras, ao analisar o desempenho de modelos univariados alternativos mais recentes, que combinam métodos lineares e não lineares. A pesquisa irá observar três horizontes de previsões, identificando, conseqüentemente, métodos que possuem melhor (ou pior) desempenho ao longo de três, seis ou doze meses.

O restante desse trabalho está organizado em mais cinco seções além dessa introdução, divididos da seguinte forma: a Seção 2 apresenta o referencial teórico da pesquisa, a Seção 3 apresenta o modelo benchmark e os métodos alternativos utilizados, apontando algumas de suas diferenças na estimação e seleção de modelos. A Seção 4 descreve as variáveis utilizadas e as modificações necessárias para se proceder as previsões, depois, na Seção 5 são discutidos os resultados obtidos nesse estudo. Por fim, a última seção apresenta as conclusões obtidas nessa pesquisa.

¹ Técnica que utiliza uma combinação de resultados de preditores fracos (modelos que, quando utilizados individualmente possuem uma acurácia abaixo do esperado), com o objetivo de produzir um melhor modelo preditivo

1.2 REVISÃO DA LITERATURA

Uma previsão econômica deve fornecer informações sobre: 1) a direção do movimento da economia; 2) o momento de qualquer ponto de virada; 3) a magnitude da mudança; e 4) o período de tempo durante o qual se espera que um movimento persista. Essas informações fornecem aos usuários das previsões uma compreensão dos pontos fortes e das limitações das previsões (FILDES; STEKLER, 2002).

Portanto, o que constitui uma boa previsão depende do custo de vários erros de previsão para a equipe ou profissional que possui esta função e, portanto, reflete tanto as preferências deste quanto a maneira pela qual as previsões são transformadas em decisões econômicas, auxiliando na melhoria das tomadas de decisão. Além disso, a identificação de um bom modelo de previsão depende das dificuldades de sua implementação.

Há muito tempo, a literatura de previsão de séries temporais tem sido dominada por métodos estatísticos lineares, como os modelos de suavização exponencial (ETS) e os modelos da família SARIMA. No entanto, Gooijer e Hyndman (2006) observam que a partir dos anos 1980 trabalhos passaram a identificar que métodos lineares não são adaptados para muitas aplicações com dados reais, como as variáveis econômicas. Neste mesmo período, diversos modelos de séries temporais não lineares foram propostas, como o modelo bilinear, por Poskitt e Tremayne (1986), os modelos autorregressivos com limiar (*threshold*) desenvolvidos em Tong (1983) e Tong (1990), e o modelo de heterocedasticidade condicional autorregressivo (ARCH) proposto por Engle (1982).

A partir da literatura mais recente de previsões, há uma mudança considerável na forma como os economistas calculam, aplicam e avaliam suas previsões. Isto se dá pelo desenvolvimento de novas técnicas que facilitam a busca e estimativa dos parâmetros de métodos de previsão cada vez mais complicados, além do acesso a novas variáveis preditoras potenciais, para os casos de modelos multivariados.

Nas últimas décadas, Palit e Popovic (2006) e Ahmed et al. (2010) destacam que os modelos de *machine learning* chamaram a atenção e se estabeleceram como sérios concorrentes aos modelos estatísticos clássicos na área de previsões. Estes métodos são exemplos de modelos não lineares não paramétricos que usam apenas dados históricos para aprender a dependência estocástica entre o passado e o futuro, apresentando, por vezes, previsões mais precisas do que os modelos tradicionais.

Estes resultados mais precisos entre as variáveis econômicas estão associados, entre outros fatores, a alta incerteza macroeconômica, estresse financeiro e estouro de bolhas imobiliárias, o que sugere que os métodos de *machine learning* são úteis para previsões macroeconômicas por capturarem não linearidades importantes que surgem no contexto de incerteza e atritos financeiros.

Portanto, a produção de previsões mais precisas se torna uma ferramenta provi-

dencial na tomada de decisão dos agentes econômicos, principalmente em períodos de maior incerteza, como o vivenciado no passado na crise do *subprime* e mais recentemente durante a pandemia do coronavírus iniciada, no Brasil, em 2020. Período mais desafiador para os profissionais e seus modelos até então utilizados, fazendo-se necessário o estudo e aplicação de modelos alternativos que podem auxiliar na melhora do desempenho das previsões das variáveis de interesse.

Além de momentos de maior incerteza, períodos eleitorais também influenciam a dinâmica das receitas e despesas e, conseqüentemente, dos resultados obtidos nas previsões dessas variáveis. Siqueira (2015) encontra que o Brasil possui política fiscal que vai de acordo com o modelo teórico de ciclos políticos orçamentários², em que, em anos eleitorais, os políticos tendem a reduzir os impostos ou elevar os gastos.

Do mesmo modo, Ebeke e Ölcer (2013) encontram resultados que apontam que, em países subdesenvolvidos, além do aumento de despesas em períodos eleitorais, também são observados fortes ajustes fiscais logo após este período, através da geração de novas receitas.

Quanto aos métodos de previsões, a literatura internacional abrange trabalhos como o de Marcellino, Stock e Watson (2003), em que os autores identificam a importância de modelos univariados autorregressivos lineares para a previsão de variáveis macroeconômicas de países da União Europeia. Neste mesmo sentido, Terui e Dijk (2002) e Teräsvirta, Dijk e Medeiros (2005) identificam que além de modelos univariados autorregressivos lineares, métodos que combinam as previsões de modelos lineares e não lineares possuem boa acurácia nas predições de variáveis macroeconômicas.

Na literatura nacional, Mendonça e Medrano (2016) utilizam um modelo de combinação de previsões para prever a arrecadação tributária no Brasil. De acordo com os resultados obtidos, os autores identificam que a combinação de previsões alcançou, na maior parte dos casos, resultado superior quando comparado com as previsões obtidas pelos métodos utilizados individualmente.

Garcia, Medeiros e Vasconcelos (2017) encontram que diversos modelos de *Machine Learning* apresentam bons resultados de previsão quando comparados com modelos já tradicionais ao se prever a inflação brasileira. Além disso, os autores preveem que a médias das previsões desses modelos de *Machine Learning* apresenta as melhores previsões entre as todas os métodos concorrentes.

Outra gama de modelos é avaliada por Cepni, Güney e Swanson (2019), que utilizam modelos de fatores dinâmicos para prever o PIB de diversas economias emergentes, entre elas o Brasil, utilizando índices de difusão local e global. É identificado que estes métodos são úteis para previsões a curto prazo e previsão do crescimento do PIB desses mercados

² Para mais detalhamento, ver Rogoff e Sibert (1988).

emergentes.

Mais recentemente, Barbosa, Ferreira e Silva (2020) utilizam modelos de séries temporais de alta dimensão com fatores estimados por componentes principais e a combinação de técnicas para avaliar como esses métodos melhoram o desempenho preditivo das variáveis econômicas brasileiras. Os autores encontram que os modelos com supervisão e aprendizado estatístico apresentam bons resultados na previsão de índices de preços e taxa de desemprego, concluindo que estes métodos são opções robustas para a previsão de variáveis macroeconômicas brasileiras.

1.3 INDICADORES FISCAIS DO GOVERNO CENTRAL

Aprovada no ano de 2000, a Lei de Responsabilidade Fiscal (LRF), tem como um dos objetivos principais controlar gastos e o endividamento excessivo da União, estados e municípios, fundamental para o bom desempenho da gestão pública no uso eficiente dos recursos orçamentários escassos. Além disso, a LRF obriga que os administradores públicos incluam nos orçamentos metas de receitas, despesas, resultado nominal e primário, essenciais para o planejamento financeiro eficaz, aderindo, assim, à causa da *accountability* pública.

Segundo Carlitz (2013) a *accountability* pública é o conjunto de mecanismos que permitem que os gestores prestem contas e sejam responsabilizados pelos seus atos. Esta prática que tem como objetivo reduzir os riscos da concentração de poder e garantir que a população participe da tomada de decisão dos governos, através de maior transparência com a sociedade em relação aos seus atos.

A previsão da arrecadação das receitas públicas federais faz parte do início da elaboração do orçamento anual do governo, como forma de balizar o montante total dos gastos públicos, contribuindo para o necessário equilíbrio nas contas públicas. Assim, uma previsão acurada desta variável proporciona aos gestores públicos, economistas, entre outros, acompanhar as alterações no cenário macroeconômico, analisando sua influência sobre os valores arrecadados e identificando as possíveis causas dos desvios dos valores arrecadados em relação ao estimado, bem como auxiliar na atividade de fiscalização (CAMPOS, 2009; VIGNOLI; FUNCIA, 2015).

Os impostos e contribuições de competência do governo central são apresentados na Tabela 1.1. O Imposto sobre a Renda (IR) e a Contribuição para o Financiamento da Seguridade Social (COFINS) são os principais impostos da receita tributária federal em 2021, atingindo 5,8% e 2,9% do PIB, respectivamente.

Tabela 1.1 – Receita do Governo Central (% do PIB)

Discriminação	2013	2014	2015	2016	2017	2018	2019	2020	2021
1. Receita Administrada pela RFB	13,5	12,8	12,8	13,1	12,7	12,9	12,8	12,0	13,8
1.1 - Imposto de importação	0,7	0,6	0,6	0,5	0,5	0,6	0,6	0,6	0,7
1.2 - IPI	0,8	0,9	0,8	0,7	0,7	0,8	0,7	0,8	0,8
1.3 - Imposto de Renda (IR)	5,1	4,9	5,0	5,4	5,2	5,2	5,4	5,1	6,0
1.4 - IOF	0,6	0,5	0,6	0,5	0,5	0,5	0,6	0,3	0,6
1.5 - COFINS	3,8	3,4	3,4	3,3	3,6	3,6	3,2	3,0	3,2
1.6 - PIS/PASEP	1,0	0,9	0,9	0,9	0,9	0,9	0,9	0,8	0,9
1.7 - CSLL	1,2	1,1	1,0	1,1	1,1	1,1	1,1	1,1	1,3
1.8 - CIDE	0,0	0,0	0,1	0,1	0,1	0,1	0,0	0,0	0,0
1.9 - Outras receitas administradas pela RFB	0,3	0,4	0,4	0,6	0,0	0,1	0,3	0,3	0,3
2 - Arrecadação Líquida para o RGPS	5,8	5,8	5,8	5,7	5,7	5,6	5,6	5,4	5,3
3 - Receitas não administradas pela RFB	2,8	2,5	2,2	2,2	2,7	2,8	3,7	2,2	3,2
Receita Total	22,1	21,1	20,8	21,0	21,0	21,3	22,1	19,7	22,3
Receita Líquida	18,7	17,7	17,4	17,4	17,5	17,6	18,2	16,1	18,2

Fonte: Secretaria do Tesouro Nacional. Elaboração própria.

Giambiagi, Alem e Pinto (2015) apontam que após a reforma ocorrida em 1988 houve um aumento da participação das receitas das contribuições - não sujeitas a partilhas com estados e municípios - como percentual do PIB. Em grande medida, este fato reflete a tentativa por parte da União de aumentar a sua receita disponível ante o aumento da descentralização das receitas tributárias no período pós-reforma 1988.

Além destas fontes de receita, o governo ainda tem a arrecadação líquida proveniente do Regime Geral de Previdência Social (RGPS) e as receitas não administradas pela RFB, como concessões, dividendos receitas de exploração de recursos naturais, entre outras fontes. Ainda são excluídos os montantes relacionados à incentivos fiscais concedidos pelo União. A soma de todas essas fontes destacadas resulta na Receita Total do Governo Central.

Para a União, a receita disponível - balanço entre o que cada esfera repassa e recebe das outras - é menor do que a bruta, pois ela transfere recursos para os demais níveis de governo. Já para os municípios, a situação é inversa, uma vez que os municípios recebem transferências tanto da União como dos estados. Nestes últimos o efeito líquido depende do balanço entre recebimentos da União e repasses para os municípios.

Além destes, impostos como o que incide sobre propriedade de veículos automotores (IPVA), imposto sobre transmissão causa mortis e doação e o imposto sobre operações relativas à circulação de mercadorias e sobre prestações de serviços de transporte interestadual, intermunicipal e de comunicação (ICMS) - este último representa, individualmente, a maior fonte de receita do país - são de competência dos estados e do Distrito Federal. Já o IPTU, o imposto sobre transmissão de bens intervivos (ITBI), o imposto sobre vendas e varejo de combustíveis líquidos e gasosos (IVV) e o imposto sobre serviços (ISS) são de competência dos municípios, não fazendo parte, portanto, do cálculo da Receita Total e Líquida do Governo Central.

As despesas totais do governo federal são caracterizadas pelos gastos necessários para promover os serviços públicos à sociedade, desconsiderando o pagamento de empréstimos e financiamentos – despesas com pessoal, encargos sociais, transferências para outros entes públicos e créditos extraordinários são exemplos de despesas que fazem parte desta variável. Nos últimos anos o montante utilizado nesta área é distribuído da seguinte maneira, como identificado na Tabela 1.2.

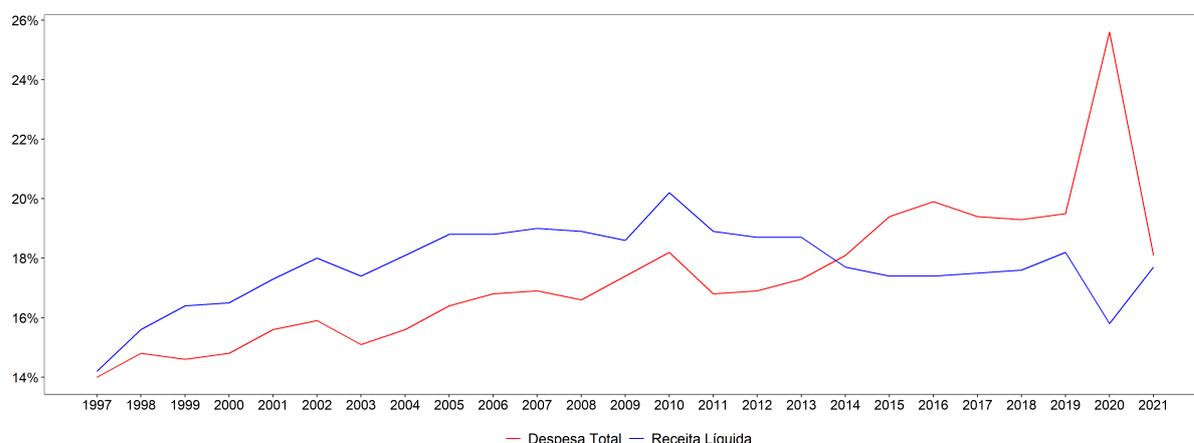
Tabela 1.2 – Despesa Total do Governo Central (% do PIB)

Discriminação	2013	2014	2015	2016	2017	2018	2019	2020	2021
Benefícios Previdenciários	6,7	6,8	7,3	8,1	8,5	8,4	8,5	8,7	8,0
Pessoal e Encargos Sociais	3,8	3,8	4,0	4,1	4,3	4,3	4,2	4,2	3,7
Outras Despesas Obrigatórias	2,7	3,0	4,1	3,4	3,0	2,9	2,6	9,5	3,4
Despesas do Poder Exec. Sujeitas à Programação Financeira	4,1	4,5	4,0	4,3	3,7	3,8	4,2	3,2	3,0
Despesa Total	17,3	18,1	19,4	19,9	19,4	19,3	19,5	25,6	18,1

Fonte: Secretaria do Tesouro Nacional. Elaboração própria.

Desde 1997 é possível identificar um crescimento dos gastos públicos em relação ao PIB, saindo de 14,8% em 1997, para patamares próximos aos 19% nos últimos anos – desconsiderando o ano de 2020, em que a despesa total esteve acima de 25% devido à créditos extraordinários referentes à pandemia – apresentando um crescimento acima do observado pelo PIB, pressionando, portanto, a carga tributária. O gasto com pessoal e encargos sociais se manteve praticamente constante neste período, oscilando entre 3,8% e 4,6%, já os benefícios previdenciários saltaram de 4,9%, em 1997, para patamares acima de 8% desde 2016, representando mais de 44% dos gastos totais do Governo Central.

Figura 1.1 – Receita Líquida e Despesa Total (% do PIB)



Fonte: Secretaria do Tesouro Nacional. Elaboração própria.

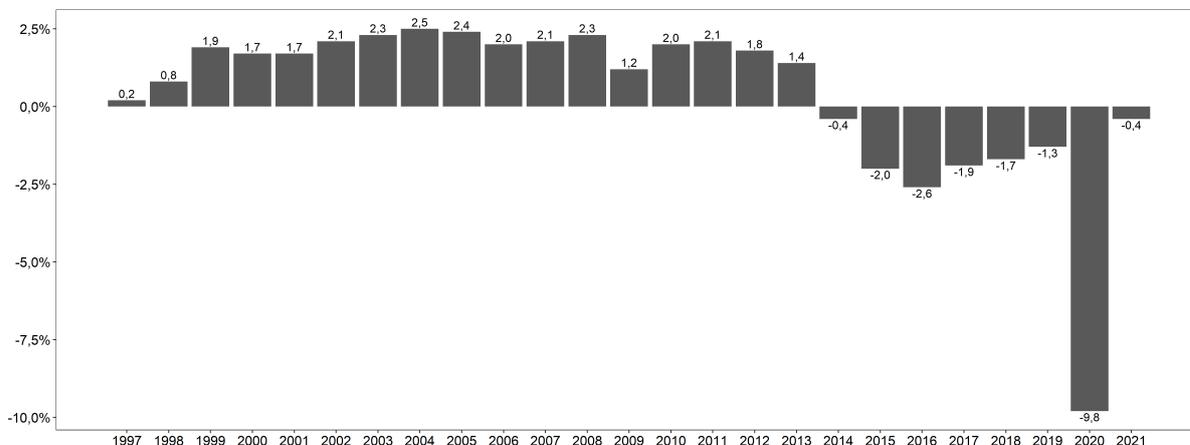
A dinâmica do gasto previdenciário está diretamente relacionada ao envelhecimento populacional e à política de valorização real do salário mínimo, que também afeta programas assistenciais, com exceção do antigo Auxílio Brasil (antigo Bolsa Família). Além disso, a rigidez orçamentária, decorrente do alto nível de gastos primários obrigatórios e a vinculação de boa parte das receitas vinculadas a determinados grupos de gastos públicos.

Desta forma, durante os últimos anos dos anos 1990 e a primeira década de 2000 o desempenho fiscal do Governo Central foi caracterizado por superávits primários anuais superiores a 2% do PIB que, segundo o Ministério da Fazenda (2018), conjuntamente com uma taxa de crescimento mais robusta da economia, contribuíram para a redução gradual da dívida pública. Entretanto, ao se decompor esse desempenho fiscal favorável por seus condicionantes, observa-se que os resultados primários positivos se deram pela elevação contínua da carga tributária, que compensava o crescimento das despesas públicas.

A partir de 2010 é possível observar uma elevação constante, em relação ao PIB, das despesas até o ano de 2016, bem como uma desaceleração e queda da receita, em proporção ao PIB de cada ano, neste mesmo período, caracterizado por um cenário de crise econômica e política, queda do PIB e aumento do desemprego, dificultando o aumento de receitas. Com este movimento, o Brasil passou de superávits primários, observados desde

o final dos anos 90, para déficits primários a partir de 2014 (COTA, 2017).

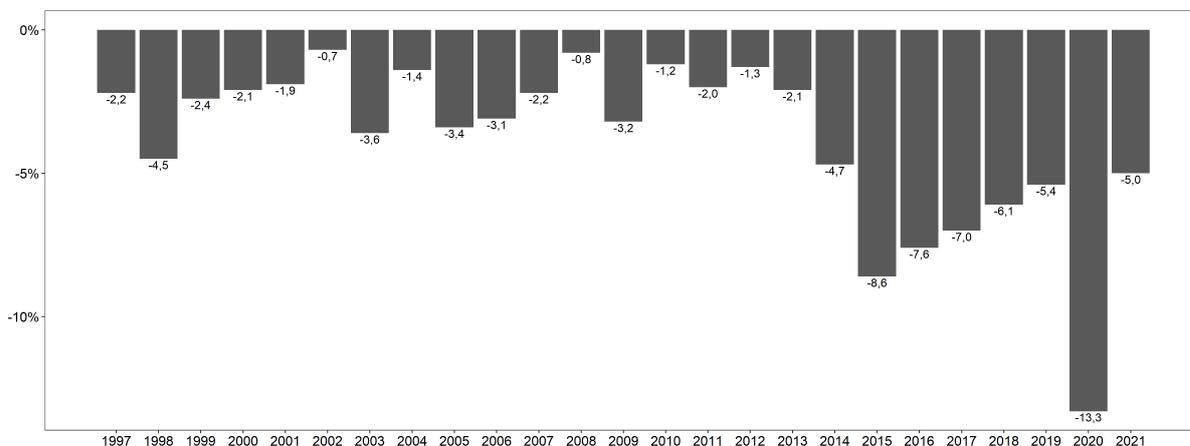
Figura 1.2 – Resultado Primário do Governo Central - acima da linha (% do PIB)



Fonte: Secretaria do Tesouro Nacional. Elaboração própria.

Do Resultado Primário, diferença entre a Receita Líquida e a Despesa Total, ao se acrescentar os juros nominais (juros reais mais a atualização monetária) incidentes sobre a dívida líquida, obtém-se o Resultado Nominal do Governo Central, sendo denominado de déficit nominal se o resultado for negativo, ou superávit nominal, se resultado positivo. Essa diferença corresponde à necessidade de financiamento do Setor Público, montante de recursos que o setor público não financeiro necessita captar com o setor privado, o setor público financeiro e o resto do mundo para fazer face aos seus dispêndios, em razão da insuficiência de suas receitas fiscais. Portanto, identifica-se na Figura 1.3 o comportamento do Resultado Nominal desde 1997.

Figura 1.3 – Resultado Nominal do Governo Central (% do PIB)



Fonte: Secretaria do Tesouro Nacional. Elaboração própria.

É possível identificar que a partir de 2016 há uma tendência de redução da trajetória do resultado nominal justificado pelo alívio gradual na despesa de juros como proporção

do PIB, que entrou em queda em função da redução dos juros básicos da economia (IFI, 2021).

Isso posto, a noção da Receita Corrente Líquida, bem como de outras variáveis orçamentárias, é crucial para as contas públicas, uma vez que é sobre ela que se assentam as bases para a verificação dos cumprimentos dos limites, por exemplo, de gastos com pessoal, da dívida consolidada líquida, das contratações de operações de crédito, de concessão de garantias e do endividamento. Além disso, a elaboração de cenários para as contas públicas - baseados em estimativas de receitas, despesas e resultado primário, por exemplo - surgem como ferramentas primordiais na etapa de formulação do orçamento público (sob uma ótica de longo prazo) ou durante sua execução (sob um olhar de curto ou médio prazo), como também contribuem com as análises relativas às metas fiscais estabelecidas na Lei de Diretrizes Orçamentárias (LDO) (TESOURO NACIONAL, 2016).

Dessa forma, Vecchia e Montoya (2020) e McNichol (2014) destacam que previsões cada vez mais acuradas servem como ferramentas para se confrontar os números definidos inicialmente na proposta orçamentária, proporcionando orçamentos mais realistas, que por sua vez estão associados a maior equilíbrio nas contas públicas. Entretanto, Bretschneider et al. (1989) e Maher e Deller (2011) constatam que a acurácia das previsões de variáveis fiscais é influenciada pelo ambiente político, com gestores públicos sentindo-se incentivados a manipular estimativas de receita, por exemplo, em períodos de pressão fiscal, surgindo como alternativa positiva a criação de mecanismos ou órgãos de monitoramento destas variáveis.

1.4 METODOLOGIA

Existem diferentes métodos de se modelar uma série temporal, iniciando por modelos mais simples e intuitivos, com pouca análise de dados, até métodos mais complexos, que envolvem um trabalho considerável de interpretação destas séries. Métodos estatísticos tradicionais, como os modelos de Suavização Exponencial e os modelos autorregressivos (AR), de médias móveis (MA) e SARIMA são lineares, uma vez que as previsões dos valores futuros são restritas a funções lineares de observações passadas, entretanto possuem relativa simplicidade de compreensão e implementação, sendo o principal foco de pesquisa e ferramentas aplicadas nas últimas décadas.

Como tentativa de superar a restrição dos modelos lineares e tentar explicar certos padrões não lineares observados em problemas reais, diversas classes de modelos não lineares foram propostos na literatura, como os modelos TAR (*threshold autoregressive*), propostos em Tong (1983) e Tong (1990), que são modelos autorregressivos que possuem um limiar, podendo ser úteis para processos em que ocorrem trocas de regimes. A classe de modelos ARCH (*autoregressive conditional heterocedasticity*), introduzida por Engle (1982), são métodos não lineares no que se refere à variância, tendo a ideia básica de que a volatilidade depende de retornos passados por meio de uma função quadrática.

Embora algumas melhorias tenham sido notadas com esses modelos não lineares, segundo Gooijer e Kumar (1992), o ganho de utilizá-los para problemas de previsão é limitado, apresentando melhor acurácia somente com horizonte de previsão curto, como o de um-passo-a-frente (*one-step-ahead*). Como esses modelos são desenvolvidos para padrões não lineares específicos, eles não são capazes de modelar outros tipos de não linearidade em séries temporais.

Mais recentemente, o Facebook desenvolveu um modelo de previsão de séries temporais, o Prophet, que, diferentemente dos modelos tradicionais de suavização exponencial, como Holt Winters e SARIMA, o Prophet adota um modelo bayesiano para suavizar e prever dados de séries temporais. Em outras palavras, o Prophet utiliza diversas funções lineares e/ou não lineares para ajustar os componentes de tendência, sazonalidade e feriados, tendo o tempo como seu único regressor (TAYLOR; LETHAM, 2018).

Apesar da ampla gama de contribuições para a literatura de previsão, Jenkins (1982), Chatfield (1988), Chatfield (2000) e Aiolfi e Timmermann (2006) destacam que é amplamente aceito que não existe um método que seja aplicável a todas as séries temporais, relatando evidências sistemáticas de que alguns modelos de previsão têm desempenho relativo variável ao longo do tempo e que alguns modelos de previsão são persistentemente ruins (ou bons) ao longo da série de tempo.

Isso se deve em grande parte ao fato de que um problema do mundo real geralmente é complexo por natureza e qualquer modelo único pode não ser capaz de capturar padrões

diferentes igualmente bem, além de que, muitas vezes é difícil determinar na prática se a série temporal em análise é gerada a partir de um processo linear ou não linear. Normalmente, vários modelos diferentes são testados e aquele com o resultado mais preciso é selecionado (ZHANG, 2003).

No entanto, um único modelo final selecionado não é necessariamente o melhor para usos futuros devido a muito fatores de influência em potencial, como variação de amostragem, incerteza do modelo e mudança de estrutura da série de tempo. Makridakis et al. (1982), Hibon e Evgeniou (2005), Tascikaraoglu e Uzunoglu (2014), Barak e Sadegh (2016) afirmam que a combinação - *ensemble* - de diferentes modelos podem melhorar a precisão da previsão em comparação com modelos individuais usados separadamente, aproveitando os métodos de previsão individuais que têm desempenhos diferentes dependendo do conjunto de dados, horizontes de previsão, bem como capacidade de capturar não linearidade.

Dentre os diversos métodos existentes de aprendizado conjunto, Zhou (2012) destaca que a combinação dada pela média das previsões individuais é a combinação mais popular e fundamental para previsões numéricas, devido à sua simplicidade e eficácia, sendo a média simples a primeira escolha em muitas aplicações reais.

Tascikaraoglu e Uzunoglu (2014) aponta que outro método utilizado de combinações de modelos é o que possui uma estrutura que leva em conta os valores de erro residual obtido a partir de um modelo de previsão. A importância desta técnica está no fato de que os efeitos negativos dos erros sistemáticos - que implicam em valores super ou subestimados na maior parte do tempo e que podem ser encontrados em todos os modelos de previsão - sobre a precisão da previsão podem ser diminuídos até certo ponto. A incorporação dos valores de erro residual à série prevista leva ao aumento da precisão da previsão.

Nesta área, a combinação de modelos lineares e/ou não lineares tem sido explorada com objetivo de melhorar as previsões quando comparadas com os modelos individuais. Modelos tradicionais como o SARIMA e o ETS são implementados para se detectar padrões lineares das séries temporais, enquanto redes neurais, máquina de vetores de suporte (*Support Vector Machine* - SVM) e *Extreme Gradient Boosting* - XGBoost detectam padrões não lineares, como em Zhang (2003), Oliveira e Ludermir (2016), Panigrahi e Behera (2017) e Li e Zhang (2018), para dados de demanda de energia elétrica, temperatura, desemprego, entre outros.

1.4.1 Modelos ARIMA

Em um modelo autorregressivo integrado de médias móveis, assume-se que o valor futuro de uma variável é uma função linear composta por p observações passadas e q erros aleatórios, assim como demonstrado na equação 1.2.

$$y_t = \sum_{i=1}^p \phi_i y_{t-i} + \epsilon_t + \sum_{j=1}^q \theta_j \epsilon_{t-j} \quad (1.1)$$

em que, y_t e ϵ_t são os valores atuais da série e o termo de erro no período t , respectivamente. ϕ_n e θ_n são os parâmetros do modelo, de forma que $i = 1, 2, \dots, p$ e $j = 1, 2, \dots, q$. Assume-se que o termo de erro, ϵ_t , é independente e identicamente distribuído com média zero e variância constante igual a σ^2 .

O modelo SARIMA é desenvolvido a partir de um modelo ARIMA, levando-se em consideração o fator sazonal, sendo caracterizado por $ARIMA(p, d, q)(P, D, Q)_s$, em que d é o número de diferenças necessários para tornar a série y estacionária e as letras maiúsculas são elementos similares do modelo ARIMA para o modelo sazonal, com s representando o número de períodos da sazonalidade, podendo representar meses, trimestres, anos, etc. Assim:

$$y_t = \sum_{i=1}^p \phi_i y_{t-i} + \sum_{j=1}^q \theta_j \epsilon_{t-j} + \sum_{i=1}^P \Phi_i y_{t-si} + \sum_{j=1}^Q \Theta_j \epsilon_{t-si} + \epsilon_t \quad (1.2)$$

Este modelo é muito semelhante ao modelo ARIMA, exceto que há um conjunto adicional de componentes autorregressivos e de média móvel, com a finalidade de se quantificar efeitos sazonais da série estudada. Esta técnica é amplamente utilizada para analisar e prever séries temporais com sazonalidade em diversas áreas, como finanças, economia e climatologia.

A seleção do modelo será dada pelo método de Box e Jenkins (1970), em que se propõe a utilização das funções de autocorrelação e autocorrelação parcial dos dados amostrais como ferramentas básicas para identificar a ordem do modelo SARIMA, após a identificação os parâmetros são estimados e posteriormente executa-se a verificação de diagnóstico da adequação do modelo, em que se verifica se as suposições do modelo sobre os erros, ϵ_t são satisfeitas.

A popularidade do modelo SARIMA deve-se às suas propriedades estatísticas, bem como a conhecida metodologia Box-Jenkins no processo de construção do modelo. Embora os modelos SARIMA sejam bastantes flexíveis, pois podemos representar vários tipos diferentes de séries temporais, sua principal limitação é a forma linear do modelo assumida previamente, portanto, estes modelos podem não ser adequados para problemas complexos não lineares do mundo real (KHASHEI; BIJARI, 2011).

1.4.2 Prophet

Introduzido por Taylor e Letham (2018), o modelo Prophet foi originalmente proposto para previsão de dados diários com sazonalidade semanal e anual. Entretanto,

posteriormente foi estendido para cobrir mais tipos de dados sazonais, funcionando melhor com séries temporais com forte sazonalidade e séries longas.

O modelo Prophet pode ser considerado um modelo de regressão não linear, seguindo a seguinte estrutura apresentada na equação 1.3.

$$y_t = g(t) + s(t) + h(t) + \epsilon_t \quad (1.3)$$

em que, $g(t)$ descreve a função de tendência que modela mudanças não periódicas no valor da série temporal, $s(t)$ representa mudanças periódicas (por exemplo, sazonalidade semanal e anual), e $h(t)$ captura os efeitos de feriados. O termo de erro, ϵ_t , representa quaisquer mudanças idiossincráticas que não são acomodadas pelo modelo, sendo normalmente distribuído³.

1.4.3 Extreme Gradient Boosting - XGBoost

O algoritmo Extreme Gradient Boosting, proposto por Chen e Guestrin (2016), é um conjunto de métodos de *machine learning* baseados em árvores de classificação e regressão que cria um modelo mais geral e poderoso. O XGBoost é aplicável em problemas de aprendizado supervisionado e visa evitar o ajuste excessivo (*overfitting*), mas também otimizar os recursos de computação. O algoritmo é baseado na ideia de "*boosting*", que combina o resultado de um conjunto de modelos mais fracos para desenvolver um resultado melhor, de forma que cada modelo vai aprendendo através do seu antecessor.

No processo de aprendizado aditivo do XGBoost, o primeiro processo de aprendizado é ajustado a toda amostra de dados e um segundo modelo é então ajustado aos resíduos deste para lidar com as desvantagens de um fraco aprendizado. Este processo de ajuste é repetido até que o critério de parada selecionado seja atingido. A previsão final do modelo é obtida pela soma da previsão de cada aprendiz. Matematicamente, a função geral para a previsão na etapa t é:

$$f_i^{(t)} = \sum_{k=1}^t f_k(x_i) = f_i^{t-1} + f_t(x_i) \quad (1.4)$$

em que, $f_t(x_i)$ é o aprendizado na etapa t , f_i^t e f_i^{t-1} são as previsões nas etapas t e $t - 1$, e x_i é a variável de entrada.

Para treinar o modelo, é preciso definir uma função objetivo para medir o quão bem o modelo está se ajustando aos dados de treinamento. A função objetivo a ser otimizada é dada por:

³ Para mais informações ver: (TAYLOR; LETHAM, 2018).

$$Obj(\theta) = \sum_{i=1}^n L(y_i, \hat{y}_i) + \sum_{k=1}^t \Omega(f_i) \quad (1.5)$$

em que, $L(y_i, \hat{y}_i)$ representa o erro de treinamento da i -ésima amostra, Ω é o termo de regularização, que controla a complexidade do modelo e previne o *overfitting*. O termo de regularização é dado por:

$$\Omega(f) = \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T w_j^2 \quad (1.6)$$

em que, w é o vetor de pontuações (*scores*) nas folhas, T é o número total de folhas, λ é o parâmetro de regularização e γ representa a perda mínima necessária para particionar ainda mais o nó da folha⁴.

1.4.4 Modelos Híbridos

Como é difícil conhecer completamente as características dos dados em um problema real, a metodologia híbrida que possui recursos de dois ou mais métodos surge como uma boa estratégia para o uso prático. Ao combinar diferentes modelos, diferentes aspectos desses dois componentes podem ser capturados, assim como nos modelos SARIMA-XGBoost e Prophet-XGBoost.

O primeiro passo desse método é estimar o modelo SARIMA/Prophet, estando nos resíduos do modelo características da série temporal que não foram capturadas no primeiro momento. Desta forma, seja e_t o resíduo no tempo t , temos que:

$$e_t = y_t - \hat{L}_t \quad (1.7)$$

em que, \hat{L}_t é a previsão pelo método ARIMA/Prophet no tempo t . Ao modelar os resíduos do primeiro método usando o método XGBoost temos que:

$$e_t = f(e_{t-1}, e_{t-2}, \dots, e_{t-n}) + \epsilon_t \quad (1.8)$$

em que f é uma função não linear determinada pelo método XGBoost e ϵ_t é o erro aleatório. Desta forma, as previsões de 1.8 podem ser determinadas como \hat{N}_t , e a combinação das previsões serão dadas por:

$$\hat{y}_t = \hat{L}_t + \hat{N}_t \quad (1.9)$$

⁴ Para informações mais detalhadas, ver (CHEN; GUESTRIN, 2016).

Ou seja, o modelo híbrido consiste em dois passos, primeiro, a estimação do método ARIMA/Prophet e segundo, o método XGBoost é desenvolvido para modelar os resíduos do primeiro passo. Portanto, pode ser vantajoso modelar padrões diferentes separadamente usando dois ou mais modelos e posteriormente combinar as previsões para melhorar a modelagem geral e o desempenho da previsão da série temporal.

1.4.5 Avaliação dos Modelos de Previsão

Para avaliação das previsões dos modelos, as métricas utilizadas serão o RMSE (raiz quadrática do erro médio ou, em inglês, *Root Mean Squared Error*) e o MAPE (erro percentual absoluto médio ou, em inglês, *Mean Absolute Percentage Error*). A primeira métrica possui como vantagens o fato de estar na mesma escala dos dados e de ser conhecido como um indicador de precisão usado para comparar erros de previsão de diferentes modelos para uma determinada variável, entretanto, essa métrica é mais sensível a *outliers*, uma vez que, conforme os valores de erros das instâncias aumentam, o índice do RMSE aumenta consideravelmente, já que os erros são elevados ao quadrado, como descrito abaixo:

$$RMSE = \sqrt{\frac{1}{T} \sum_{t=1}^T (Y_t - \hat{Y}_t)^2} \quad (1.10)$$

A segunda métrica, o MAPE, tem a vantagem de ser independente da escala dos dados, podendo ser utilizado para se comparar a performance das previsões de diferentes dados, medindo a precisão das previsões como uma porcentagem, assim:

$$MAPE = \frac{1}{T} \sum_{t=1}^T \left| \frac{Y_t - \hat{Y}_t}{Y_t} \right| \quad (1.11)$$

Entretanto, segundo Kim e Kim (2016), essa métrica possui restrições para quando a série analisada possui valores zeros, ou próximos de zero, resultando em uma divisão por zero ou valores do índice do MAPE que tendem ao infinito, como é o caso das séries de Resultado Primário e Resultado Nominal. Logo, essa métrica não será utilizada na avaliação destas duas séries.

Nas fórmulas acima, T representa o tamanho da amostra a ser prevista, Y_t e \hat{Y}_t são os valores reais e previstos para o tempo t , respectivamente. De forma geral, quanto menor o valor desse indicador, mais precisa a previsão do modelo, em que um valor de 0 irá indicar uma previsão perfeita.

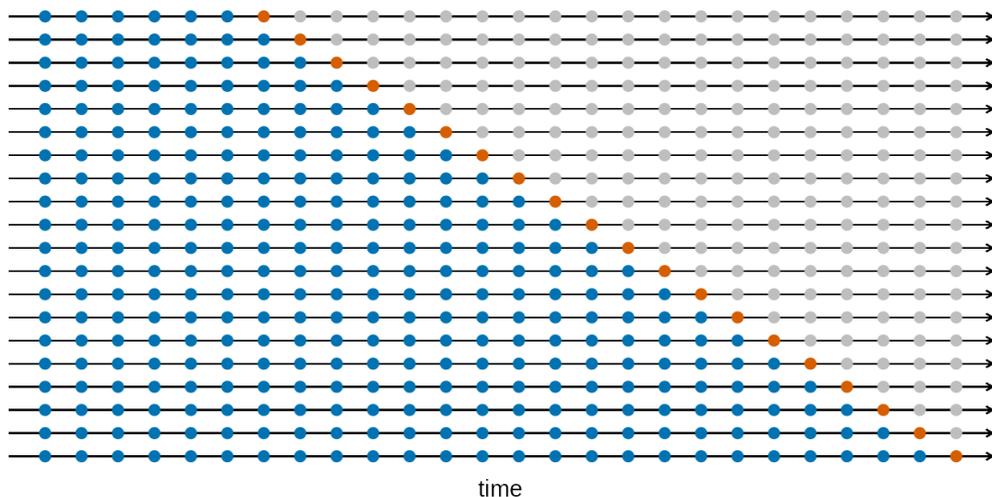
Os modelos serão avaliados utilizando a estratégia de validação cruzada contínua k -passos-à-frente (*cross-validation on a rolling basis k-step-ahead*), em que k corresponde ao horizonte de previsões consideradas e as métricas finais são dadas pelas médias das pontuações obtidas em cada passo e, como forma de representar o valor típico da amostra,

também será calculada a mediana das pontuações obtidas em cada passo - uma vez que esse último cálculo não é distorcido pelos valores extremamente altos ou baixos. Para esse estudo, $k = \{3, 6, 12\}$ meses, abrangendo horizontes mais curtos e mais longos na análise.

A estratégia de validação cruzada em séries temporais consiste em começar com um pequeno subconjunto de dados como amostra de treinamento, fazer previsões para os k pontos de dados posteriores e, em seguida, verificar a precisão dos dados previstos. Posteriormente, os valores reais desses pontos de dados previstos são incluídos como parte do próximo subconjunto de dados de treinamento, efetuando-se a previsão dos k períodos subsequentes.

A Figura 1.4 ilustra a série de conjuntos de treinamento e teste para o caso exemplificativo de $k = 1$, onde as observações azuis formam os conjuntos de treinamento e as observações laranja formam os conjuntos de teste. Assim, a precisão da previsão é calculada pela média ou pela mediana dos conjuntos de teste.

Figura 1.4 – Validação cruzada k-passos-a-frente



Fonte: Hyndman e Athanasopoulos (2021)

Esta estratégia será utilizada pois, segundo Hyndman e Athanasopoulos (2021), as previsões de apenas uma etapa podem não ser tão relevantes quanto as previsões de várias etapas. Nesse caso, o procedimento de validação cruzada contínua permite um resultado mais robusto quanto a precisão das previsões dos modelos.

1.5 DADOS

Neste estudo são utilizados dados das principais variáveis fiscais brasileiras, a saber: i) Arrecadação das Receitas Federais; ii) Receita Líquida do Governo Central; iii) Despesa Total do Governo Central; iv) Resultado Primário do Governo Central acima da linha; e v) Resultado Nominal do Governo Central.

Os dados mensais da série histórica da Arrecadação das Receitas Federais, extraídos da Receita Federal do Brasil, representando a soma das receitas administradas pela Receita Federal com as receitas administradas por outros órgãos. Essa variável considera as receitas tributárias federais, como o Imposto de Renda (Pessoas Física e Jurídica), Imposto sobre Produtos Industrializados - IPI, impostos sobre importações e exportações, e as contribuições previdenciárias (para trabalhadores do setor privado e servidores públicos), entre outros, a Figura A.1 no Apêndice A representa a evolução temporal da Arrecadação Federal.

A Receita Líquida do Governo Central representa todas as receitas tributárias e não tributárias - como receitas de concessões ou dividendos, por exemplo - do Governo Central, excluindo as transferências constitucionais para estados e municípios, demonstrada na Figura A.2.

A Despesa Total do Governo Central considera a soma de todas as despesas primárias do Governo Central - abrangendo o Governo Federal (incluindo a Previdência Social) e o Banco Central -, incluindo, entre outras, os benefícios previdenciários, os subsídios, os investimentos e as despesas de pessoal, representada na Figura A.3 no Apêndice A.

O Resultado Primário do Governo Central representa a diferença entre a Receita Líquida e a Despesa Total do Governo Central. Já o Resultado Nominal é resultado do Resultado Primário mais ajustes metodológicos, discrepâncias estatísticas e os juros nominais incidentes sobre a dívida líquida. Estas séries estão disponíveis no site do Tesouro Nacional ⁵ e estão representadas nas Figuras A.4 e A.5.

A amostra de dados para as séries vai de Janeiro de 1997 até Abril de 2022, totalizando 304 observações. Como estratégia, as séries temporais serão divididas em duas partes, a amostra de treino, usada para estimativas iniciais dos parâmetros, seleção do modelo, dentre outros processos, e a amostra de teste, usada para avaliação do desempenho das previsões de cada modelo estimado.

A amostra de teste corresponderá a 80% dos dados e o restante é utilizado na parte da amostra de teste. Para os modelos de *machine learning* e híbridos as amostras de treino são novamente divididas em conjunto de treino (70% da série original) e conjunto de validação (correspondente a 10% da série original), como forma de obter os melhores

⁵ <<https://www.tesourotransparente.gov.br/visualizacao/series-temporais-do-tesouro-nacional>>

parâmetros dos modelos, evitar o *overfitting* e avaliar o desempenho desses modelos de uma maneira mais robusta.

As amostras de treino começam em janeiro de 1997 e vão até março de 2017, totalizando 243 observações para cada série estudada, as amostras de teste têm seu início em abril de 2017 e vão até abril de 2022, totalizando 61 observações. As amostras de validação dos modelos de *machine learning* terão 31 observações e irão de setembro de 2014 até março de 2017.

Como forma de observar a estacionariedade das séries, foi implementado o teste Dickey-Fuller Aumentado (ADF), desenvolvido em Dickey e Fuller (1981) e Said e Dickey (1984), que possui hipótese nula de não estacionariedade. Caso a hipótese nula seja rejeitada, a série será diferenciada até que a hipótese nula não possa ser rejeitada. Como robustez, a estacionariedade da série temporal será verificada por meio do teste de estacionariedade de Kwiatkowski-Phillips-Schmidt-Shin (KPSS), desenvolvido em Kwiatkowski et al. (1992).

As séries de Arrecadação Federal, Receita Líquida e Despesa Total foram transformadas para termos reais, aplicou-se o logaritmo e a primeira diferença das séries. As séries de Resultado Primário e Nominal foram utilizadas em termos reais e aplicadas a primeira diferença.

1.6 RESULTADOS

Essa seção apresenta os resultados obtidos para as previsões das variáveis econômicas selecionadas. Como o modelo SARIMA é tratado como o *benchmark*, os resultados das métricas estão apresentados de forma relativa ao obtido por este modelo em cada variável e para cada período de tempo. Além disso, os resultados que apresentam melhor acurácia do que o método padrão estão destacados de negrito na Tabela 1.3.

A assimetria dos resultados dos métodos relativamente ao modelo padrão, quando visualizados a média e a mediana das métricas, pode estar sendo influenciada, em partes, pelos pontos discrepantes que podem ter surgido por exemplo no período da pandemia, época em que houve forte mudança na dinâmica econômica, retraindo a atividade - e, conseqüentemente, a arrecadação pelo poder público. Além desse fator, houve também a postergação do pagamento de tributos, o aumento das despesas através da criação de dívida pública e a conseqüente influência nos resultados de todas as variáveis utilizadas nesta pesquisa. Efeitos desse tipo tornam a tarefa de prever séries macroeconômicas uma desafio ainda maior, uma vez que um evento dessa escala nunca foi observado diretamente na amostra disponível, assim como observado em Ludvigson, Ma e Ng (2020) e Schorfheide e Song (2021).

Na primeira parte da tabela são apresentados os resultados referentes à média das métricas (RMSE e MAPE) obtidas em cada uma das previsões para os horizontes de previsões de 3, 6 e 9 meses ($k = 3, 6, 9$). Na parte de baixo da tabela estão presentes os valores medianos das métricas citadas acima, com a finalidade de desconsiderar a influência de resultados excepcionalmente altos ou baixos.

Para a Arrecadação das Receitas Federais, destaca-se que os modelos Prophet-XGBoost, a média dos valores previstos por todos os modelos considerados e a mediana desses valores possuem melhor acurácia do que o modelo padrão SARIMA, principalmente quanto maior o horizonte das previsões efetuadas. Desses modelos, são obtidos resultados até 30% mais precisos do que o obtido pelo modelo *benchmark*.

Ao se observar a acurácia dos modelos para a variável da Receita Líquida do Governo Central identifica-se que apenas o SARIMA-XGBoost não possui maior acurácia do que o modelo padrão, principalmente se considerado apenas a mediana das métricas, indicando que os modelos alternativos possuem melhor acurácia do que o modelo padrão quando desconsiderada as distorções de valores excepcionalmente altos ou excepcionalmente baixos.

Tabela 1.3 – Comparação das previsões

Métrica	Modelo	Arrecadação			Receita			Despesa			Res. Primário			Res. Nominal		
		3	6	9	3	6	9	3	6	9	3	6	9	3	6	9
RMSE	Prophet	1,28	1,23	1,18	1,13	1,17	1,23	1,04	1,14	1,14	1,09	1,08	1,07	1,11	1,08	1,08
	XGBoost	1,24	1,30	1,51	0,99	1,04	1,10	1,29	1,62	1,85	0,86	0,98	0,97	1,00	0,98	1,60
	Arima-XGBoost	1,05	1,07	1,06	1,05	1,14	1,34	1,46	2,09	3,56	0,95	1,00	1,11	1,14	1,16	1,52
	Prophet-XGBoost	1,01	0,97	0,97	0,94	1,00	1,09	1,03	1,12	1,12	0,94	0,99	1,07	1,01	1,33	1,50
	Média	0,98	0,96	0,98	0,90	0,96	1,04	1,00	1,20	1,55	0,88	0,92	0,97	0,96	0,99	1,21
Mediana das métricas	1,01	0,96	0,95	0,94	1,00	1,08	0,98	1,09	1,17	0,87	0,97	1,00	0,98	0,99	1,20	
MAPE	Prophet	1,22	1,18	1,15	1,11	1,22	1,32	1,09	1,19	1,24	-	-	-	-	-	-
	XGBoost	1,26	1,36	1,63	0,97	1,11	1,22	1,30	1,68	2,10	-	-	-	-	-	-
	Arima-XGBoost	1,03	1,07	1,09	1,07	1,20	1,44	1,44	2,09	3,77	-	-	-	-	-	-
	Prophet-XGBoost	0,98	0,97	0,97	0,93	1,05	1,17	1,05	1,16	1,18	-	-	-	-	-	-
	Média	0,95	0,96	0,98	0,88	1,00	1,10	0,99	1,20	1,63	-	-	-	-	-	-
Mediana das métricas	0,97	0,96	0,95	0,92	1,04	1,14	0,98	1,11	1,23	-	-	-	-	-	-	
RMSE	Prophet	1,60	1,62	1,23	1,08	0,92	0,94	0,99	0,86	1,23	1,32	1,20	1,21	1,43	1,09	1,21
	XGBoost	1,38	1,74	1,64	0,89	0,81	0,91	0,94	1,12	1,84	0,91	1,09	1,17	1,21	1,01	2,01
	Arima-XGBoost	1,08	1,36	0,98	1,14	1,07	1,31	1,37	1,49	2,51	0,97	1,41	1,53	1,30	1,18	2,11
	Prophet-XGBoost	1,05	1,09	0,79	0,80	0,67	0,78	0,84	0,86	1,05	1,00	1,11	1,24	1,20	1,46	1,85
	Média	1,22	1,08	0,83	0,81	0,83	0,88	0,76	0,71	1,07	1,00	1,00	1,01	1,12	0,94	1,24
Mediana das métricas	1,03	1,09	0,83	0,94	0,78	0,90	0,83	0,66	1,00	1,00	1,11	1,21	1,14	0,95	1,35	
MAPE	Prophet	1,40	1,48	1,26	0,97	1,02	1,15	1,14	1,16	1,28	-	-	-	-	-	-
	XGBoost	1,40	1,82	1,95	0,71	0,88	1,17	1,07	1,37	1,92	-	-	-	-	-	-
	Arima-XGBoost	1,08	1,50	1,08	0,93	1,03	1,91	1,37	1,84	2,72	-	-	-	-	-	-
	Prophet-XGBoost	1,05	1,04	0,71	0,71	0,79	0,89	0,86	1,01	1,09	-	-	-	-	-	-
	Média	0,98	1,05	0,77	0,68	0,90	1,00	0,89	0,77	1,09	-	-	-	-	-	-
Mediana das métricas	1,04	0,99	0,83	0,65	0,84	1,08	0,80	0,93	1,16	-	-	-	-	-	-	

Fonte: Elaboração própria

Resultado semelhante pode ser observado para a Despesa Total do Governo Central, principalmente quando observada apenas a tendência central das métricas escolhidas. Particularmente para a Despesa Total, a melhor acurácia dos modelos alternativos em relação ao modelo padrão se dá nos menores horizontes de tempo, enquanto que na Receita Líquida esse resultado pode ser observado em todos os horizontes de previsão. Os modelos alternativos apresentam previsões mais precisas em até 33% e 34% para a Receita Líquida e a Despesa Total, respectivamente.

No Resultado Primário do Governo Central acima da linha tem-se que, pela média das métricas obtidas na validação cruzada, apenas o método Prophet não obtém uma melhor acurácia do que o SARIMA, além disso, identifica-se que os métodos alternativos são mais precisos principalmente em horizontes de previsão menores, sendo observado resultado até 13% mais preciso.

Ao se observar a mediana das métricas em comparação aos resultados das métricas obtidos pela média, identifica-se que os resultados dos modelos analisados são influenciados por alguns resultados excepcionalmente baixos, uma vez que há uma piora nos resultados desses métodos em relação ao SARIMA.

Para o Resultado Nominal do Governo Central, identifica-se que nenhum dos modelos é consistentemente melhor do que o modelo SARIMA, uma vez que, dentre os resultados que apresentam maior precisão, esses resultados são sempre próximos do modelo padrão.

De modo geral, os resultados provenientes da média e da mediana das previsões, além do modelo Prophet-XGBoost são os métodos de previsão que mais vezes performam melhor do que o modelo benchmark, além de apresentarem as previsões mais precisas. Em pouquíssimas ocasiões o método SARIMA-XGBoost obteve melhor acurácia do que o método SARIMA, o que indica que, neste caso, a combinação desses dois métodos tornou a previsão pior do que o próprio modelo SARIMA de forma independente.

1.7 CONSIDERAÇÕES FINAIS

Este trabalho teve como objetivo identificar modelos de séries temporais univariados que melhorem o desempenho das previsões de variáveis macroeconômicas brasileiras, como a arrecadação das receitas federais, despesas totais e resultado primário do Governo Central. Para tanto, foram comparados cinco modelos, o primeiro sendo o modelo SARIMA – já tradicional na literatura e considerado, neste trabalho, como o modelo de referência do estudo –, e mais quatro modelos não lineares, utilizando metodologias mais recentes de *machine learning*. Além destes modelos, também foram considerados métodos de combinações de previsões pela média e pela mediana das previsões dos 5 modelos estudados nesta pesquisa.

De forma geral, considerando os resultados obtidos nas 5 séries macroeconômicas avaliadas, tem-se que o desempenho das combinações das previsões – seja pela média ou pela mediana dos valores previstos – estão entre os melhores desempenhos entre os modelos alternativos analisados, principalmente quanto maior forem os horizontes de previsão considerados, independente da série que está sendo analisada. Estes métodos apresentam resultados melhores do que o modelo *benchmark* de forma corriqueira.

Além destas duas metodologias, o método Prophet-XGBoost é outro modelo que apresenta melhores previsões que o modelo SARIMA. Assim, considera-se que, para as variáveis de finanças públicas consideradas nesta pesquisa, o modelo Prophet-XGBoost em conjunto com os métodos citados no parágrafo acima possuem melhor acurácia do que o modelo padrão SARIMA. Em alguns casos, esses modelos obtêm resultados até 35% mais precisos do que o método padrão.

De forma individualizada, encontra-se que a metodologia baseada em árvores de regressão, o XGBoost, possui boa acurácia na previsão da variável Receita Líquida e para o Resultado Primário do Governo Central, principalmente nos menores horizontes de previsão. Além destes, o método Prophet – algoritmo desenvolvido pelo Meta, antigo Facebook – apresenta alguns bons resultados na previsão da Receita Líquida, entretanto, apenas quando observada a mediana das métricas consideradas, indicando ser um modelo que possui mais resultados excepcionalmente altos em relação ao *benchmark*.

Assim, os resultados obtidos nesta pesquisa indicam que os métodos de previsão de séries temporais não lineares mais recentes melhoram o desempenho preditivo das variáveis relacionadas às finanças públicas brasileiras estudadas. Dentre os métodos observados, destacam-se o Prophet-XGBoost – método híbrido resultado da combinação de dois métodos já existentes – e os valores médios e medianos das previsões de todos os modelos considerados. Esses métodos obtiveram resultados mais precisos do que o método padrão de forma regular.

Estes melhores resultados obtidos nas previsões das finanças públicas brasileiras

podem justificar a sua utilização, mesmo com o fato de que, no geral, esses modelos alternativos possuem um maior custo computacional do que o modelo SARIMA. Ainda, a partir destes resultados, trabalhos futuros podem responder se o emprego de variáveis exógenas nesses modelos alternativos pode melhorar ainda mais a predição destas variáveis financeiras.

2 INCERTEZA E ATIVIDADE ECONÔMICA: UM ÍNDICE DE INCERTEZA BASEADO NAS PESQUISAS DO GOOGLE

2.1 INTRODUÇÃO

Desde sua criação, a internet está se tornando progressivamente mais importante para muitos aspectos das vidas das pessoas, das sociedades e dos negócios - independentemente de estas serem usuárias ou não dessa rede de conexões globais que permite o compartilhamento instantâneo de diversos tipos de dados. Isto se dá através da criação de novas plataformas de internet para pesquisa, comércio eletrônico, mídia *online*, negociação financeira, redes sociais e outras atividades nas últimas 3 décadas (LEVIN, 2011).

Atualmente, com o rápido desenvolvimento destas plataformas, uma quantidade muito grande de informações é produzida e coletada diariamente sobre os usuários dessa rede, incluindo dados de mídia social, registros de chamada de telefones celulares, hábitos de consumo, informações geográficas voluntárias, dados de mecanismos de pesquisa e dados de trajetória de táxi, por exemplo.

Este grande volume de informações, conhecido como *Big Data*, tornou-se bastante atrativa para a academia, para a indústria e para governos em todo o mundo. Através do *Big Data* os agentes buscam entender melhor o comportamento da sociedade no presente e mudanças que possam ocorrer nestes comportamentos, bem como aprimorar previsões sobre os mais variados temas, como saúde pública, desenvolvimento econômico e segurança (JIN et al., 2015).

O uso cada vez mais generalizado da Internet por empresas e consumidores levou à criação de uma fonte de dados potencialmente útil: a informação sobre o comportamento de pesquisa na Internet. Os provedores de mecanismos de pesquisa mantêm um registro das pesquisas inseridas em seu site e algumas dessas informações são disponibilizadas publicamente permitindo que os usuários rastreiem a popularidade de uma ampla variedade de termos de pesquisa, este vasto banco de dados pode ser usado para analisar vários problemas.

O comportamento de buscas na Internet pode ser interpretado como uma medida das expectativas reveladas. Presumivelmente, as pessoas pesquisam informações sobre tópicos que desejam aprender mais ou sobre coisas que as preocupam. Por exemplo, se alguém não está se sentindo bem, pode procurar informações sobre os sintomas de saúde que está enfrentando. Da mesma forma, se alguém está preocupado com suas despesas domésticas e acredita que o nível geral de preços está subindo ou que os preços podem

subir em breve, pode procurar informações sobre a inflação.

Entretanto, se alguém não estiver preocupado com a perspectiva de aumento de preços, provavelmente não procurará informações sobre esta, já que a maior parte das pessoas tem mais medo da inflação do que da deflação. Assim, as mudanças no volume de consultas de pesquisa de uma determinada palavra-chave podem ser interpretadas como uma medida de incerteza – ou interesses ou das expectativas reveladas – do público em geral para, como nos exemplos acima, mudanças futuras no nível de preços ou informações sobre questões de saúde.

Para responder essa pergunta, construímos e utilizamos diversas *proxies* para a incerteza doméstica que refletem a ideia de que incerteza trata de eventos que mal sabemos descrever, muito menos computar as suas probabilidades de realização. Incerteza, portanto, é recorrentemente associada a períodos de dispersão de expectativas e volatilidade nos preços dos ativos, além de ser tema recorrente em jornais e na mídia em geral.

Segundo McLaren e Shanbhogue (2011), os dados de pesquisa na Internet têm o potencial de ser úteis para a formulação de políticas econômicas. O monitoramento próximo da atividade econômica atual é um aspecto importante na tomada de decisões, mas a divulgação de estatísticas econômicas oficiais muitas vezes fica para trás, sendo publicada com defasagem. As pesquisas dos usuários da Internet são publicadas mais rapidamente do que as pesquisas oficiais e têm sido normalmente utilizadas para monitorar, entre outras coisas, a atividade econômica atual e o nível de desemprego.

Nesse contexto, Askitas e Zimmermann (2009) e Choi e Varian (2012) destacam que dados do Google *Trends* - uma ferramenta gratuita do Google que permite acompanhar a evolução do número de buscas dos usuários por uma determinada palavra-chave ao longo do tempo e em tempo real, em periodicidade diária, semanal ou mensal, além de estar disponível para muitos países, regiões e até mesmo algumas grandes cidades – para as pesquisas de determinadas palavras-chave estão correlacionados com vários indicadores econômicos, servindo, assim, como uma variável preditora para formulações de boas políticas públicas.

Um pré-requisito para uma boa formulação de políticas macroeconômicas é a informação oportuna sobre o estado atual da economia, especialmente quando a atividade econômica está mudando rapidamente. Dado que variáveis como o PIB e o desemprego estão disponíveis ou trimestralmente ou mensalmente - mas com defasagem na divulgação de até dois meses -, surge como uma ferramenta adicional para os formuladores de políticas e os analistas indicadores criados a partir dos dados disponibilizados pelo Google *Trends* que podem auxiliar na construção de estimativas em tempo real do crescimento do PIB, do desemprego ou outras variáveis (WOLOSZKO, 2020).

A plataforma do Google será utilizada pois, segundo dados da Global Stats¹, o Google se tornou o mecanismo de procura número um em todo o mundo, com trilhões de pesquisas todos os anos e uma participação de mercado de pesquisas on-line de 92,42% no mundo e 96,11% no Brasil em setembro de 2022. Como os dados de pesquisa do Google capturam a maioria das atividades de coleta de informações, ele se qualifica como um *proxy* preciso para quantificar a incerteza no Brasil.

A criação de medidas de incerteza, como a advinda de dados de pesquisas na internet, surge com a difícil tarefa de medir a incerteza, particularmente ao longo dos tempos, uma vez que diante da incerteza os agentes não sabem o que vai acontecer. Dado seu conceito amplo, uma vez que se relaciona com fenômenos macro como o crescimento do PIB e fenômenos micro como a taxa de crescimento das empresas, bem como outros eventos como eleições, guerras e mudanças climáticas (AHIR; BLOOM; FURCERI, 2022).

Assim, o objetivo desse trabalho é construir um índice que quantifique a incerteza econômica no Brasil, utilizando dados mensais de pesquisas em tempo real do Google. O indicador abrange o período de janeiro de 2006 a dezembro de 2019, combinando palavras-chave com conotações de incerteza que sejam pesquisadas pelos usuários. O artigo considera termos retirados das Atas do Comitê de Política Monetária – Copom, que analisam a evolução recente e as perspectivas para a economia brasileira e mundial – portanto, palavras-chave relacionadas ao mundo econômico e financeiro.

Em seguida, este índice será comparado à medidas de incerteza já existentes para o país, a fim de avaliá-lo frente a estes através de uma análise de correlação e como o índice se correlaciona com eventos globais e locais específicos. Por fim, através de um modelo de Vetores Autorregressivos Estrutural (SVAR) com restrição de sinais, observar como variáveis como a atividade econômica e a inflação reagem a um choque de incerteza medido pelo índice GTU-Br.

A inovação desta pesquisa na literatura nacional consiste na construção de um índice de incerteza da economia para o Brasil pesquisando palavras-chave relacionadas à incerteza que são pesquisadas na Internet, diferentemente de outros índices que são construídos a partir de *surveys* ou informações coletadas nos principais jornais no país e a frequência de termos utilizados nas matérias e veiculações que remetem à incerteza, assim como descrito em Filho (2014). Diferentemente destes, o banco de dados do Google *Trends* está disponível gratuitamente e quantifica a popularidade relativas às buscas no mecanismo do Google por agentes que podem representar não somente os consumidores, mas também o lado das firmas e até do governo.

Outras vantagens dessa metodologia para a criação do índice de incerteza é a disponibilidade em tempo real dos dados do Google *Trends*, bem como o fato de dispor

¹ Global Stats. Search Engine Market Share Worldwide, 2022. <<https://gs.statcounter.com/search-engine-market-share>>

informações em termos regionais/estaduais e permitir que estas informações possam ser usadas para mensurar incertezas por categorias (financeira, econômica, política, etc.). (BONTEMPI et al., 2021)

Além dessa primeira seção introdutória, este trabalho possui mais cinco seções. A segunda seção apresenta a revisão da literatura referente aos temas abordados na pesquisa, como índices de incerteza e a utilização das informações de ferramentas de pesquisa na Internet como variáveis explicativas na estimação de variáveis econômicas. A terceira seção detalha como será construído o índice de incerteza GTU-BR. A seção seguinte descreve a metodologia e os dados que são utilizados na pesquisa, com a quinta seção apresentando os resultados obtidos na estimação do modelo principal do VAR, bem como os testes de robustez deste modelo. No final, a última seção conclui esta pesquisa.

2.2 REVISÃO DA LITERATURA

2.2.1 Medidas de Incerteza

Incerteza na economia refere-se à clareza, ou falta dela, sobre a atividade econômica futura. Ela incorpora tanto o “risco” - caso em que as probabilidades de resultados potenciais são conhecidas, porém não é conhecido qual resultado ocorrerá -, bem como a incerteza do ponto de vista de Knight (1921), primeiro autor a introduzir a incerteza no contexto econômico, em que nem as probabilidades de resultado nem o resultado final são conhecidos. Na prática, os dois são difíceis de se separar e compreender a incerteza econômica é bastante difícil porque não é diretamente observável, sendo um vasto campo da literatura econômica a tentativa de se medir a incerteza e avaliar como uma alta incerteza afeta a economia (WATKINS, 1922); (CAGLIARINI; HEATH et al., 2000); (MOORE, 2017).

Em Bloom (2009), o autor constata que um grande choque temporário de incerteza – como ataques terroristas e choques de petróleo - causa uma rápida queda, recuperação e superação do emprego e da produção industrial. De forma que, imediatamente após um choque, um maior nível de incerteza faz com que as empresas pausem temporariamente seus investimentos e contratações, resultando em queda da produtividade. No médio prazo, o aumento da volatilidade do choque induz a um excesso de produção, emprego e produtividade. Assim, choques de incerteza geram recessões e recuperações curtas e agudas.

Além disso, Bloom (2014) observa que países em desenvolvimento tendem a ter um maior nível de incerteza, com países da África e da América do Sul possuindo taxas de crescimento do PIB, mercados de ações e taxas de câmbios mais voláteis do que países de renda alta. O autor atribui o maior nível de incerteza a três fatores, são eles: i) o fato de que países em desenvolvimento tendem a ter economias menos diversificadas, estando mais expostos à variação do pequeno número de produtos dos quais dependem; ii) o fato de que muitos dos bens nos quais estes países se concentram também tem preços mais voláteis, como *commodities* como a borracha, o açúcar, o petróleo e o cobre; e iii) o fato de que os países de baixa renda parecem ter mais choques políticos internos como golpes, revoluções e guerras, além de serem mais suscetíveis a desastres naturais como epidemias e inundações, bem como políticas de estabilização fiscal e monetária menos eficazes do que as observadas nos países de alta renda.

Bontempi, Golinelli e Squadrani (2016) destacam que, maiores níveis de incerteza estão relacionados, principalmente, à má qualidade das informações disponíveis, o que alimenta a necessidade dos indivíduos de coletar mais informações para preencher a lacuna entre o que é conhecido e o que precisa ser conhecido para poder tomar melhores decisões. Desta forma, níveis mais altos de incerteza levam a uma maior demanda por conhecimento

e, conseqüentemente, resultam em um uso mais intenso de ferramentas capazes de coletar mais informações.

Assim, existem algumas formas de se tentar quantificar a incerteza na literatura econômica. O primeiro dos casos - utilizado em Filho (2014), Baker, Bloom e Davis (2016) e Moore (2017) para o Brasil, os Estados Unidos e a Austrália, respectivamente – captura a incerteza baseado em artigos de jornais que fazem referência à incerteza econômica. A ideia por trás dessa metodologia é de que as informações sobre incerteza econômica podem estar contidas na cobertura do jornal. Esse tipo de medida baseada em jornais capturam uma ampla gama de incertezas (ao contrário de medidas baseadas em finanças, por exemplo) e são extremamente oportunas, uma vez que podem ser executadas até diariamente.

Outra medida, baseada nas finanças, utilizada em Romer (1990), Bekaert, Hoerova e Duca (2013), Baker e Bloom (2013) e Caggiano, Castelnuovo e Groshenny (2014), recorrem a volatilidade do mercado de ações como uma *proxy* para a incerteza. Essa metodologia tem como benefício o fato de ser possível obter uma série temporal muito longa de forma que, mesmo que os indivíduos não participem do mercado acionário, eles podem utilizar esse índice como um sinal do estado da economia. Entretanto, a volatilidade de ações está apenas indiretamente conectada à atividade econômica, limitando de certa forma a utilização deste índice.

Uma terceira *proxy* utilizada para se quantificar a incerteza é a dispersão entre previsores para variáveis econômicas, entre elas a do crescimento econômico. Segundo Moore (2017), o aumento da incerteza econômica amplia a distribuição potencial dos resultados, logo, isso deve ser identificado com maior dispersão entre as previsões dos analistas. Usadas em Sill et al. (2012), Bachmann, Elstner e Sims (2013), Banerjee, Kearns e Lombardi (2015) e Jurado, Ludvigson e Ng (2015), medidas desse tipo tem como benefício o fato de estarem inteiramente ligadas conceitualmente à atividade econômica.

Mais recentemente, através de técnicas de análise textual, diversos índices foram estimados na literatura, entre eles o de incerteza, a partir de relatórios como, por exemplo, os disponibilizados por diversos Bancos Centrais, uma vez que sua transparência é vista como fundamental para controlar as expectativas dos agentes econômicos. Metodologia de certa forma recente, o avanço dessa temática ainda é embrionário, destacando-se alguns trabalhos como: Graminho et al. (2015), Nopp e Hanbury (2015), Ferreira et al. (2017), Godeiro e Lima (2017), Barboza e Zilberman (2018) e Silva et al. (2020).

2.2.2 Choques de Incerteza

Nesta seção serão discutidos os canais de transmissão da incerteza em diversas variáveis da economia, destacando diferenças entre o cenário nacional e internacional, se houver. A partir do que for ponderado nessa seção serão definidos quais as restrições de

sinal a serem aplicadas no modelo SVAR apresentado na Seção 2.4 e, conseqüentemente, nas funções impulso-resposta das variáveis que serão analisadas.

2.2.2.1 Atividade econômica

A literatura teórica enfatiza diversos canais de transmissão da incerteza sobre a atividade econômica. Uma constatação ao se observar a literatura é de que não existe um consenso sobre quais os meios que a incerteza impacta sobre as variáveis macroeconômicas, além disso, também não há consenso em relação à direção deste impacto, podendo ter efeitos que ocasionem em recessão econômica ou crescimento econômico².

Nessa vasta literatura, Bloom (2014) destaca quatro mecanismos no impacto da incerteza sobre o ciclo econômico, sendo o mecanismo de *real options* (opções reais) e o mecanismo *precautionary savings* (poupanças precaucional) mecanismo que identificam um efeito contracionista em períodos de aumento de incerteza. Além desses, o autor destaca mais dois mecanismos em que a incerteza pode potencialmente ter um efeito positivo no crescimento de longo prazo, sendo o mecanismo de *growth options* (opções de crescimento) e o mecanismo conhecido como o efeito Oi-Hartman-Abel.

O canal *real options* destaca que a incerteza afeta os investimentos e a contratação de trabalhadores, reduzindo-os, e, conseqüentemente, impacta de forma negativa sobre o crescimento econômico. Ou seja, quando o investidor tem uma informação imperfeita sobre os retornos futuros de projetos de longo prazo, e considerando um cenário de aumento de incerteza, a escolha intuitiva é de adiar as decisões de investimento e esperar por mais informações antes de agir (BERNANKE, 1983).

Bloom (2009) desenvolve um modelo de choque de incerteza e custos de ajustamento que identifica que após um choque temporário sobre a incerteza, o crescimento do emprego, do produto e da produtividade cai drasticamente e a taxa de investimento diminuem. Ainda sobre o mecanismo de opções reais, Caggiano, Castelnuovo e Groshenny (2014) identificam maior volatilidade do mercado de ações, o que geralmente leva a prejuízos maiores por parte dos investidores e diminuição da confiança em relação à economia por parte das firmas, reduzindo o nível de emprego significativamente.

O mecanismo de poupanças precaucionais – destacado, por exemplo, em Bansal e Yaron (2004) – argumenta que um aumento no risco de incerteza leva os consumidores a aumentar a poupança preventiva, reduzindo os gastos de consumo. No curto prazo os efeitos sobre uma economia se apresentam como contracionistas, entretanto, os autores afirmam que os efeitos no longo prazo não são tão claros, afinal, em teoria, um menor consumo e maior poupança podem permitir um aumento no investimento, o que pode beneficiar o crescimento de longo prazo.

² Bloom (2014) e Dou et al. (2020) apresentam uma vasta revisão desta literatura.

Quanto ao mecanismo de opções de crescimento, o argumento é de que a incerteza pode encorajar o investimento se aumentar o tamanho do prêmio potencial. Bar-Ilan e Strange (1996) observam que, se as empresas demoram muito para concluir os projetos – talvez devido ao tempo de construção ou de desenvolvimento de algum produto –, a incerteza pode ter um efeito positivo no investimento. Nessa situação, um aumento no risco pode significar maior lucro esperado quando o produto for lançado no mercado ou na finalização da construção.

Segundo Bloom (2014), esse mecanismo foi muito invocado para explicar o boom das empresas pontocom no final dos anos 1990, indicando que as empresas não tinham certeza sobre a Internet, mas essa incerteza incentivou o investimento. Kraft, Schwartz e Weiss (2018) mostram como as opções de crescimento são particularmente importantes para empresas intensivas em pesquisa e desenvolvimento, tanto que uma maior incerteza pode aumentar o valor de suas ações.

O outro mecanismo que segue a linha de que a incerteza pode potencialmente aumentar o crescimento no longo prazo é conhecido como efeito Oi-Hartman-Abel, desenvolvido em Oi (1961), Hartman (1972) e Abel (1983), e destaca a possibilidade de que, se as empresas puderem expandir para explorar bons resultados e contrair para se proteger contra resultados ruins, elas podem amar o risco. No entanto, para que esse mecanismo funcione, as empresas precisam ser capazes de expandir ou contrair facilmente em resposta a boas ou más notícias. Para Bloom (2014) o efeito Oi-Hartman-Abel normalmente não é muito forte no curto prazo devido aos custos de ajuste, sendo mais poderoso no longo prazo.

2.2.2.2 Inflação

Faltam evidências que estabeleçam concretamente como a inflação reage a um aumento da incerteza. Como abordado na Seção 2.2.2.1, em períodos marcados por futuro mais incerto, as famílias adiam o consumo no período t para aumentar a poupança e as empresas adiam suas decisões de investimento, o que leva a uma diminuição da produção econômica no curto prazo. Em relação a inflação, a literatura não apresenta resultados tão homogêneos.

Leduc e Liu (2016) apontam – através de um modelo DSGE e de evidências empíricas – que choques de incerteza ocasionam em queda considerável na inflação e um crescimento significativo do desemprego. Em contrapartida, Fasani e Rossi (2018) mostram que a inflação reage positivamente a choques de incerteza e Born e Pfeifer (2014) destacam que, sob a ótica da oferta, empresas sujeitas à rigidez de preços têm o incentivo de definir preços acima do seria definido em um cenário com ausência de incerteza para evitar a perda de lucros caso condições econômicas favoráveis se concretizarem no futuro.

A falta de um resultado concreto acerca do movimento da inflação em período de

elevação da incerteza pode ser identificado em Meinen e Roehe (2018), que estimam os impactos sobre a evolução do nível geral de preços a partir de choques de incerteza nos Estados Unidos e na Zona do Euro e concluem que a resposta dos preços é ambígua em relação ao choque, não sendo possível determinar um efeito positivo ou negativo sobre a inflação. Resultado semelhante é observado para o Brasil em Duarte (2021).

2.2.2.3 Juros

Caggiano, Castelnuovo e Nodari (2017) comprovam que, em um cenário de elevada incerteza, as taxas de juros diminuem tanto em períodos de recessão como em períodos de crescimento econômico. Os autores detectam que a diferença existente está relacionada à dinâmica e a resposta quantitativa dessa redução. Em tempos de expansão, a diminuição da taxa de juros é menor e possui uma reação de curta duração, já em períodos de recessão, a redução da taxa de juros é maior e mais longa.

Na mesma linha, Castelnuovo (2019) argumenta que, seguindo o mecanismo de poupança precaucionaria, em períodos de maior incerteza nos Estados Unidos, parte dos recursos das famílias são transferidos do consumo para a poupança e pode acontecer que os títulos governamentais de curto prazo possam ser percebidos como os de menor risco, recebendo maior parte da demanda desse aumento por poupança e, conseqüentemente, apreciando o seu valor e reduzindo os juros.

Duarte (2021) argumenta que, no Brasil, o comportamento dos juros brasileiros pode se dá de forma diferente da observada nos países desenvolvidos, sendo observado a elevação dos prêmios de risco dos títulos do Tesouro, o que sugere um aumento na taxa de juros de longo prazo. Argumenta ainda que se espera que em um cenário de baixo crescimento econômico a expectativa é de uma política monetária mais estimulativa, ou seja, ao menos uma manutenção dos juros de curto prazo em um primeiro momento após o choque, variável de interesse deste trabalho.

Similarmente, Besarria et al. (2021) encontram para o Brasil que o aumento da incerteza posterga a tomada de decisão a respeito do processo de produção, o que provoca efeitos que reduzem o emprego, o capital e a taxa de juros, seguindo o canal de opções reais.

2.2.2.4 Mercado de ações

Gilchrist, Sim e Zakrajšek (2014) destacam que mercados financeiros são afetados pela incerteza, podendo ter impactos negativos para a demanda por ativos e, também, impactos adversos no lado da oferta da economia. Os autores destacam que durante períodos de maior incerteza os investidores exigem maiores compensações como seguro contra riscos futuros, o que implica em redução dos preços dos ativos e da riqueza dos

investidores que já possuem esses ativos. Além disso, os autores detectam que os preços dos ativos tendem a ser mais voláteis durante esse período.

Mais especificamente sobre o mercado acionário, Brown, Harlow e Tinic (1988) desenvolveram a *Uncertain Information Hypothesis* – UIH (Hipótese de Informação Incerta), teoria que parte da concepção de que, em momentos de aumento de incerteza, investidores racionais e que são avessos ao risco reagem estabelecendo preços dos ativos significativamente abaixo do valor esperado em função do aumento do risco associado a esses ativos e, conseqüentemente, do retorno esperado. Pástor e Veronesi (2013) destacam que quanto maior for a incerteza, maior será o retorno exigido pelo investidor, ou seja, maior será o prêmio de risco de mercado.

Caldara et al. (2016) encontram que o aumento da incerteza econômica é uma notícia ruim para o mercado de ações. Uma vez que o aumento persistente da incerteza e o aperto associado das condições do mercado financeiro levam a uma queda acentuada e imediata no mercado de ações.

2.3 ÍNDICE GOOGLE DE INCERTEZA ECONÔMICA - GTU-BR

Esta seção ilustra a metodologia por trás da construção do índice de incerteza do Google *Trends*. A suposição central deste índice é que os agentes econômicos, representados pelos internautas, procuram informações online quando estão incertos. Essa suposição implica que a frequência de busca de termos que podem estar associados a eventos futuros incertos é alta quando o nível de incerteza é alto. Portanto, construir esse índice requer especificar os meios pelos quais a frequência de busca é medida e definir os termos associados à incerteza econômica.

Primeiro serão selecionadas subjetivamente um amplo conjunto de palavras-chave que são frequentemente citadas nas Atas da Reunião do Comitê de Política Monetária - Copom, divulgadas pelo Banco Central do Brasil, em correspondência às incertezas sobre as condições econômicas futuras. Exemplos dessas palavras são “inflação”, “reformas econômicas”, “*commodities*” e “preço dos combustíveis”. Segundo Castelnovo e Tran (2017), esses termos estão relacionados a decisões de política ou eventos que afetam os mercados financeiros que provavelmente injetam incerteza no sistema econômico e, portanto, sugerem uma pausa no investimento e uma queda na demanda de trabalho por parte dos empresários.

Posteriormente, as pesquisas desse grupo de palavras são condicionadas a uma determinada área geográfica, todo o território brasileiro, e, através do Google *Trends* são recuperadas a frequência de pesquisa de cada uma dessas palavras na amostra de interesse. Em seguida, os resultados de toda pesquisa individuais são agregados por mês, obtendo assim o índice de incerteza do Google *Trends* para o Brasil, o GTU-BR.

Os dados do Google *Trends* fornecem a frequência de um determinado termo de pesquisa em relação ao volume total de pesquisa. Para isso, o Google *Trends* divide cada ponto de dados brutos $R_{i,j,m}$ - ou seja, a frequência de uma palavra i em um grupo de palavras pesquisadas j em um determinado mês m - pelo total de pesquisas T no mesmo mês, ou seja:

$$S_{i,j,m} = \frac{R_{i,j,m}}{T_m} \quad (2.1)$$

Os números resultantes são então redimensionados para variar entre 0 e 100, sendo este último valor imposto à palavra i mais pesquisada no grupo de palavras j pesquisadas pelos usuários de internet. Formalmente, a frequência relativa de uma palavra i em um conjunto de palavras pesquisadas j é dada por:

$$FI_{i,j} = 100 * \frac{S_{i,j,m}}{\max(S_{i,j,m})} \quad (2.2)$$

Dado que o Google só permite inserir no máximo 5 termos de pesquisa diferentes no Google *Trends* ao mesmo tempo, um termo de referência foi escolhido para fins de agregação. Isto é feito em três etapas, com a primeira sendo a inclusão de cinco termos de busca no Google *Trends*, e um desses termos – associado à frequência FI_y^* – foi escolhido para ser o *benchmark*. Em segundo lugar, o termo de referência foi incluído junto com outros quatro novos termos selecionados aleatoriamente do conjunto de palavras-chave escolhido, a busca continuou com quatro novos termos mais o de referência até a rodada final, onde os últimos z termos (com $z \leq 4$) foram pesquisados juntamente com o termo de referência.

Ao finalizar este passo, é possível observar que para cada rodada j de buscas de palavras-chaves, a frequência do termo de referência $FI_{y,j}$ pode ser potencialmente diferente daquelas calculadas nas rodadas anteriores, pois o termo de pesquisa mais alto na nova combinação de cinco termos de pesquisa (ou menos, na última rodada) é definido automaticamente para ter um máximo de 100. Assim, na terceira e última etapa, a frequência $FI_i^{x,*}$ de uma palavra x que usamos para construir o índice foi obtida pelo cálculo da razão:

$$FI_i^{x,*} = FI_{i,j}^x * \frac{FI_y^*}{FI_{y,j}} \quad (2.3)$$

Esta última equação remove a ligação existente da frequência de cada termo i do conjunto de palavras-chave com o qual foi pesquisado. Assim, o índice GTU-BR é obtido através de:

$$GTU - BR = \sum_{i=1}^N FI_{i,m,c}^{x,*} \quad (2.4)$$

Em que N representa o número de palavras-chave pesquisadas, que no caso desse trabalho é de 77 palavras³. Após a construção do índice, abordamos a potencial sazonalidade nas buscas de palavras ajustando sazonalmente o índice GTU-BR, bem como a normalização da série para que a média seja igual a 100 e o desvio-padrão igual a 30.

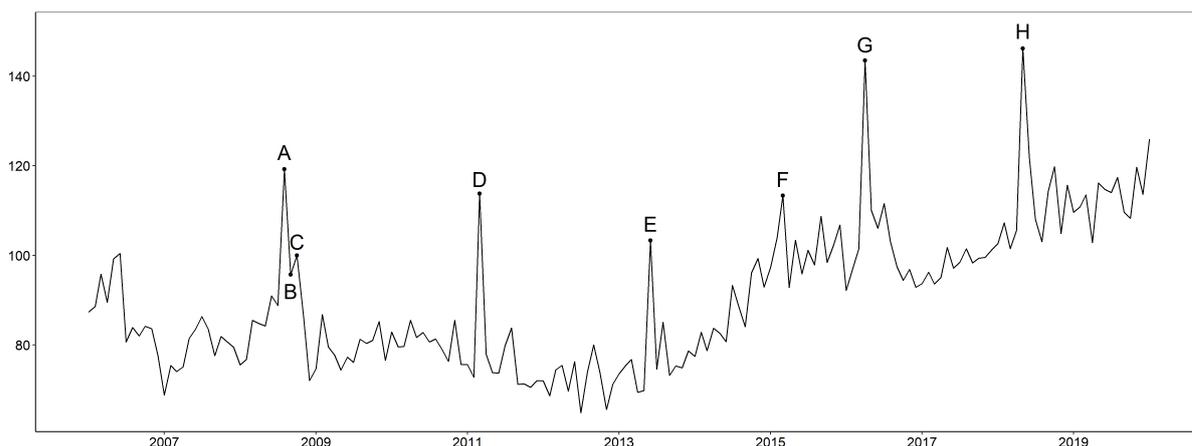
Assim, a Figura 2.1 mostra a evolução do índice de incerteza construído para o Brasil para o período iniciado em janeiro de 2006 e finalizado em dezembro de 2019. Este período analisado se deve ao fato de que para a análise do modelo VAR, a série relativa à mercado de trabalho utilizada tem como fonte o CAGED, pesquisa que foi descontinuada em dezembro de 2019, não havendo pesquisa do mercado de trabalho que possua mais observações disponíveis do que esta série.

Os maiores picos observados estão relacionados a termos como “China”, “Estados Unidos”, “Gasolina”, “Protesto”, “Impeachment”, “Pandemia”, “Japão”, entre outros.

³ As palavras utilizadas para criação do índice GTU-BR estão disponíveis no Apêndice B.

Existem vários saltos significativos no índice, a maioria deles está relacionada a choques externos de incerteza econômica e financeira nos mercados globais, proveniente, por exemplo, de eventos como a crise do subprime iniciada nos Estados Unidos ou eventos naturais de grande impacto, como terremotos e tsunamis que atingiram o Japão em 2011 e seus efeitos colaterais na economia brasileira.

Figura 2.1 – Índice de Incerteza - GTU-BR



Fonte: Elaboração própria

Os pontos A, B e C da Figura 2.1 são os meses de agosto, setembro e outubro de 2008, seus picos estão relacionados ao aumento de procura, principalmente, pelas palavras "China", "Estados Unidos", "Europa" e "Bovespa", período caracterizado pela crise do *subprime* que iniciou ainda em 2007 nos Estados Unidos, mas que se estendeu gradualmente afetando o sistema financeiro de todo o mundo. Entre agosto e outubro de 2008 diversas instituições entraram com pedido de recuperação judicial, sendo necessária a utilização de pacotes de ajuda e nacionalização de bancos em diversos desses países, além de fortes estímulos monetários nas economias.

O ponto D, março de 2011, está relacionado ao forte interesse pelas palavras "Japão", "China" e "Gasolina", mês em que o Japão foi atingido por terremoto, que gerou tsunamis e um acidente nuclear no país. O ponto E, representa o pico de pesquisas por palavras como "Protesto" e "Seca" em junho de 2013, época em que houve uma série de protestos populares no Brasil.

Em março de 2015, ponto F, as palavras mais pesquisadas foram "Desemprego", "China" e "Euro". Já o ponto G, que representa abril de 2016, está relacionado à pesquisas da palavra "Impeachment", mês em que houve abertura do processo contra a então presidente do país. A última data relevante, maio de 2018, está fortemente relacionada à palavras como "Gasolina", "Combustível" e "Seca".

Além dos eventos internacionais, o índice é capaz de identificar importantes eventos nacionais que estão relacionados com o aumento da incerteza econômica, como a grande

procura por combustíveis – principalmente a gasolina – em períodos de grande elevação do preço destes, ou períodos marcados por protestos sociais que tiveram impacto na atividade econômica daquele período. Ou seja, o índice GTU-BR correlaciona com sucesso importantes eventos específicos da economia global e brasileira.

2.3.1 Correlação com índices de incerteza existentes

A Tabela 2.1⁴ mostra a correlação entre o índice GTU-BR e uma variedade de diferentes *proxies* para a incerteza propostos na literatura e disponíveis em frequência mensal. Os índices selecionados foram o Índice de Incerteza Econômica (IIE-BR) da FGV, desenvolvido por Ferreira et al. (2017), a volatilidade do Ibovespa (IBOV)⁵, o Índice de Incerteza da Política Econômica (IIPE), derivado das atas do Copom, desenvolvido por Silva et al. (2020) e o índice EPU⁶ (Economic Policy Uncertainty) que segue a metodologia utilizada em Baker, Bloom e Davis (2016) - e utiliza arquivos de textos do Jornal Folha de São Paulo, contando o número de artigos que contêm os termos "incerto", "incerteza", "econômico" ou "economia" e um ou mais dos seguintes termos relevantes para a política: regulação, déficit, orçamento, imposto, banco central, alvorada, planalto, congresso, senado, câmara dos deputados, legislação, lei e/ou tarifa.

Tabela 2.1 – Correlação entre os índices de incerteza

	GTU-BR	IIE-Br	IBOV	EPU	IIPE
IIE-Br	0,82	1			
IBOV	0,44	0,48	1		
EPU	0,39	0,53	0,18	1	
IIPE	0,10	0,10	0,04	0,25	1

Fonte: Elaboração própria

As correlações entre o GTU-BR e os demais índices são todas positivas num intervalo entre 0,10 (com o índice IIPE) e 0,82 (com o índice IIE-BR). A associação com o índice IIPE - que mensura o sentimento de incerteza contido nas atas de reunião do Copom - representa o menor grau de relação do GTU-BR com os índices observados. No geral, interpreta-se que o sinal positivo dessas correlações sinaliza que o índice gerado é capaz de captar a mesma tendência ou sinal dos demais índices existentes e vastamente utilizados na literatura, uma vez que são observadas associações positivas do GTU-BR com estes índices.

⁴ Como os dados do IIPE são trimestrais, os demais índices que são mensais foram transformados para trimestrais usando a média dos 3 meses que fazem parte do trimestre.

⁵ Essa variável foi calculada através da variância mensal do índice diária de fechamento do Ibovespa

⁶ Disponível em: <www.policyuncertainty.com>

2.4 MODELO VAR COM RESTRIÇÕES DE SINAIS

Uma vez desenvolvido o índice, busca-se analisar o impacto que um choque no índice de incertezas na economia real. Para tanto, será utilizado um modelo de Vetores Autorregressivos (VAR) - desenvolvido inicialmente por Sims (1980) e difundido rapidamente como ferramenta para se investigar empiricamente a macroeconomia, em consonância com diversos estudos da literatura como Bloom (2009) e Bachmann, Elstner e Sims (2013). Neste modelo serão incorporados o índice de incerteza do Google *Trends* mais um conjunto de variáveis macroeconômicas do Brasil, descritas posteriormente, que são tratadas como endógenas.

De forma que, os modelos VAR são sistemas de equações simultâneas que capturam a existência de relações de interdependência entre variáveis e examinam relações lineares entre cada variável e os valores defasados dela própria e de todas as demais variáveis assim os modelos VAR estruturais podem ser expressos genericamente a seguir:

$$AX_t = \sum_{i=1}^p A_i X_{t-p} + \epsilon_t \quad \forall \quad t = 0, 1, \dots, T \quad (2.5)$$

Em que, do lado esquerdo da equação, A é uma matriz $n \times n$ de impactos contemporâneos, X_t é um vetor-coluna de variáveis endógenas, do lado direito, A_i representa uma matriz $n \times n$ de coeficientes das variáveis endógenas defasadas X_{t-p} , e ϵ_t denota o resíduo, que segue um processo $i.i.d \sim (0, \sigma)$.

Ao assumir que a matriz de impactos contemporâneos A da Equação 2.5 seja não singular, então pode-se obter o modelo VAR na sua forma reduzida ao se multiplicar ambos os lados da equação por A^{-1} , assim:

$$X_t = \sum_{i=1}^p B_i X_{t-p} + u_t \quad \forall \quad t = 0, 1, \dots, T \quad (2.6)$$

Em que $B_i = A^{-1}A_i$ e $u_t = A^{-1}\epsilon_t$, além da matriz de covariância dos resíduos do modelo ser representada por $\Omega = E[u_t u_t']$.

É necessária a imposição de algumas restrições sobre os coeficientes da Equação 2.5 para conseguir recuperar as informações do sistema primitivo através do modelo reduzido (Equação 2.6). Entre as alternativas existentes, esta pesquisa utilizará a decomposição recursiva de Cholesky para identificar as restrições. Como proposto em Sims (1980), a decomposição de Cholesky impõe restrições sobre a matriz de impactos contemporâneos, tornando-a uma matriz triangular inferior, o que permite a obtenção dos valores dos choques primitivos estruturais (ϵ_t) através dos resíduos estimados através do vetor u_t .

Dentro desse contexto, Uhlig (2005) desenvolveu um método de identificação na qual se dispensa a imposição de restrições sobre toda a matriz A , esse método ficou

conhecido por modelo VAR com identificação agnóstica de sinais. Esse método consiste em mostrar que a matriz A , tal que $\hat{\Sigma} = AA'$, pode ser definida como $A = \tilde{A}Q$, onde Q é uma matriz ortogonal e \tilde{A} é a decomposição de Cholesky da matriz de variância dos resíduos ($\hat{\Sigma}$) abordada anteriormente.

Em seguida, deve-se identificar uma única coluna a , conhecida como vetor impulso, pertencente à matriz A da Equação 2.5, e que contém as respostas contemporâneas das n -variáveis endógenas a um determinado choque. Com isto, o novo problema se torna a determinação do vetor α associado ao vetor m -dimensional, de modo que:

$$a = \tilde{A}\alpha \quad (2.7)$$

Uhlig (2005) demonstrou que dado o vetor impulso (a), é possível calcular a resposta apropriada ao se admitir que $r_i(k)$ é o impulso-resposta no período k do i -ésimo choque obtido por meio da decomposição de Cholesky. Desse modo, a função impulso-resposta para k períodos pode ser representada por:

$$r_a = \sum_{i=1}^m a_i r_i(k) \quad (2.8)$$

Em que $r_i(K)$ é o vetor de resposta no horizonte k ao i -ésimo choque em uma decomposição de Cholesky de $\hat{\Sigma}$. Assim, é possível identificar o vetor de impulso correspondente ao choque investigado, no caso desse trabalho, de incerteza.

A utilização desse método de restrições de sinais tem como benefício o fato de que não é necessário a imposição de restrições sobre toda a matriz A da Equação 2.5. Além disso, a utilização desse método tem como propósito identificar o choque estrutural, ao impor restrição ao sinal da resposta impulsiva de algumas variáveis, sendo possível manter o sinal irrestrito nas variáveis de interesse com o objetivo de observar a sua resposta ao impulso.

A análise de impulso-resposta, ferramenta que proporciona desvendar as relações entre as variáveis de um modelo SVAR, deste trabalho irá investigar o impacto de choques de incerteza – medido pelo índice de incerteza proposto – na economia brasileira, mais especificamente na taxa básica de juros da economia, na inflação, no saldo de empregos e na atividade econômica do país.

2.4.1 Descrição dos dados

Nesta subseção são apresentadas as 5 variáveis macroeconômicas que irão receber o impacto dos choques no indicador de incerteza proposto. O período de coleta dos dados vai de janeiro de 2006, primeiro mês em que é possível se obter os resultados de pesquisas pela plataforma Google *Trends*, até dezembro de 2019, restrito pelo tamanho da série

relativa à emprego. A Tabela 2.2 informa quais as variáveis foram utilizadas, além de suas fontes, unidades e algumas estatísticas descritivas.

Tabela 2.2 – Descrição dos dados utilizados no modelo principal

Série	Acrônimo	Fonte	Unidade de medida
Indicador de incerteza	GTU-BR	Elaboração própria	Índice
Variância de fechamento do Ibovespa	Ibov	B3	Índice
Índice Nacional de Preços ao Consumidor Amplo	IPCA	IBGE	Índice
Saldo de empregados	Empregados	CAGED	Pessoas
Taxa básica de juros	Selic	BACEN	Índice
Índice de Atividade Econômica do Banco Central	IBC-Br	BACEN	Índice

Fonte: Elaboração própria

As séries referentes ao saldo de empregados e o IPCA acumulado em 12 meses foram coletadas através da plataforma do Instituto de Pesquisa Econômica Aplicada – IPEA, o IPEADATA. Já os dados referentes à Atividade Econômica (IBC-Br) e à taxa Selic foram extraídas do Sistema Gerenciador de Séries Temporais (SGS) do Banco Central do Brasil. Por fim, o índice Ibovespa, principal indicador de desempenho das ações negociadas na B3, reunindo as empresas mais importantes do mercado de capitais brasileiro e correspondendo a cerca de 80% do número de negócios, foi obtido no website da B3⁷. A Figura C.1 (Apêndice C) apresenta a evolução temporal das séries de dados utilizadas nesta pesquisa.

A taxa de juros utilizada na pesquisa foi a Over/Selic acumulado no mês divulgada em % a.a. Esta variável representa a média apurada do Sistema Especial de Liquidação e Custódia (Selic), que incide sobre operações de financiamento por um dia útil (overnight), lastreados por títulos públicos registrados no Sistema, servindo como um indexador para os investimentos. Desta forma, um aumento dessa taxa representa um encarecimento do crédito e, conseqüentemente, redução do consumo e da produção, além de maiores retornos para incentivar os investimentos.

O saldo de empregados refere-se a diferença entre o total de admissões e dispensa de empregados que estão sob o regime da Consolidação das Leis trabalhistas – CLT. O IBC-Br tem como objetivo mensurar a evolução contemporânea da atividade econômica do país, incorporando variáveis consideradas como proxies para desempenho dos setores da economia.

A Tabela C.1 (Apêndice C) apresenta a análise de correlação entre as séries descritas na Tabela 2.2. A partir desta é possível identificar que a incerteza possui alta correlação negativa com a produção industrial e com o saldo de empregados e alta correlação positiva com a inflação, conferindo um caráter contracíclico à incerteza, resultado que está em linha com o encontrado para outros países em Bloom (2009) e para o Brasil em análise feita com outros indicadores de incerteza em Filho (2014).

⁷ Disponível em <https://www.b3.com.br/pt_br/market-data-e-indices/indices/indices-amplos/indice-ibovespa-ibovespa-estatisticas-historicas.htm>

Todas as séries possuem periodicidade mensal. As séries referentes ao GTU-BR, índice de atividade econômica e índice Ibovespa foram logaritimizadas. Por apresentarem comportamento sazonal, as variáveis referentes à taxa Selic, ao Ibovespa, atividade econômica, empregados e inflação foram ajustadas sazonalmente, utilizando o método X-13 ARIMA. Com o intuito de testar a presença de raiz unitária na base de dados, foram realizados alguns dos testes de raiz unitária conhecidos na literatura como o Dickey e Fuller (1979), e Kwiatkowski et al. (1992), ADF e KPSS, respectivamente.

A partir do resultado dos testes de raiz unitária (Tabela C.3 no Apêndice C), não foi possível rejeitar a hipótese de existência de raiz unitária. Para tanto, optou-se por eliminar os termos determinísticos destas variáveis, através do filtro Hodrick-Prescott (HP), transformando-as em desvios em relação à tendência, ou seja, a parte cíclica das séries, assim como utilizado em Bloom (2009). Por se tratar de dados de frequência mensal, o parâmetro lambda utilizado no filtro HP foi o valor padrão para séries mensais, assim: $\lambda = 129.600$.

Para os testes de robustez do modelo base, serão realizadas alterações nas medidas de incerteza, na variável que mede a atividade econômica e no período de duração das restrições de sinais impostas. Para medida de incerteza, serão considerados dois dos indicadores de incerteza discutidos no cálculo de correlação na Seção 2.3.1, são eles, o IIE-Br, medido pelo IBRE/FGV e o índice EPU, que serão utilizados no lugar do índice de incerteza proposto neste trabalho.

Além desses, será criada uma variável *dummy* do índice de incerteza GTU-BR, em que o valor 1 é atribuído quando o valor do índice de incerteza ultrapassa 1,65 desvios-padrões acima da variável de tendência do índice obtida pela aplicação do filtro HP. Essa forma de introduzir a incerteza tem como objetivo garantir que a identificação venha apenas de aumentos significativos de incerteza, e não de flutuações menores em andamento⁸.

Outros indicadores utilizados, agora no lugar do índice de atividade econômica, serão variáveis referente à produção industrial da indústria geral, medida pela Pesquisa Industrial Mensal – Produção Física (PIM-PF/IBGE) e às vendas reais no varejo ampliado, medida através da Pesquisa Mensal do Comércio (PMC/IBGE), que capta as atividades de revenda, venda sem transformação significativa, de bens de consumo novos e usados, preponderantemente para o consumidor final, representando, portanto, o último elo da cadeia de distribuição.

Para tanto, a estratégia de identificação de sinais apresentado na Tabela 2.3 foi extraída a partir da análise da teoria econômica sobre choques de incerteza e seus impactos na Seção 2.2.2.

⁸ No caso da variável *dummy*, 8 observações receberam o valor um, pontos que estão destacados na Figura 2.1 que apresenta a série histórica do índice GTU-BR.

Tabela 2.3 – Restrições de sinais impostas nas respostas estruturais

Variável	Choque de Incerteza
Incerteza	+
Ibovespa	-
Inflação	Livre
Juros	-
Saldo de empregados	-
Atividade econômica	Livre

Fonte: Elaboração própria

2.5 RESULTADOS

Esta seção apresenta os resultados com as análises sobre as funções impulso-resposta produzidas em cada caso de estudo. Primeiro serão discutidas as impulso-resposta referentes ao modelo base da pesquisa, que utiliza um SVAR com restrições de sinais para alguns dados macroeconômicos brasileiros.

No tópico seguinte, serão analisados os testes de robustez ao modelo base, em que serão testados outros índices de incerteza, bem como outras variáveis referentes a atividade econômica.

Em ambos os casos, vale destacar que as bandas de erro dos gráficos são de 68% em torno da mediana (representada pela linha sólida), ou seja, a linha tracejada inferior representa o 16º quantil enquanto a linha tracejada superior representa o 84º quantil da distribuição utilizada. Esta prática consiste com a utilizada em Uhlig (2005) - em uma distribuição normal esses quantis representariam uma banda de um desvio padrão. Os resultados das impulso-resposta são considerados até 25º mês após a incidência do choque de incerteza.

Assim como identificada na Tabela 2.3, as variáveis Ibovespa, saldo de empregados e juros foram restritos para serem não positivas e o índice GTU-Br (variável que sofrerá o choque) é restrita para ser não negativa. Além dessas, inflação e atividade econômica foram deixadas livre por haver evidências teóricas que indicam trajetórias distintas após um choque de incerteza. As imposições de sinais serão válidas nos 6 primeiros meses após a incidência do choque.

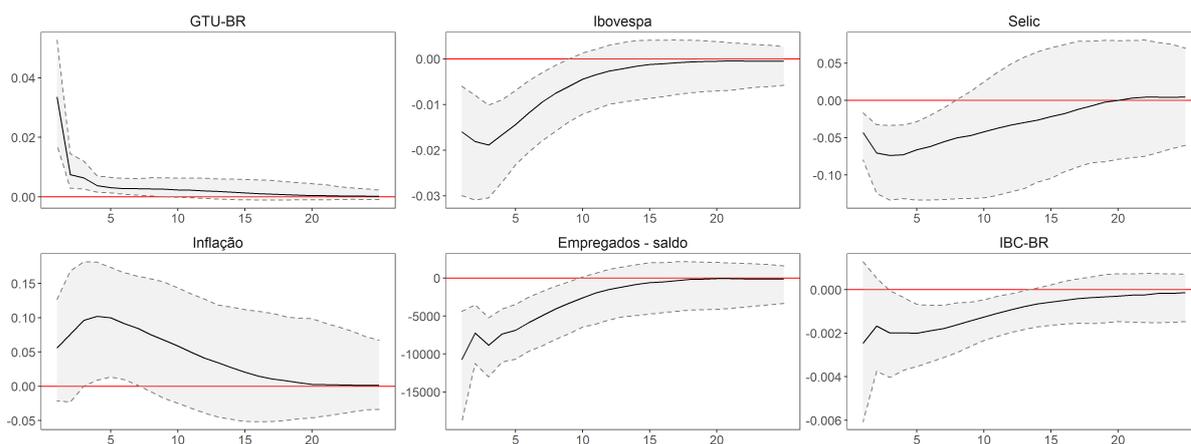
No geral, os resultados encontrados estão de acordo com os mecanismos de transmissão que identificam um efeito contracionista em períodos de aumento de incerteza, sendo observado um aumento da inflação e do desemprego e retração da produção do país. Resultado em linha com os observados tanto internacionalmente - como por exemplo em Bloom (2014), Caggiano, Castelnuovo e Groshenny (2014), Baker, Bloom e Davis (2016) e Moore (2017) - tanto na literatura nacional, como apresentado em Filho (2014), Ferreira et al. (2017) e Barboza e Zilberman (2018).

A retração da atividade econômica no curto prazo se dá por meio da redução dos investimentos e contratação de trabalhadores por partes das empresas, e do consumo das famílias, que alocam uma maior quantidade de dinheiro em uma poupança precaucional, resultando em um impacto negativo sobre o crescimento econômico.

2.5.1 Resultados do modelo principal

No modelo principal, além da variável relacionada à incerteza proposta e desenvolvida na Seção 2.3, serão utilizadas as variáveis relacionadas ao Ibovespa, ao IPCA, à taxa de juros Selic, ao saldo de empregados do período e à atividade econômica, medida pelo IBC-Br. As funções impulso-resposta ao choque de incerteza (GTU-BR) sobre estas variáveis do modelo principal estão representadas na Figura 2.2. Os resultados são descritos a seguir.

Figura 2.2 – Função impulso-resposta do modelo principal



Fonte: Elaboração própria

É possível identificar que, inicialmente, o choque de incerteza eleva o índice GTU-Br, mantendo uma trajetória positiva em relação à tendência até pelo menos o décimo mês após o choque, só após esse período que o índice reverte seu curso, voltando a trajetória da tendência. Logo após o aumento da incerteza, o Ibovespa apresenta uma redução de 0,015%, essa trajetória negativa em relação a tendência é mantida por 16 meses, tendo o período de maior retração por volta do segundo e terceiro mês.

Esse movimento do mercado de ações é explicado pelo fato de que investidores racionais e que são avessos ao risco estabelecem preços menores dos ativos, significativamente abaixo do valor esperado, uma vez que houve aumento do risco associado a esses ativos, sendo necessário um maior retorno para que haja o investimento por parte do investidor, assim como observado em Pástor e Veronesi (2013) e Caldara et al. (2016).

Em relação à taxa de juros, à inflação e ao saldo de empregados, o choque de incerteza possui um efeito contracionista no número de empregados da economia, além de

provocar uma política monetária que estimule mais a economia. Além disso, é possível observar um crescimento da inflação, tendo uma trajetória positiva em relação à tendência por 19 meses após o choque de incerteza.

Conforme discutido anteriormente, esse resultado observado pode ser explicado por dois mecanismos de transmissão, o opções reais, que destaca que a incerteza afeta negativamente os investimentos e a contratação de funcionários e tem como consequência um efeito contracionista sobre a atividade econômica, assim como observado em Bernanke (1983), Bloom (2009) e Caggiano, Castelnuovo e Groshenny (2014). Além deste mecanismo, o canal de poupanças precaucionais também aborda o efeito contracionista de um choque de incerteza sobre a atividade econômica no curto prazo, uma vez que os consumidores aumentam sua poupança preventiva e reduzem seu consumo, contraindo a atividade econômica no curto prazo, porém esta relação de mais poupança e menos consumo pode, no longo prazo, beneficiar o crescimento econômico.

Os resultados encontrados estão em linha com os observados em Filho (2014), Barboza e Zilberman (2018) e Duarte (2021), que identificam um efeito contracionista de um choque de incerteza no Brasil. Na literatura internacional, os resultados obtidos estão em linha com Bloom (2009) para os Estados Unidos, Castelnuovo e Tran (2017) para a Austrália e Bilgin et al. (2019) para a Turquia.

2.5.2 Teste de robustez

Esta seção tem como objetivo avaliar a sensibilidade dos resultados que foram encontrados na Seção 2.5.1, assim sendo, o modelo foi estimado mais oito vezes - três vezes modificando os índices de incerteza, duas vezes modificando o índice de atividade econômica, assim como descrito na Seção 2.4.1 e mais três estimações diferenciando o tempo de restrição dos sinais do modelo (k) -, sendo consideradas uma de cada vez, por haver evidências de que afetam a atividade econômica no Brasil e devido ao fato de que modelos VAR sofrem da maldição da dimensionalidade, impedindo que sejam todas consideradas juntas no modelo básico. Todas as impulsos-respostas referentes ao teste de robustez estão apresentadas no Apêndice D.

2.5.2.1 Diferentes medidas para a incerteza

Inicialmente, duas medidas alternativas são utilizadas para medir a incerteza. A primeira medida de incerteza utilizada é o EPU, indicador que utiliza a metodologia desenvolvida por Baker, Bloom e Davis (2016), construído a partir de textos do Jornal Folha de São Paulo e que possuem determinados termos. A segunda medida utilizada nesta seção é o IIE-BR, indicador desenvolvido em Ferreira et al. (2017) e que é composto por dois outros indicadores que refletem a incerteza incidente em artigos publicados em seis dos principais jornais do país e que refletem a dispersão das previsões de especialistas

para a taxa de câmbio e Selic, 12 meses à frente, e o IPCA acumulado para os próximos 12 meses que são divulgados pelo Banco Central.

Ao analisar as Figuras D.2 e D.1, nota-se que todas as impulso-resposta seguem o padrão apresentado na Figura 2.2 que representa o modelo principal. Estes resultados fortalecem o ponto de vista que o mecanismo de transmissão de um choque de incerteza na economia tem caráter contracionista, assim como abordado pelo mecanismo de opções reais. Além desse resultado, pode-se identificar que a reação do Ibovespa aos choques desses dois indicadores de incerteza alternativos possui uma magnitude maior do que ao indicador GTU-Br, entretanto, o choque de incerteza do modelo principal possui um elevado grau de persistência sobre o Ibovespa.

Sobre a taxa de juros, identifica-se maior persistência além de uma proporção maior da política monetária expansionista. Em relação a inflação, detecta-se que a reação do nível geral de preços ao choque do GTU-Br é maior do que a reação observado nos índices alternativos, desta forma, aumentos de incerteza detectados por esse índice apresentam maiores índices de preços no curto prazo.

Para o trabalho, nota-se menor persistência do choque sobre o saldo de empregados. Reação semelhante a observada no índice IBC-Br, entretanto, a retração da economia é maior nos modelos que consideram os índices alternativos.

Além destes indicadores alternativos, utilizou-se também como indicador de incerteza uma variável *dummy* representando os períodos em que o valor do índice GTU-Br ultrapassa 1,65 desvio-padrões acima da variável de tendência desse índice, que é obtida pela aplicação do filtro HP e tem como finalidade identificar apenas o efeito de aumentos mais inesperados e significativos da incerteza. Os resultados referentes são apresentados na Figura D.3. Ibovespa e Selic apresentam retrações maiores do que as observadas no modelo principal, sendo maior a persistência do choque sobre a Selic nesse teste de robustez. Diferentemente do modelo principal, a atividade econômica neste teste apresenta um leve crescimento nos primeiros meses após o choque, entretanto esse movimento é logo revertido nos meses seguintes.

2.5.2.2 Diferentes medidas para a atividade econômica

Como forma de observar a sensibilidade do modelo principal a outras medidas de atividade econômica, foram estimados novos modelos SVAR utilizando à produção industrial da indústria geral, através da PIM-PF, e as vendas reais do varejo ampliado, medidas pela PMC. De modo geral, nos dois novos casos analisados, não houve mudanças nas conclusões do impacto do choque de incerteza sobre as variáveis relativas à inflação e atividade econômica.

A produção industrial apresenta a maior retração de seu índice após o choque de

incerteza, com a resposta máxima acontecendo no terceiro e quarto mês, a partir do qual a resposta inflexiona e converge para zero. A dinâmica do varejo é similar, porém apresenta uma resposta menor e que converge de forma mais rápida para a tendência do índice.

No índice IBC-Br do modelo principal, que considera os diversos setores da economia, o efeito é menor do que os observados nas vendas reais do varejo e na produção da indústria geral. Isso pode indicar que setores não observados nesta análise devem ter reações ao choque de incerteza mais amenas (ou até não ter reações negativas) do que as observadas no Varejo e na produção industrial.

2.5.2.3 Choque de incerteza com diferentes durações

Esta última seção do teste de robustez tem como objetivo verificar a sensibilidade do modelo a mudanças na duração das restrições de sinais impostas. Para tanto, além da restrição inicial do modelo principal ($k = 6$), foram testados modelos menos e mais restritos do que o modelo *benchmark*, assim as restrições impostas foram de 1, 3 e 9 meses. Os resultados obtidos estão apresentados na Figura D.6 no Apêndice D.

De modo geral, as dinâmicas sobre as variáveis são parecidas. Ou seja, em virtude do choque de incerteza seguiu-se uma contração do mercado acionário brasileiro, acompanhado por um relaxamento da política monetária e um aumento da inflação como consequência do aumento da incerteza. Além disso, constata-se o aumento do desemprego e a retração da atividade econômica brasileira, seguindo o mecanismo de opções reais já destacado anteriormente.

Porém, é possível identificar uma contração menos significativa da taxa de juros ao longo do tempo, de forma que, na medida em que a incerteza se torna mais duradoura, os preços tendem a aumentar mesmo com juros que sejam menos estimulantes. Este resultado está de acordo com o encontrado em Duarte (2021), que explica que algumas possibilidades para esse efeito altista nos preços pode ser a presença de estímulos fiscais realizados pelo Governo ou por choques de preços de *commodities*, por exemplo. Para o índice IBC-Br, detecta-se que quanto maior a duração da incerteza, mais significativa e persistente será o impacto.

2.6 CONSIDERAÇÕES FINAIS

Este trabalho buscou observar o impacto de um aumento da incerteza sobre variáveis macroeconômicas, a partir de um novo índice de incerteza criado para a economia brasileira, captando o sentimento das pessoas diretamente de suas procuras na Internet. De forma que, aumentos inesperados na busca por palavras específicas podem indicar um aumento de incerteza no Brasil.

Os resultados encontrados ajudam a explicar o desempenho negativo observado na economia brasileira em períodos de aumento da incerteza nos períodos mais recentes provenientes por diversos fatores, entre eles políticos, conflitos externos e desastres naturais. Para gerar esses resultados foi utilizado um modelo de vetor autorregressivo estrutural (SVAR) com restrições de sinais, com a decomposição de Cholesky para identificar os choques de incerteza.

Esses resultados podem indicar a necessidade de maior diversificação da economia, sendo menos exposto às variações de um pequeno número de produtos, bem como políticas de estabilização fiscal e monetária mais eficazes, uma vez que Bloom (2014) constata que países em desenvolvimento, como o Brasil, tendem a ter um maior nível de incerteza quando comparados com países de alta renda.

Cabe ressaltar que as principais contribuições deste trabalho podem ser resumidas em dois importantes tópicos que são complementares. O primeiro objeto construído na pesquisa foi um índice de incerteza, chamado de GTU-BR, que utiliza um conjunto de dados de pesquisas efetuadas na Internet e que se correlacionam com eventos específicos e são correlacionados com outros índices de incerteza utilizados com frequência na literatura, para o período de janeiro de 2006 até dezembro de 2019, utilizando palavras-chave específicas para o Brasil.

O segundo aspecto desenvolvido neste trabalho foi a investigação do impacto de um aumento da incerteza em variáveis macroeconômicas nacionais. A pesquisa mostrou que o aumento da incerteza afeta negativamente a atividade econômica nacional e o mercado acionário, além do aumento do desemprego e da inflação, resultado compatível com os mecanismos de opções reais e poupanças precaucionárias.

Em suma, esses tópicos podem gerar efeitos positivos ao contribuir para a literatura nacional com um indicador de incerteza que possa auxiliar na tomada de decisão dos agentes econômicos principalmente em períodos de aumento da incerteza. Este trabalho pode, ainda, ser estendido em diversos aspectos. Primeiro, choques de incerteza podem ser originados por razões diversas, assim, sugere-se investigar se choques de incerteza originados por razões diferentes possuem impactos diferentes sobre as variáveis macroeconômicas, utilizando choques de incerteza externos ou choques de incerteza gerados internamente, além de diferenciar, se possível, incerteza proveniente de motivos fiscais e incerteza proveniente de

motivos políticos, por exemplo.

REFERÊNCIAS

- ABEL, A. B. Optimal investment under uncertainty. *The American Economic Review*, JSTOR, v. 73, n. 1, p. 228–233, 1983.
- AHIR, H.; BLOOM, N.; FURCERI, D. *The world uncertainty index*. [S.l.], 2022.
- AHMED, N. K. et al. An empirical comparison of machine learning models for time series forecasting. *Econometric reviews*, Taylor & Francis, v. 29, n. 5-6, p. 594–621, 2010.
- AIOLFI, M.; TIMMERMANN, A. Persistence in forecasting performance and conditional combination strategies. *Journal of Econometrics*, Elsevier, v. 135, n. 1-2, p. 31–53, 2006.
- ASKITAS, N.; ZIMMERMANN, K. F. Google econometrics and unemployment forecasting. *Applied Economics Quarterly*, DUNCKER UND HUMBLOT, v. 55, n. 2, p. 107–120, 2009.
- BACHMANN, R.; ELSTNER, S.; SIMS, E. R. Uncertainty and economic activity: Evidence from business survey data. *American Economic Journal: Macroeconomics*, v. 5, n. 2, p. 217–49, 2013.
- BAKER, S. R.; BLOOM, N. Does uncertainty reduce growth? using disasters as natural experiments. 2013.
- BAKER, S. R.; BLOOM, N.; DAVIS, S. J. Measuring economic policy uncertainty. *The quarterly journal of economics*, Oxford University Press, v. 131, n. 4, p. 1593–1636, 2016.
- BANERJEE, R.; KEARNS, J.; LOMBARDI, M. J. (why) is investment weak? *BIS Quarterly Review March*, 2015.
- BANSAL, R.; YARON, A. Risks for the long run: A potential resolution of asset pricing puzzles. *The journal of Finance*, Wiley Online Library, v. 59, n. 4, p. 1481–1509, 2004.
- BAR-ILAN, A.; STRANGE, W. C. Investment lags. *The American Economic Review*, JSTOR, v. 86, n. 3, p. 610–622, 1996.
- BARAK, S.; SADEGH, S. S. Forecasting energy consumption using ensemble arima–anfis hybrid algorithm. *International Journal of Electrical Power & Energy Systems*, Elsevier, v. 82, p. 92–104, 2016.
- BARBOSA, R. B.; FERREIRA, R. T.; SILVA, T. M. d. Previsão de variáveis macroeconômicas brasileiras usando modelos de séries temporais de alta dimensão. *Estudos Econômicos (São Paulo)*, SciELO Brasil, v. 50, p. 67–98, 2020.
- BARBOZA, R. d. M.; ZILBERMAN, E. Os efeitos da incerteza sobre a atividade econômica no brasil. *Revista Brasileira de Economia*, SciELO Brasil, v. 72, p. 144–160, 2018.
- BEKAERT, G.; HOEROVA, M.; DUCA, M. L. Risk, uncertainty and monetary policy. *Journal of Monetary Economics*, Elsevier, v. 60, n. 7, p. 771–788, 2013.

- BERGE, T. J. Predicting recessions with leading indicators: Model averaging and selection over the business cycle. *Journal of Forecasting*, Wiley Online Library, v. 34, n. 6, p. 455–471, 2015.
- BERNANKE, B. S. Irreversibility, uncertainty, and cyclical investment. *The quarterly journal of economics*, MIT Press, v. 98, n. 1, p. 85–106, 1983.
- BESARRIA, C. C. da N. et al. Incerteza macroeconômica e seus efeitos fiscais: Uma análise a partir de processamento de linguagem natural e modelos dinâmicos estocásticos de equilíbrio geral (dsge). *CADERNOS DE FINANÇAS PÚBLICAS*, v. 21, n. 1, 2021.
- BILGIN, M. H. et al. A novel index of macroeconomic uncertainty for turkey based on google-trends. *Economics Letters*, Elsevier, v. 184, p. 108601, 2019.
- BLOOM, N. The impact of uncertainty shocks. *econometrica*, Wiley Online Library, v. 77, n. 3, p. 623–685, 2009.
- BLOOM, N. Fluctuations in uncertainty. *Journal of Economic Perspectives*, v. 28, n. 2, p. 153–76, 2014.
- BOHN, L.; BUENO, N. P. Um procedimento para prever recessões no brasil a partir de indicadores antecedentes. *Estudos Econômicos (São Paulo)*, SciELO Brasil, v. 45, p. 215–247, 2015.
- BONTEMPI, M. E. et al. Eurq: A new web search-based uncertainty index. *Economica*, Wiley Online Library, v. 88, n. 352, p. 969–1015, 2021.
- BONTEMPI, M. E.; GOLINELLI, R.; SQUADRANI, M. A new index of uncertainty based on internet searches: A friend or foe of other indicators? Quaderni-Working Paper DSE, 2016.
- BORN, B.; PFEIFER, J. Policy risk and the business cycle. *Journal of Monetary Economics*, Elsevier, v. 68, p. 68–85, 2014.
- BOX, G. E.; JENKINS, G. M. Time series analysis. forecasting and control. *Holden-Day Series in Time Series Analysis*, 1970.
- BRETSCHNEIDER, S. I. et al. Political and organizational influences on the accuracy of forecasting state government revenues. *International Journal of Forecasting*, Elsevier, v. 5, n. 3, p. 307–319, 1989.
- BROWN, K. C.; HARLOW, W. V.; TINIC, S. M. Risk aversion, uncertain information, and market efficiency. *Journal of financial economics*, Elsevier, v. 22, n. 2, p. 355–385, 1988.
- CAGGIANO, G.; CASTELNUOVO, E.; GROSHENNY, N. Uncertainty shocks and unemployment dynamics in us recessions. *Journal of Monetary Economics*, Elsevier, v. 67, p. 78–92, 2014.
- CAGGIANO, G.; CASTELNUOVO, E.; NODARI, G. Uncertainty and monetary policy in good and bad times. Melbourne Institute Working Paper, 2017.
- CAGLIARINI, A.; HEATH, A. et al. The optimal control problem with bewley preferences| rdp 2000-10: Monetary policy-making in the presence of knightian uncertainty. *Reserve Bank of Australia Research Discussion Papers*, n. December, 2000.

- CALDARA, D. et al. The macroeconomic impact of financial and uncertainty shocks. *European Economic Review*, Elsevier, v. 88, p. 185–207, 2016.
- CAMPOS, C. Métodos econométricos aplicados à previsão da arrecadação de receitas federais. *Receita Federal do Brasil. ESAF*, 2009.
- CARLITZ, R. Improving transparency and accountability in the budget process: An assessment of recent initiatives. *Development Policy Review*, Wiley Online Library, v. 31, p. s49–s67, 2013.
- CASTELNUOVO, E. Yield curve and financial uncertainty: Evidence based on us data. *Australian Economic Review*, Wiley Online Library, v. 52, n. 3, p. 323–335, 2019.
- CASTELNUOVO, E.; TRAN, T. D. Google it up! a google trends-based uncertainty index for the united states and australia. *Economics Letters*, Elsevier, v. 161, p. 149–153, 2017.
- CEPNI, O.; GÜNEY, I. E.; SWANSON, N. R. Nowcasting and forecasting gdp in emerging markets using global financial and macroeconomic diffusion indexes. *International Journal of Forecasting*, Elsevier, v. 35, n. 2, p. 555–572, 2019.
- CHATFIELD, C. What is the ‘best’ method of forecasting? *Journal of Applied Statistics*, Taylor & Francis, v. 15, n. 1, p. 19–38, 1988.
- CHATFIELD, C. *Time-series forecasting*. [S.l.]: Chapman and Hall/CRC, 2000.
- CHEN, T.; GUESTRIN, C. Xgboost: A scalable tree boosting system. In: *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*. [S.l.: s.n.], 2016. p. 785–794.
- CHOI, H.; VARIAN, H. Predicting the present with google trends. *Economic record*, Wiley Online Library, v. 88, p. 2–9, 2012.
- COTA, R. T. C. Uma imersão nas contas públicas brasileiras. Instituto de Pesquisa Econômica Aplicada (Ipea), 2017.
- DICKEY, D. A.; FULLER, W. A. Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American statistical association*, Taylor & Francis, v. 74, n. 366a, p. 427–431, 1979.
- DICKEY, D. A.; FULLER, W. A. Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica: journal of the Econometric Society*, JSTOR, p. 1057–1072, 1981.
- DOU, W. W. et al. Macroeconomic models for monetary policy: A critical review from a finance perspective. *Annual Review of Financial Economics*, Annual Reviews, v. 12, p. 95–140, 2020.
- DUARTE, C. A. d. P. Os efeitos de choques de incerteza na economia brasileira: uma abordagem var com sign restrictions. 2021.
- EBEKE, M. C.; ÖLCER, M. *Fiscal policy over the election cycle in low-income countries*. [S.l.]: International Monetary Fund, 2013.
- ELLIOTT, G.; TIMMERMANN, A. Economic forecasting. *Journal of Economic Literature*, v. 46, n. 1, p. 3–56, 2008.

- ENGLE, R. F. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica: Journal of the econometric society*, JSTOR, p. 987–1007, 1982.
- FASANI, S.; ROSSI, L. Are uncertainty shocks aggregate demand shocks? *Economics Letters*, Elsevier, v. 167, p. 142–146, 2018.
- FERREIRA, P. C. et al. Medindo a incerteza econômica no brasil. *Economia Aplicada*, 2017.
- FILDES, R.; STEKLER, H. The state of macroeconomic forecasting. *Journal of macroeconomics*, Elsevier, v. 24, n. 4, p. 435–468, 2002.
- FILHO, A. E. d. C. Incerteza e atividade econômica no brasil. *Economia Aplicada*, SciELO Brasil, v. 18, p. 421–453, 2014.
- GARCIA, M. G.; MEDEIROS, M. C.; VASCONCELOS, G. F. Real-time inflation forecasting with high-dimensional models: The case of brazil. *International Journal of Forecasting*, Elsevier, v. 33, n. 3, p. 679–693, 2017.
- GIAMBIAGI, F.; ALEM, A.; PINTO, S. G. B. *Finanças públicas - Teoria e Prática no Brasil*. 5. ed. [S.l.]: Elsevier Brasil, 2015.
- GILCHRIST, S.; SIM, J. W.; ZAKRAJŠEK, E. *Uncertainty, financial frictions, and investment dynamics*. [S.l.], 2014.
- 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.
- GOOIJER, J. G. D.; HYNDMAN, R. J. 25 years of time series forecasting. *International journal of forecasting*, Elsevier, v. 22, n. 3, p. 443–473, 2006.
- GOOIJER, J. G. D.; KUMAR, K. Some recent developments in non-linear time series modelling, testing, and forecasting. *International Journal of Forecasting*, Elsevier, v. 8, n. 2, p. 135–156, 1992.
- GRAMINHO, F. M. et al. Sentimento e macroeconomia: uma análise dos índices de confiança no brasil. *Brazilian Central Bank, trabalhos para discussão*, v. 408, 2015.
- HARTMAN, R. The effects of price and cost uncertainty on investment. *Journal of economic theory*, Elsevier, v. 5, n. 2, p. 258–266, 1972.
- HIBON, M.; EVGENIOU, T. To combine or not to combine: selecting among forecasts and their combinations. *International journal of forecasting*, Elsevier, v. 21, n. 1, p. 15–24, 2005.
- HOLLAUER, G.; ISSLER, J. V. Construção de indicadores antecedentes para a atividade industrial brasileira e comparação de metodologias. Instituto de Pesquisa Econômica Aplicada (Ipea), 2006.
- HYNDMAN, R. J.; ATHANASOPOULOS, G. *Forecasting: principles and practice*. [S.l.]: OTexts: Melbourne, Australia, 2021. v. 3.

- IFI, I. F. I. *Relatório de Acompanhamento Fiscal (RAF)*. [S.l.]: Brasília: Senado Federal, 2021. <https://www2.senado.leg.br/bdsf/bitstream/handle/id/592401/RAF55_AGO2021.pdf>. Acessado em: 20/12/2022.
- JENKINS, G. M. Some practical aspects of forecasting in organizations. *Journal of forecasting*, Wiley Online Library, v. 1, n. 1, p. 3–21, 1982.
- JIN, X. et al. Significance and challenges of big data research. *Big data research*, Elsevier, v. 2, n. 2, p. 59–64, 2015.
- JURADO, K.; LUDVIGSON, S. C.; NG, S. Measuring uncertainty. *American Economic Review*, v. 105, n. 3, p. 1177–1216, 2015.
- KHASHEI, M.; BIJARI, M. A novel hybridization of artificial neural networks and arima models for time series forecasting. *Applied soft computing*, Elsevier, v. 11, n. 2, p. 2664–2675, 2011.
- KIM, S.; KIM, H. A new metric of absolute percentage error for intermittent demand forecasts. *International Journal of Forecasting*, Elsevier, v. 32, n. 3, p. 669–679, 2016.
- KNIGHT, F. H. Risk, uncertainty and profit, hart. *Schaffner & Marx*, Houghton Mifflin Company, 1921.
- KRAFT, H.; SCHWARTZ, E.; WEISS, F. Growth options and firm valuation. *European Financial Management*, Wiley Online Library, v. 24, n. 2, p. 209–238, 2018.
- KWIATKOWSKI, D. et al. Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *Journal of econometrics*, Elsevier, v. 54, n. 1-3, p. 159–178, 1992.
- LEDUC, S.; LIU, Z. Uncertainty shocks are aggregate demand shocks. *Journal of Monetary Economics*, Elsevier, v. 82, p. 20–35, 2016.
- LEVIN, J. D. The economics of internet markets. *National Bureau of Economic Research*, 2011.
- LI, P.; ZHANG, J.-S. A new hybrid method for china's energy supply security forecasting based on arima and xgboost. *Energies*, MDPI, v. 11, n. 7, p. 1687, 2018.
- LUDVIGSON, S. C.; MA, S.; NG, S. *COVID-19 and the macroeconomic effects of costly disasters*. [S.l.], 2020.
- MAHER, C. S.; DELLER, S. C. Measuring municipal fiscal condition: Do objective measures of fiscal health relate to subjective measures? *Journal of Public Budgeting, Accounting & Financial Management*, Emerald Publishing Limited, 2011.
- MAKRIDAKIS, S. Forecasting: its role and value for planning and strategy. *International journal of forecasting*, Elsevier, v. 12, n. 4, p. 513–537, 1996.
- MAKRIDAKIS, S. et al. The accuracy of extrapolation (time series) methods: Results of a forecasting competition. *Journal of forecasting*, Wiley Online Library, v. 1, n. 2, p. 111–153, 1982.

- MARCELLINO, M.; STOCK, J. H.; WATSON, M. W. Macroeconomic forecasting in the euro area: Country specific versus area-wide information. *European Economic Review*, Elsevier, v. 47, n. 1, p. 1–18, 2003.
- MCLAREN, N.; SHANBHOGUE, R. Using internet search data as economic indicators. *Bank of England Quarterly Bulletin*, n. 2011, p. Q2, 2011.
- MCNICHOL, E. C. Improving state revenue forecasting: best practices for a more trusted and reliable revenue estimate. *Center on Budget and Policy Priorities-CBPP*, set, 2014.
- MEINEN, P.; ROEHE, O. To sign or not to sign? on the response of prices to financial and uncertainty shocks. *Economics Letters*, Elsevier, v. 171, p. 189–192, 2018.
- MENDONÇA, M. J. C. d.; MEDRANO, L. A. T. Um modelo de combinação de previsões para arrecadação de receita tributária no brasil. Instituto de Pesquisa Econômica Aplicada (Ipea), 2016.
- Ministério da Fazenda. *Panorama Fiscal Brasileiro*. 2018. <<https://www.gov.br/fazenda/pt-br/centrais-de-conteudos/publicacoes/analises-e-estudos/arquivos/2018/panorama-fiscal-brasileiro.pdf>>. Acessado em: 15/08/2022.
- MOORE, A. Measuring economic uncertainty and its effects. *Economic record*, Wiley Online Library, v. 93, n. 303, p. 550–575, 2017.
- NOPP, C.; HANBURY, A. Detecting risks in the banking system by sentiment analysis. In: *Proceedings of the 2015 conference on empirical methods in natural language processing*. [S.l.: s.n.], 2015. p. 591–600.
- OI, W. Y. The desirability of price instability under perfect competition. *Econometrica: journal of the Econometric Society*, JSTOR, p. 58–64, 1961.
- OLIVEIRA, J. F. D.; LUDERMIR, T. B. A hybrid evolutionary decomposition system for time series forecasting. *Neurocomputing*, Elsevier, v. 180, p. 27–34, 2016.
- PALIT, A. K.; POPOVIC, D. *Computational intelligence in time series forecasting: theory and engineering applications*. [S.l.]: Springer Science & Business Media, 2006.
- PANIGRAHI, S.; BEHERA, H. S. A hybrid ets–ann model for time series forecasting. *Engineering applications of artificial intelligence*, Elsevier, v. 66, p. 49–59, 2017.
- PÁSTOR, L.; VERONESI, P. Political uncertainty and risk premia. *Journal of financial Economics*, Elsevier, v. 110, n. 3, p. 520–545, 2013.
- POSKITT, D. S.; TREMAYNE, A. The selection and use of linear and bilinear time series models. *International Journal of Forecasting*, Elsevier, v. 2, n. 1, p. 101–114, 1986.
- ROGOFF, K.; SIBERT, A. Elections and macroeconomic policy cycles. *The review of economic studies*, Wiley-Blackwell, v. 55, n. 1, p. 1–16, 1988.
- ROMER, C. D. The great crash and the onset of the great depression. *The Quarterly Journal of Economics*, MIT Press, v. 105, n. 3, p. 597–624, 1990.
- SAID, S. E.; DICKEY, D. A. Testing for unit roots in autoregressive-moving average models of unknown order. *Biometrika*, Oxford University Press, v. 71, n. 3, p. 599–607, 1984.

- SCHORFHEIDE, F.; SONG, D. *Real-time forecasting with a (standard) mixed-frequency VAR during a pandemic*. [S.l.], 2021.
- SILL, K. et al. Measuring economic uncertainty using the survey of professional forecasters. *Federal Reserve Bank of Philadelphia Business Review*, v. 92, p. 16–27, 2012.
- SILVA, P. H. N. et al. Mensurando o sentimento de incerteza da política econômica: uma análise a partir da comunicação do banco central do brasil. Universidade Federal da Paraíba, 2020.
- SIMS, C. A. Macroeconomics and reality. *Econometrica: journal of the Econometric Society*, JSTOR, p. 1–48, 1980.
- SIQUEIRA, F. d. F. *Política fiscal e ciclo político no Brasil: uma análise empírica*. Tese (Doutorado) — Universidade de São Paulo, 2015.
- SYNTETOS, A. A.; BOYLAN, J. E.; DISNEY, S. M. Forecasting for inventory planning: a 50-year review. *Journal of the Operational Research Society*, Springer, v. 60, n. 1, p. S149–S160, 2009.
- TASCIKARAOGLU, A.; UZUNOGLU, M. A review of combined approaches for prediction of short-term wind speed and power. *Renewable and Sustainable Energy Reviews*, Elsevier, v. 34, p. 243–254, 2014.
- TAYLOR, S. J.; LETHAM, B. Forecasting at scale. *The American Statistician*, Taylor & Francis, v. 72, n. 1, p. 37–45, 2018.
- TERÄSVIRTA, T.; DIJK, D. V.; MEDEIROS, M. C. Linear models, smooth transition autoregressions, and neural networks for forecasting macroeconomic time series: A re-examination. *International Journal of Forecasting*, Elsevier, v. 21, n. 4, p. 755–774, 2005.
- TERUI, N.; DIJK, H. K. V. Combined forecasts from linear and nonlinear time series models. *International Journal of Forecasting*, Elsevier, v. 18, n. 3, p. 421–438, 2002.
- TESOURO NACIONAL. *Manual de Estatísticas Fiscais do Resultado do Tesouro Nacional*. 2016. <<https://www.tesourotransparente.gov.br/publicacoes/manual-de-estatisticas-fiscais-do-resultado-do-tesouro-nacional/2019/30>>. Acessado em: 21/09/2022.
- TONG, H. *Threshold Models in Non-linear Time Series Analysis*. New York, NY: Springer New York, 1983. v. 21. (Lecture Notes in Statistics, v. 21). ISBN 9780387909189 9781468478884. Disponível em: <<http://link.springer.com/10.1007/978-1-4684-7888-4>>.
- TONG, H. *Non-linear time series: a dynamical system approach*. Reprinted. Oxford: Clarendon Press, 1990. (Oxford statistical science series, 6). ISBN 9780198523000.
- UHLIG, H. What are the effects of monetary policy on output? results from an agnostic identification procedure. *Journal of Monetary Economics*, Elsevier, v. 52, n. 2, p. 381–419, 2005.
- VECCHIA, E. D.; MONTTOYA, M. A. Orçamento e planejamento municipal: um estudo de caso. *Brazilian Journal of Political Economy*, SciELO Brasil, v. 22, p. 334–354, 2020.

VIGNOLI, F. H.; FUNCIA, F. R. *Planejamento e orçamento público*. [S.l.]: Editora FGV, 2015.

WATKINS, G. P. Knight's risk, uncertainty and profit. *The Quarterly Journal of Economics*, MIT Press, v. 36, n. 4, p. 682–690, 1922.

WOLOSZKO, N. Tracking activity in real time with google trends. OECD, 2020.

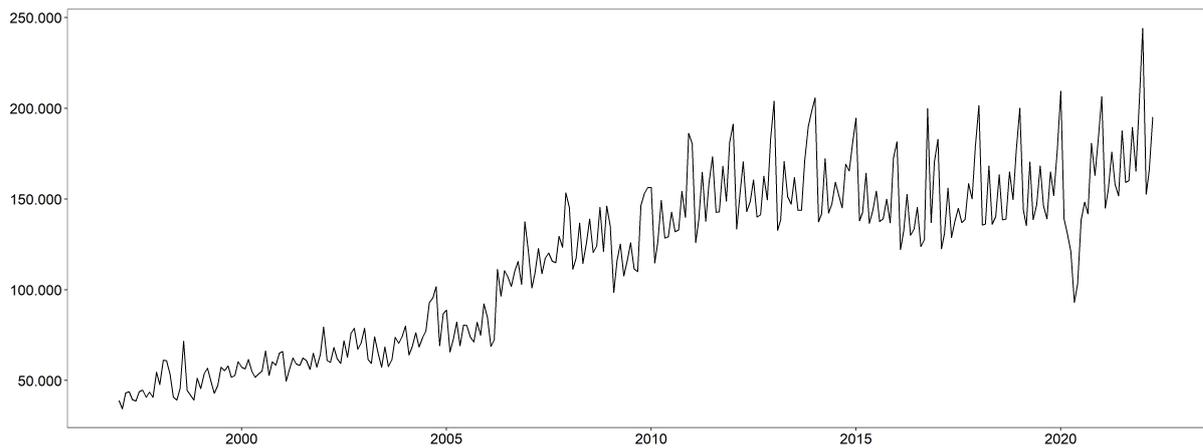
ZHANG, G. P. Time series forecasting using a hybrid arima and neural network model. *Neurocomputing*, Elsevier, v. 50, p. 159–175, 2003.

ZHOU, Z.-H. *Ensemble methods: foundations and algorithms*. [S.l.]: CRC press, 2012.

Apêndices

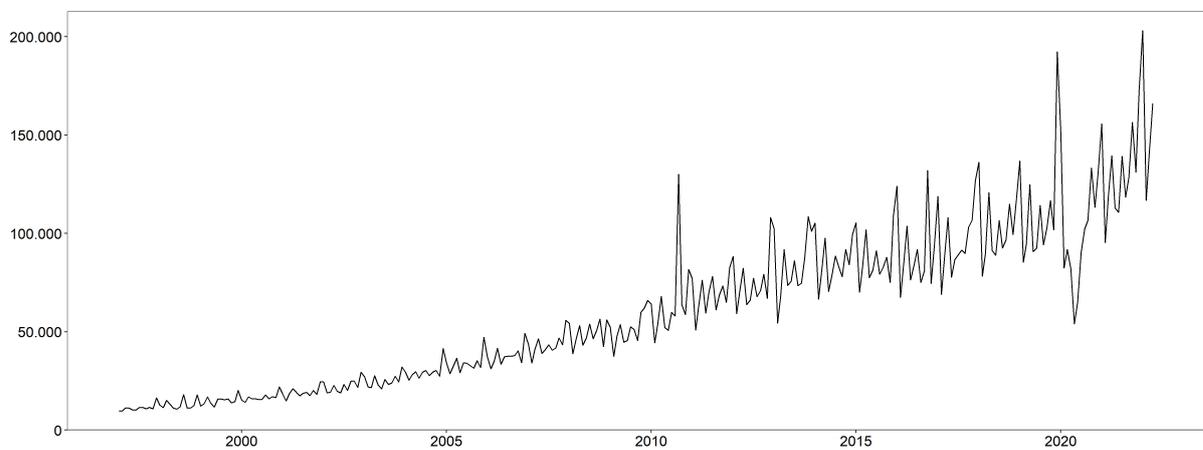
APÊNDICE A – SÉRIE TEMPORAL DAS VARIÁVEIS PREVISTAS

Figura A.1 – Arrecadação das Receitas Federais - em R\$ milhões



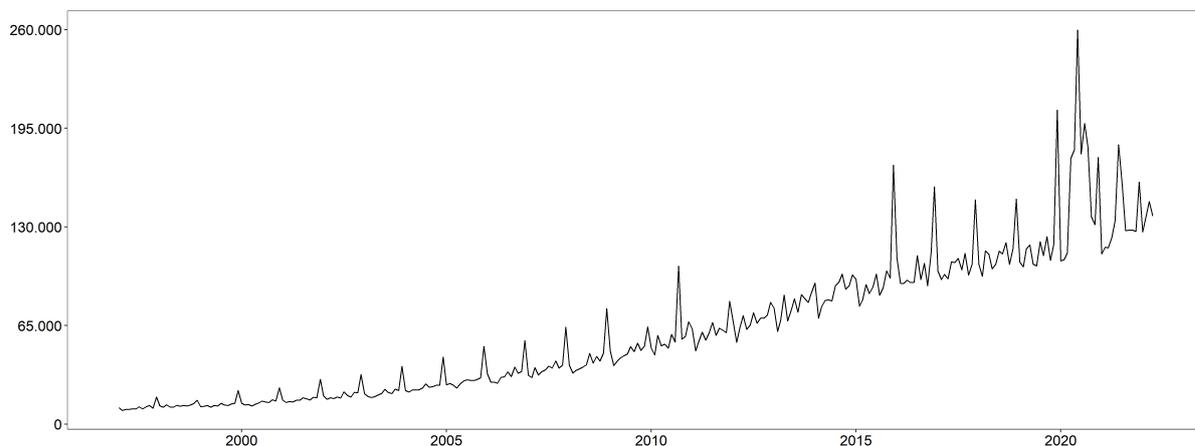
Fonte: Receita Federal do Brasil. Elaboração Própria

Figura A.2 – Receita Líquida do Governo Central - em R\$ milhões



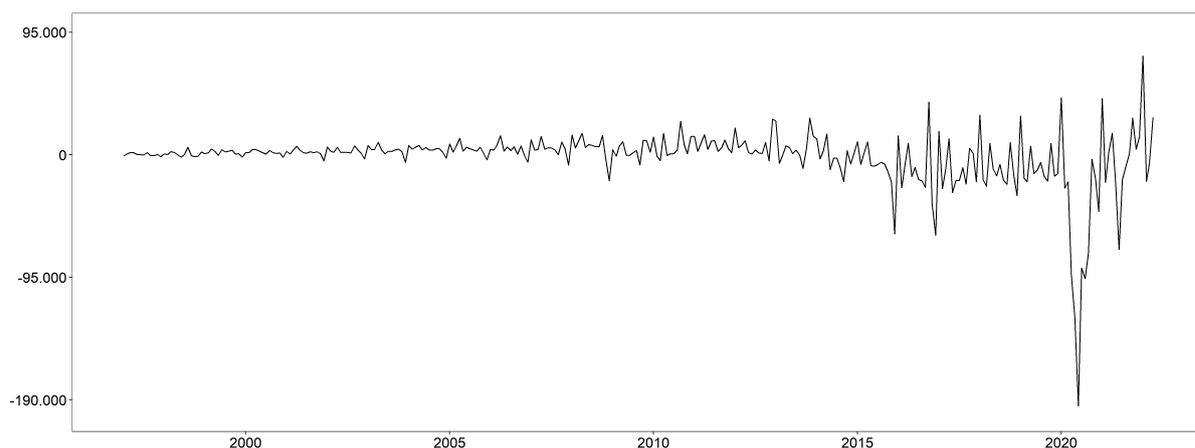
Fonte: Receita Federal do Brasil. Elaboração Própria

Figura A.3 – Despesa Total do Governo Central - em R\$ milhões



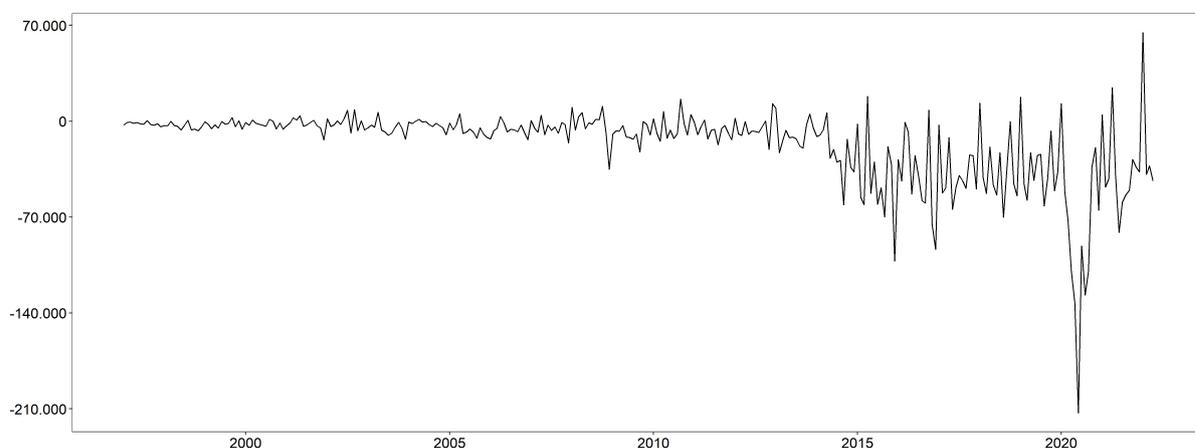
Fonte: Receita Federal do Brasil. Elaboração Própria

Figura A.4 – Resultado Primário do Governo Central - em R\$ milhões



Fonte: Receita Federal do Brasil. Elaboração Própria

Figura A.5 – Resultado Nominal do Governo Central - em R\$ milhões



Fonte: Receita Federal do Brasil. Elaboração Própria

APÊNDICE B – PALAVRAS-CHAVE UTILIZADAS NA CRIAÇÃO DO ÍNDICE GTU-BR

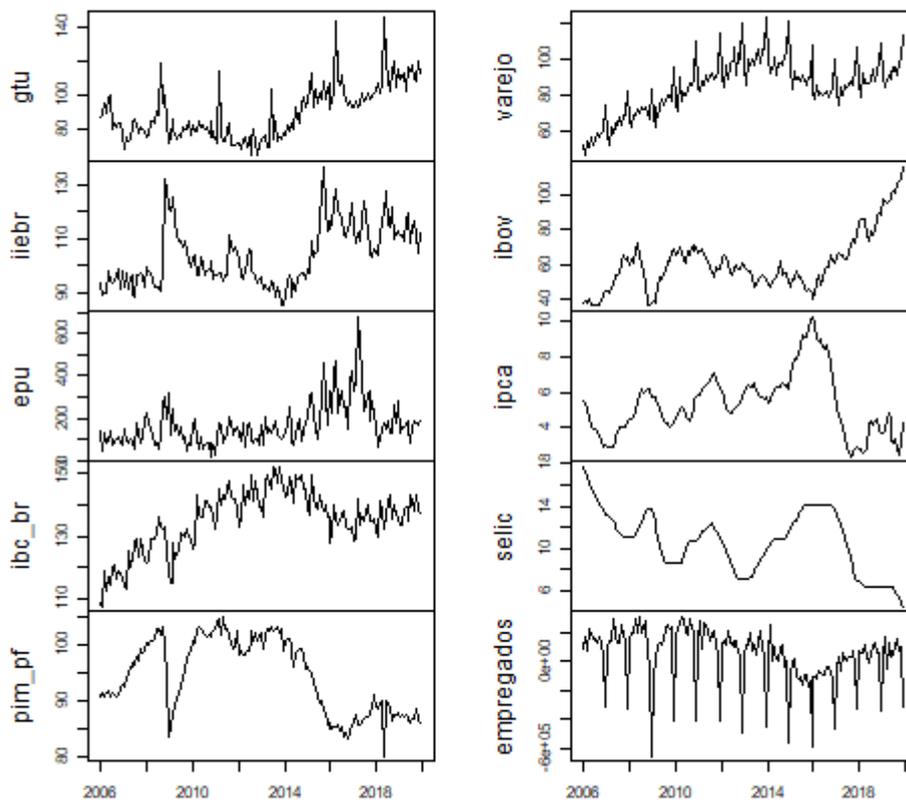
Tabela B.1 – Palavras-chave do índice GTU-BR

Ajustes fiscal	Economia brasileira	Pleno emprego
Alumínio	Economia da China	Política fiscal
América Latina	Economia dos Estados Unidos	Política monetária
Apagão	Economia mundial	Política monetária contracionista
Aperto monetário	Economias emergentes	Produção industrial
Arcabouço fiscal	Eleição presidencial	Protesto
Atividade econômica	Energia elétrica	Recessão
Banco Central Europeu	Estabilidade de preços	Reforma fiscal
Bandeira tarifária	Estados Unidos	Reforma previdenciária
Bovespa	Euro	Retomada da atividade econômica
Brexit	Europa	Risco Brasil
Câmbio	Expectativas de inflação	Risco de crédito
Caminhoneiros	FED	Salário mínimo
Cenário econômico mundial	Gás natural	Seca
China	Gasolina	Selic
Cobre	Gastos públicos	Soja
Combustível	Impeachment	Subprime
Commodities	Inflação	Taxa de juros dos Estados Unidos
Commodities internacionais	Japão	Telecomunicações
Conflito na Europa	Mercado de trabalho	Transporte
Conjuntura econômica	Mercado financeiro internacional	Tributos
Covid-19	Meta de inflação	Trigo
Crescimento econômico	Milho	
Demanda agregada	Níquel	
Desemprego	Pandemia	
Dólar	Petróleo	

Fonte: Elaboração própria

APÊNDICE C – TRAJETÓRIA E TESTES ECONOMÉTRICOS APLICADOS AS SÉRIES TEMPORAIS UTILIZADAS NO MODELO VAR

Figura C.1 – Trajetória temporal das variáveis



Fonte: Elaboração própria

Tabela C.1 – Correlação entre as variáveis utilizadas no trabalho

	GTU-Br	IIE-Br	EPU	IBC-Br	PIM-PF	Varejo	Ibov	IPCA	Selic	Empregados
IIE-Br	0,54	1								
EPU	0,40	0,58	1							
IBC-Br	0,07	-0,03	0,09	1						
PIM-PF	-0,64	-0,70	-0,51	0,32	1					
Varejo	0,08	0,05	0,11	0,84	0,15	1				
Ibov	0,44	0,20	-0,03	0,29	-0,20	0,37	1			
IPCA	0,02	0,18	0,27	0,30	0,04	0,22	-0,43	1		
Selic	-0,13	-0,03	0,19	-0,47	-0,09	-0,55	-0,75	0,48	1	
Empregados	-0,19	-0,31	-0,37	-0,05	0,38	-0,45	0,06	-0,30	-0,10	1

Fonte: Elaboração própria

Tabela C.2 – Estatística descritiva das variáveis

	GTU-Br	IIE-Br	EPU	IBC-Br	PIM-PF	Varejo	Ibov	IPCA	Selic	Empregados
Média	89,88	104,03	169,6	135,3	94,52	85,08	60,79	5,33	10,62	50.547
Desv. Padrão	15,31	11,09	100,56	8,89	6,60	13,26	16,31	1,78	2,85	100.394
Mediana	85,38	101,65	149,6	137,3	95,52	88,88	57,69	5,15	10,90	54.789
Máximo	146,13	136,80	677	149,7	105	105,34	115,65	10,22	17,65	234.306
Mínimo	64,89	85,10	22,30	113,20	80,12	54,14	36,23	2,43	4,59	-177.888
Curtose	3,47	2,49	7,52	2,73	1,54	2,58	4,10	3,17	2,22	2,50
Assimetria	0,80	0,58	1,78	-0,70	-0,14	-0,72	1,05	0,67	-0,04	-0,44

Fonte: Elaboração própria

Tabela C.3 – Testes de raiz unitária*

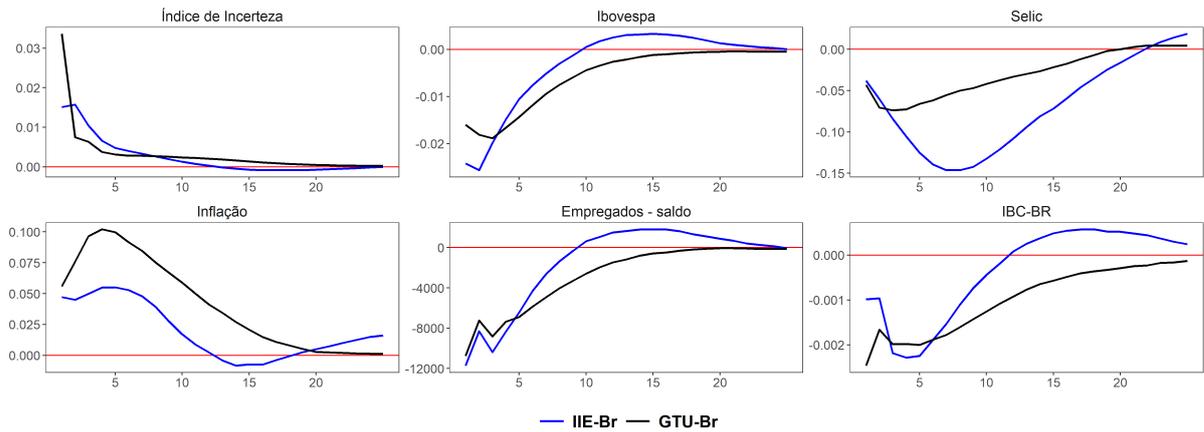
Variáveis	Est. ADF	Est. KPSS	Resultado
GTU-Br	-0,122	2,040	Raiz Unitária
IIE-Br	-0,010	1,153	Raiz Unitária
EPU	-1,951	1,227	Raiz Unitária
IBC-Br	1,256	1,696	Raiz Unitária
PIM-PF	-0,336	1,409	Raiz Unitária
Varejo	1,562	2,004	Raiz Unitária
Ibov	1,401	1,395	Raiz Unitária
IPCA	-0,686	0,440	Raiz Unitária
Selic	-1,474	0,874	Raiz Unitária
Saldo-empregados	-1,822	1,583	Raiz Unitária

*Valor crítico do teste ADF: -2,58, Valor crítico do teste KPSS: 0,739

Fonte: Elaboração própria

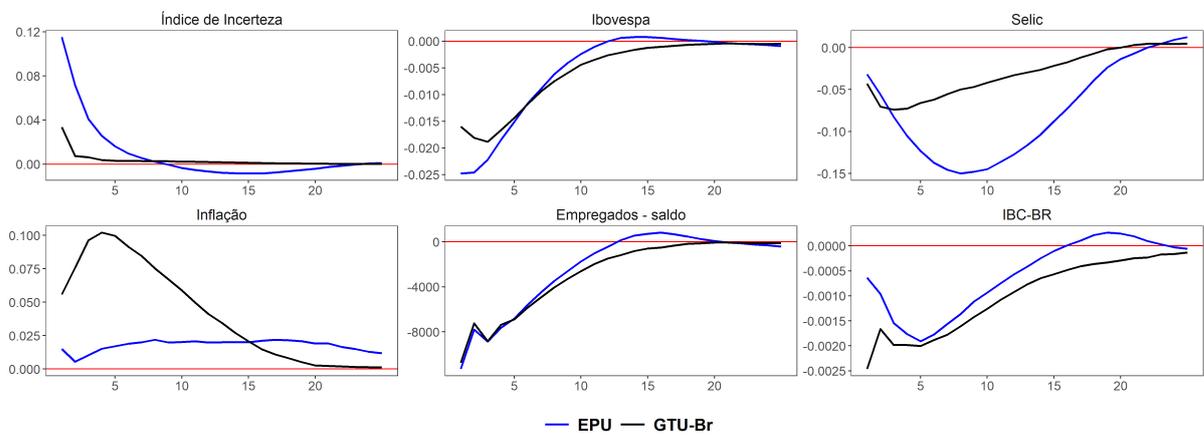
APÊNDICE D – GRÁFICOS DO TESTE DE ROBUSTEZ

Figura D.1 – Função impulso-resposta com índice de incerteza IIE-BR



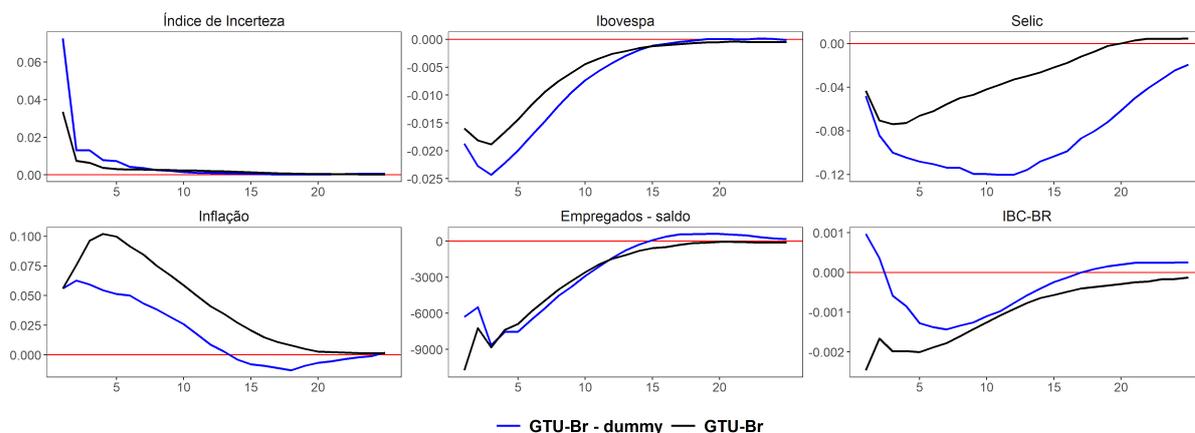
Fonte: Elaboração própria

Figura D.2 – Função impulso-resposta com índice de incerteza EPU



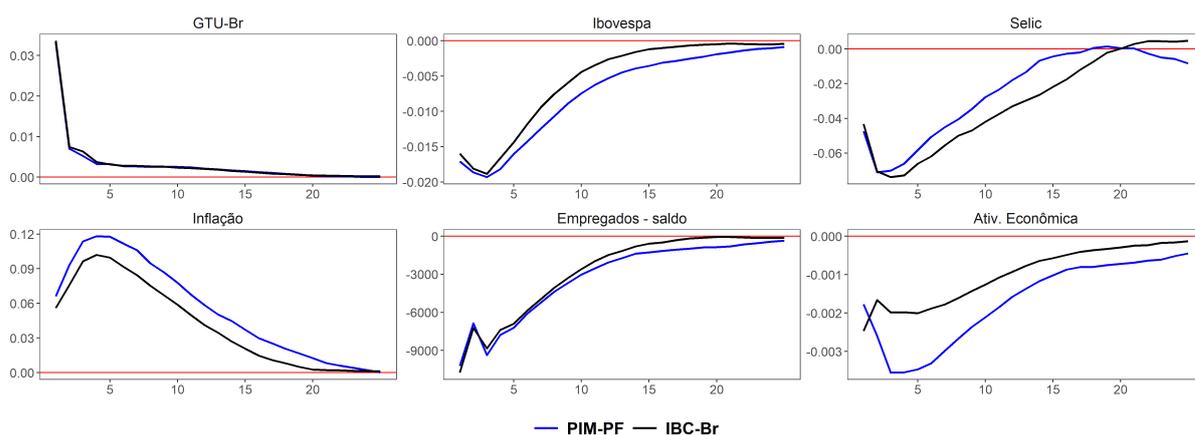
Fonte: Elaboração própria

Figura D.3 – Função impulso-resposta com índice de incerteza como variável dummy



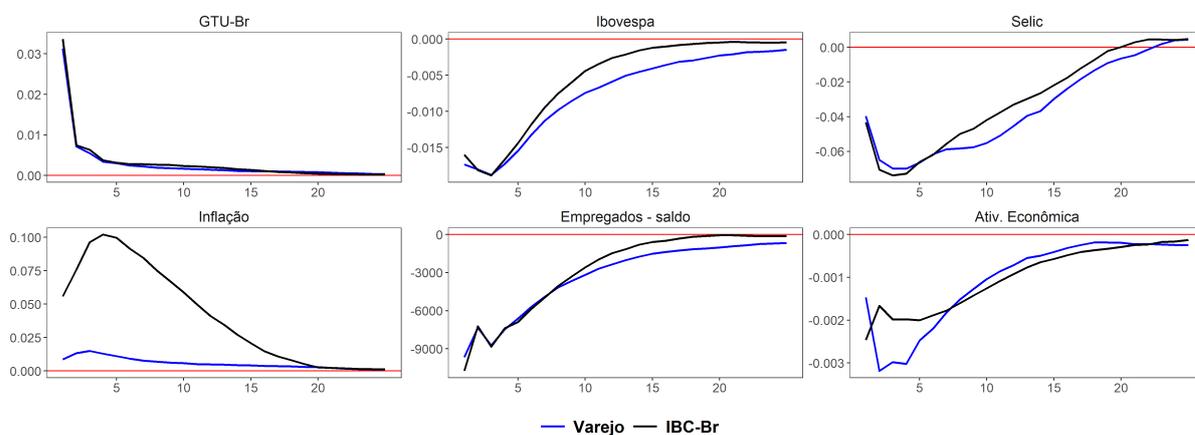
Fonte: Elaboração própria

Figura D.4 – Função impulso-resposta para PIM-PF



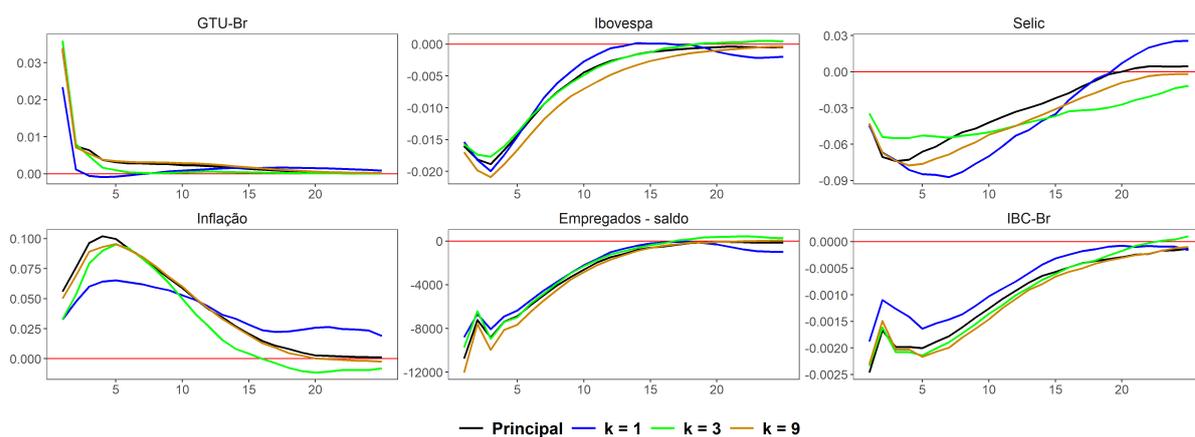
Fonte: Elaboração própria

Figura D.5 – Função impulso-resposta para Varejo



Fonte: Elaboração própria

Figura D.6 – Função impulso-resposta para diferentes durações de restrição



Fonte: Elaboração própria