



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

Rennan Kertlly de Medeiros

Ensaio em macroeconomia aplicada: choques
petrolíferos e previsão

João Pessoa

2022

Rennan Kertlly de Medeiros

**Ensaio em macroeconomia aplicada: choques
petrolíferos e previsão**

Tese de Doutorado apresentada ao Programa de Pós-Graduação em Economia da Universidade Federal da Paraíba, como requisito parcial para obtenção do título de Doutor em Economia.

Orientador: Prof. Dr. Edilean Kleber da Silva Bejarano Aragón

Coorientador: Prof. Dr. Cássio da Nóbrega Besarria

João Pessoa

2022

Catálogo na publicação
Seção de Catalogação e Classificação

M488e Medeiros, Rennan Kertlly de.

Ensaio em macroeconomia aplicada: choques petrolíferos e previsão / Rennan Kertlly de Medeiros. - João Pessoa, 2022.

68 f. : il.

Orientação: Edilean Kleber da Silva Bejarano Aragón.

Coorientação: Cássio da Nóbrega Besarria.

Tese (Doutorado) - UFPB/CCSA.

1. Macroeconomia. 2. Choques do petróleo. 3. Sentimento textual. 4. Projeção local. 5. Previsão fiscal. 6. Arrecadação. I. Aragón, Edilean Kleber da Silva Bejarano. II. Besarria, Cássio da Nóbrega. III. Título.

UFPB/BC

CDU 330.101.541(043)

Rennan Kertlly de Medeiros

**Ensaaios em macroeconomia aplicada: choques petrolíferos e
previsão**

Tese de doutorado apresentada ao Programa de Pós-Graduação em Economia da Universidade Federal da Paraíba - PPGE/UFPB, em cumprimento às exigências de conclusão do Curso de Doutorado em Economia. Submetida e **APROVADA** pela banca examinadora abaixo assinada.

Defesa realizada no Campus I da UFPB em João Pessoa-PB, em 21 de dezembro de 2022.

**Prof. Dr. Edilean Kleber da Silva
Bejarano Aragón**
Orientador

Prof. Dr. Cássio da Nóbrega Besarria
Coorientador

**Prof. Dr. José Luis da Silva Netto
Junior**
Avaliador Interno

**Prof. Dr. Luiz Renato Régis de
Oliveira Lima**
Avaliador Interno

Prof. Dr. Lucas Lúcio Godeiro
Avaliador Externo

Prof. Dr. Roberto Tatiwa Ferreira
Avaliador Externo

João Pessoa

2022

À minha mãe.

AGRADECIMENTOS

A Deus, pela saúde e perseverança.

Aos familiares que estiveram ao meu lado, em especial, minha mãe Maria do Céu.

Aos professores Edilean Aragón e Cássio Besarria, pelas orientações valiosas para desenvolvimento deste trabalho, pelo conhecimento transmitido, bem como pela amizade construída ao longo dos anos.

Ao professor Hélio Ramos, por me acompanhar/orientar na fase inicial de carreira (graduação), com ensinamentos que foram importantes para minha trajetória acadêmica.

A todo corpo docente do Programa de Pós-Graduação em Economia (PPGE) e do Departamento de Economia, da UFPB, pelo inestimável compromisso com o ensino.

À coordenação/secretaria do PPGE, pela presteza em atender as demandas dos alunos.

Aos amigos, pelo convívio.

À banca avaliadora, pelas sugestões e correções.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pelo fomento e incentivo fornecido a esta pesquisa.

RESUMO

Capítulo 1 - Efeitos do sentimento do mercado de petróleo sobre variáveis macroeconômicas

Este ensaio objetiva avaliar os efeitos dos choques de preços de petróleo em variáveis macroeconômicas, para as economias dos Estados Unidos e Brasil. Desenvolvemos uma variável que mensura a volatilidade de preços de petróleo, a partir de uma análise de sentimento textual. Avaliamos os choques de preços de petróleo a partir do método de Projeção Local. Nossos resultados sugerem que alterações nos preços de petróleo provocam impactos maiores na economia norte-americana, comparado aos efeitos na economia brasileira. As respostas das variáveis dos EUA e Brasil foram semelhantes quanto ao uso do indicador de sentimento ou o índice de volatilidade VIX. Por fim, constatamos que a diminuição de frequência das variáveis, juntamente com mudança de método, não altera as trajetórias de resposta das variáveis macroeconômicas.

Palavras-chave: Sentimento textual; Choques de petróleo; Projeção Local.

Capítulo 2 - Estratégias de previsão fiscal: um estudo empírico para a economia brasileira

Este ensaio objetiva realizar a previsão da arrecadação federal do Brasil a partir de diversos métodos de aprendizado de máquina, para diferentes amostras. Para previsão da variável de interesse foram utilizadas 34 variáveis explicativas. A metodologia de estimação utilizada está dividida em três categorias de aprendizagem: shrinkage, ponderação e fator. Os resultados sugerem que o modelo Elastic Net possui a maior acurácia para previsões mensais com uso de 20% e 30% da amostra para teste, para períodos curtos. Para previsão de períodos acumulados, o modelo LASSO tem alta performance. Por fim, constatamos que o modelo Bagging é limitado com aumento de períodos de previsão e, sobretudo, com diminuição de frequência das variáveis.

Palavras-chave: Previsão fiscal; Arrecadação; Modelos de machine learning; Elastic Net.

ABSTRACT

Chapter 1 - Effects of oil market sentiment on macroeconomic variables

This essay aims to evaluate the effects of oil price shocks on macroeconomic variables, for the economies of the United States and Brazil. We develop a variable that measures the volatility of oil prices, from a textual sentiment analysis. We evaluate oil price shocks using the Local Projection method. Our results suggest that changes in oil prices cause larger impacts on the US economy, compared to the effects on the Brazilian economy. The responses of the US and Brazilian variables were similar when using the sentiment indicator or the VIX volatility index. Finally, we find that decreasing the frequency of the variables, together with changing the method, does not change the response trajectories of the macroeconomic variables.

Keywords: Textual sentiment; Oil Shocks; Local Projection.

Chapter 2 - Fiscal forecasting strategies: an empirical study for the Brazilian economy

This article aims to forecasting Brazil's federal tax revenues from different machine learning methods, for different samples. To forecasting the variable of interest, 34 explanatory variables were used. The estimation methodology used is divided into three learning categories: shrinkage, ensemble and factor. The results suggest that the Elastic Net model has the highest accuracy for monthly forecasts using 20% and 30% of the test sample, for short periods. For forecasts of accumulated periods, the LASSO model has high performance. Finally, we verify that the Bagging model is limited with an increase in forecast periods and, above all, with a decrease in the frequency of variables.

Keywords: Textual sentiment; Oil Shocks; Local Projection.

LISTA DE ILUSTRAÇÕES

Figura 1.1 – Produção de Óleo Bruto dos EUA e Brasil (Toneladas)	15
Figura 1.2 – Consumo de Óleo Bruto dos EUA e Brasil (Toneladas)	16
Figura 1.3 – Trajetória de preços reais do petróleo WTI (linha sólida) e recessões nos Estados Unidos (área sombreada) (Barril/US\$)	17
Figura 1.4 – Nuvem de palavras do texto utilizado para construção do sentimento	22
Figura 1.5 – Trajetória da variável de sentimento do mercado de petróleo (Sent)	23
Figura 1.6 – Diagrama do processo de coleta e estimação do sentimento textual	24
Figura 1.7 – Respostas ao impulso em choques de petróleo - EUA	29
Figura 1.8 – Respostas ao impulso em choques de petróleo - Brasil	30
Figura 2.1 – Trajetória da arrecadação fiscal (Em Milhões de R\$)	36
Figura 2.2 – Correlação das previsões para Arrecadação de novembro de 2017 a outubro de 2021	49
Figura 2.3 – Correlação das previsões para Arrecadação de dezembro de 2015 a outubro de 2021	51
Figura A.1 – Respostas ao impulso em choques de petróleo com estimação por SVAR	59
Figura A.2 – Respostas ao impulso em choques de petróleo com frequência trimestral	60
Figura A.3 – Respostas ao impulso de petróleo WTI (Cavalcanti e Jalles (2013)) - Estimação por Projeção Local	61

LISTA DE TABELAS

Tabela 1.1 – Resumo das variáveis utilizadas	27
Tabela 2.1 – Resumo das variáveis utilizadas na previsão de arrecadação	38
Tabela 2.2 – Modelos de Previsão	47
Tabela 2.3 – RMSE para fora da amostra - Previsão de arrecadação - Novembro/2017 a outubro/2021	48
Tabela 2.4 – RMSE para fora da amostra - Previsão de arrecadação - Dezembro/2015 a outubro/2021	50
Tabela 2.5 – RMSE para fora da amostra - Previsão de arrecadação para fora da amostra - Primeiro trimestre de 2018 ao terceiro trimestre de 2021	52
Tabela A.1 – Estatísticas descritivas das variáveis utilizadas	62
Tabela A.2 – Defasagem ótima das variáveis endógenas com a inclusão da variável de choque Sent - Estimação por Projeção Local e SVAR - EUA	62
Tabela A.3 – Defasagem ótima das variáveis endógenas com a inclusão da variável de choque VIX - Estimação por Projeção Local e SVAR - EUA	62
Tabela A.4 – Optimal lag of the exogenous variable Sent	63
Tabela A.5 – Defasagem ótima da variável exógena VIX	63
Tabela A.6 – Defasagem ótima das variáveis endógenas - Estimação por Projeção Local para variáveis mensais (sem a variável de choque) - Brasil	63
Tabela A.7 – Defasagem ótima das variáveis endógenas - Estimação por Projeção Local para variáveis trimestrais (com inclusão da variável de choque VIX) - EUA	64
Tabela B.1 – Estatísticas descritivas das variáveis utilizadas	65

LISTA DE ABREVIATURAS E SIGLAS

VAR	Vetores Auto-Regressivos
OVX	Índice de Volatilidade de Petróleo
ANP	Agência Nacional de Petróleo
OPEP	Organização dos Países Exportadores de Petróleo
PIB	Produto Interno Bruto
FAVAR	Vetor Auto-Regressivo com Fator Aumentado
ASEAN-5	Associação das Nações do Sudeste Aisático
IRF	Funções de Impulso-Respostas
VIX	Índice de Volatilidade S&P 500 do mercado de ações norte-americano
OLS	Mínimos Quadrados Ordinários
IPCA	Índice Nacional de Preços ao Consumidor Amplo
SVAR	Vetores Auto-Regressivos na forma estrutural
IBGE	Instituto Brasileiro de Geografia e Estatística
SELIC	Taxa de juros básica do Brasil
AIC	Critério de Informação de Akaike
BIC	Critério Bayesiano de Schwarz
HQ	Critério de Hannan-Quin
CRS	Regressões de Subconjunto Completadas
RMSE	Raiz do Erro Quadrático Médio
Sent	Sentimento do mercado de petróleo

SUMÁRIO

1	Efeitos do sentimento do mercado de petróleo sobre variáveis macroeconômicas	12
1.1	Introdução	12
1.2	Mercado de petróleo e macroeconomia	14
1.2.1	Revisão da literatura empírica	18
1.3	Construção da variável de sentimento para mercado de petróleo	21
1.4	Metodologia e dados	24
1.4.1	Estimação por Projeção Local	24
1.4.2	Banco de dados	26
1.5	Resultados e discussão	27
1.6	Resultados adicionais	31
1.7	Conclusões	33
2	Estratégias de previsão fiscal: um estudo empírico para a economia brasileira	34
2.1	Introdução	34
2.2	Banco de Dados	36
2.3	Metodologia e Modelos de aprendizado de máquina	39
2.3.1	Metodologia	39
2.3.2	Shrinkage	41
2.3.2.1	Ridge	41
2.3.2.2	LASSO	42
2.3.2.3	LASSO Adaptativo	43
2.3.2.4	Elastic Net	43
2.3.3	Métodos de ponderação (<i>ensemble</i>)	44
2.3.3.1	Bagging	44
2.3.3.2	CSR	45

2.3.4	Modelos Fatoriais	46
2.3.4.1	Fatores <i>Target</i>	46
2.3.4.2	Fatores Boosting	47
2.3.5	Acurácia dos modelos	47
2.4	Resultados e discussão	48
2.5	Teste de robustez	51
2.6	Conclusões	52
	Referências	54
	Apêndices	58
	APÊNDICE A – Capítulo 1	59
	APÊNDICE B – Capítulo 2	65
	Anexos	66
	ANEXO A – Capítulo 2	67

1 EFEITOS DO SENTIMENTO DO MERCADO DE PETRÓLEO SOBRE VARIÁVEIS MACROECONÔMICAS

1.1 INTRODUÇÃO

O comportamento dos preços mundiais de petróleo reflete na atividade econômica por diversos canais. [Kilian \(2009\)](#) sinaliza que a trajetória de preços petrolíferos afeta a economia em diversos segmentos, desde orçamentos familiares a lucros das empresas. Nesse sentido, as grandes recessões mundiais possuem forte correlação com aumentos bruscos nos preços de petróleo.

Em geral, a literatura econômica tem diversos estudos sobre impactos de choques nos preços de petróleo em variáveis macroeconômicas. Em um estudo pioneiro para a economia norte-americana, [Hamilton \(1983\)](#) indica uma possível causalidade entre aumentos de preços de petróleo e atividade econômica. Posteriormente, vários trabalhos relacionam choques nos preços de petróleo sobre variáveis macroeconômicas a partir da técnica de Vetores Auto-Regressivos (VAR), como [Bernanke et al. \(1997\)](#), que testa a identificação da resposta da política monetária a choques petrolíferos.¹ Outros autores avaliam os impactos de choques de petróleo por meio de modelos alternativos, como o modelo Dinâmico Estocástico de Equilíbrio Geral (DSGE), estudo realizado por [Zhao et al. \(2016\)](#). Para mensuração de choques do petróleo, [Kocaaslan \(2019\)](#) utiliza o Índice de Volatilidade de Petróleo (VOX) como *proxy* de incerteza para avaliar seus efeitos sobre a atividade econômica. Por sua vez, [Kocaarslan, Soytaş e Soytaş \(2020\)](#) analisam choques de petróleo a partir de uma variável de incerteza, construída a partir de um modelo GARCH sobre a série de preços de petróleo.

Ao longo da história dos Estados Unidos, as grandes recessões estiveram fortemente

¹ Para outros estudos que utilizam extensões de modelos de vetores auto-regressivos para análise de choques de preços de petróleo sobre atividade econômica, ver [Burbidge e Harrison \(1984\)](#), [Hamilton e Herrera \(2004\)](#), [Kilian \(2009\)](#), [Aastveit, Bjørnland e Thorsrud \(2015\)](#), dentre outros.

ligadas com choques de oferta ou demanda de petróleo, o que provoca impacto nos preços do petróleo e, conseqüentemente, nos seus derivados. Conforme Peersman e Robays (2009) e Kilian (2009), existe uma redução no impacto de choques de preços do petróleo na atividade econômica norte-americana ao longo do tempo, derivado do surgimento de combustíveis alternativos. Aqueles autores indicam que choques de preços petrolíferos são conseqüências de choques indiretos de demanda e oferta de petróleo.

Por sua vez, ainda não existe consenso sobre os impactos de choques nos preços de petróleo sobre a economia brasileira. Cavalcanti e Jalles (2013) sinalizam que choques petrolíferos têm pouco impacto sobre a atividade econômica e a inflação, bem como sobre a volatilidade do PIB brasileiro, durante o período de 1980 a 2010. Entretanto, a produção de petróleo no Brasil cresce a taxas superiores a 2% desde 2011, de acordo com Agência Nacional de Petróleo (ANP). Ainda de acordo com essa agência, o Brasil está entre os 10 maiores produtores dessa *commodity*. Resultados futuros podem expressar uma maior dependência da economia brasileira em relação ao comportamento de preços do petróleo. Ressaltamos que, desde 2016, o Brasil adota uma regra de paridade de preços internacionais do petróleo para precificação dos combustíveis e demais derivados do petróleo.

Assim, o objetivo desse estudo é analisar os efeitos de choques de petróleo sobre a atividade econômica dos Estados Unidos e do Brasil. A análise para esses dois países é relevante porque eles podem reagir de maneiras diferentes a choques petrolíferos, devido às composições heterogêneas de agregados econômicos, instituições e políticas distintas, bem como, a dependência das importações de petróleo. Ademais, a maior parcela da produção de petróleo bruto no mundo está concentrada nos Estados Unidos. Por sua vez, o Brasil é o maior produtor de petróleo da América Latina, segundo IEA (2020a). Nesse caso, serão analisadas as respostas de três variáveis macroeconômicas: PIB, inflação e taxa de juros. A escolha das variáveis de resposta foi baseada em Bernanke et al. (1997).

Este ensaio contribui para a literatura sobre choques de petróleo por meio da construção de um índice de sentimento textual para o mercado de petróleo, baseado nos relatórios mensais da Organização dos Países Exportadores de Petróleo (OPEP), no

período de janeiro de 2001 a dezembro de 2020. Os relatórios explorados tratam dos movimentos de preços do petróleo produzido pelos países membros da OPEP, denominado na literatura de Basket. O desenvolvimento de variáveis a partir de técnicas de sentimento textual ainda é limitado nessa literatura de choques de petróleo.² [Medeiros et al. \(2022\)](#) desenvolvem um indicador que capta o sentimento do mercado de petróleo, para o período de 1995 a 2017, entretanto, o índice de petróleo é baseado em relatórios da Agência Internacional de Energia. Além do desenvolvimento de um novo índice para o mercado petrolífero, estimamos as respostas das variáveis macroeconômicas aos choques petrolíferos a partir do método de projeção local, empregado por [Jordà \(2005\)](#).

Além dessa introdução, este ensaio está dividido em mais seis seções. A Seção 1.2 é reservada a uma contextualização da relação de petróleo com variáveis macroeconômicas, como inflação, PIB e taxa de juros nominal. A Seção 1.3 mostra a construção detalhada do índice de sentimento textual para o mercado petrolífero. A Seção 1.4 é reservada para a abordagem empírica e os dados utilizados. A Seção 1.5 apresenta os resultados obtidos pelo estudo. A Seção 1.6 é dedicada para resultados adicionais (robustez). E, por fim, a Seção 1.7 mostra as principais conclusões.

1.2 MERCADO DE PETRÓLEO E MACROECONOMIA

De acordo com a [IEA \(2020b\)](#), o petróleo tem grande impacto na economia mundial, em que responde por mais de 31% da energia primária global, sendo mais utilizado no setor de transportes. Nesse sentido, [Hamilton \(1996\)](#) indica que choques no mercado de petróleo impactam variáveis macroeconômicas por diversos canais de transmissão, por exemplo, com aumento nos custos de transporte, o que provoca aumento de inflação na economia.

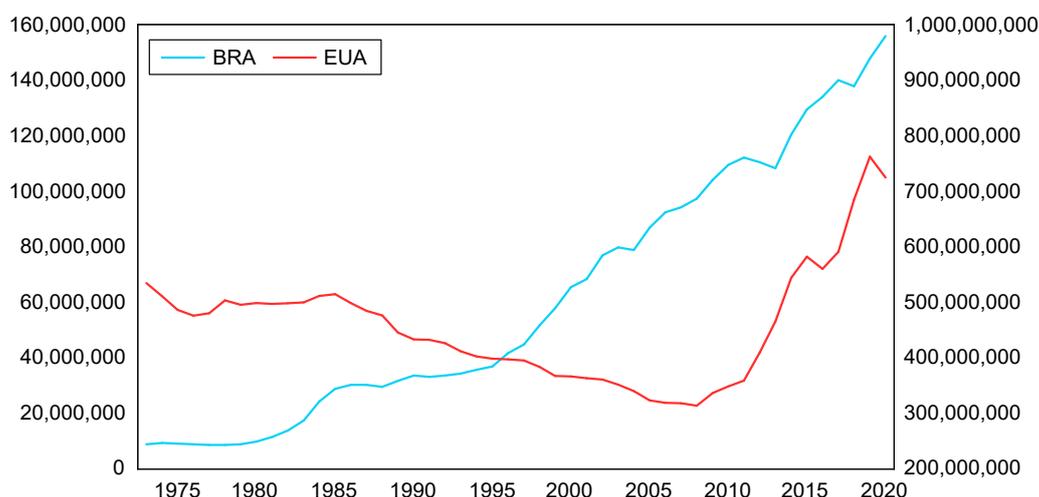
Conforme [IEA \(2020a\)](#), o petróleo é responsável por 32% da energia fornecida no mundo, seguido por carvão (27%) e gás natural (23%). Embora a participação do petróleo na geração de eletricidade tenha diminuído ao longo do tempo (25% em 1973 para 3% em

² Ainda existe uma carência na literatura acerca de choques de petróleo a partir de análise de sentimento textual, entretanto, para outras análises econométrica e macroeconômica há diversos estudos, como os trabalhos de [Fève e Guay \(2019\)](#), [Algaba et al. \(2020\)](#) e [Correa et al. \(2021\)](#).

2019), ainda é a principal fonte energética em outros segmentos, por exemplo, responde por 91% do consumo de energia no setor de transporte e 70% do consumo de matéria-prima nos setores químico e petroquímico.

Os maiores produtores de petróleo bruto no mundo, são: Estados Unidos (1º), Arábia Saudita (2º), Rússia (3º), Brasil (8º), dentre outros. Segundo a IEA (2020a), os Estados Unidos produziram mais de 725 milhões de toneladas de óleo bruto, em 2020. Por sua vez, o Brasil produziu mais de 156 milhões de toneladas de petróleo, para o mesmo período. A Figura 1.1 mostra a evolução da produção de óleo bruto dos Estados Unidos e Brasil, para o período de janeiro de 1973 a dezembro de 2020.

Figura 1.1 – Produção de Óleo Bruto dos EUA e Brasil (Toneladas)



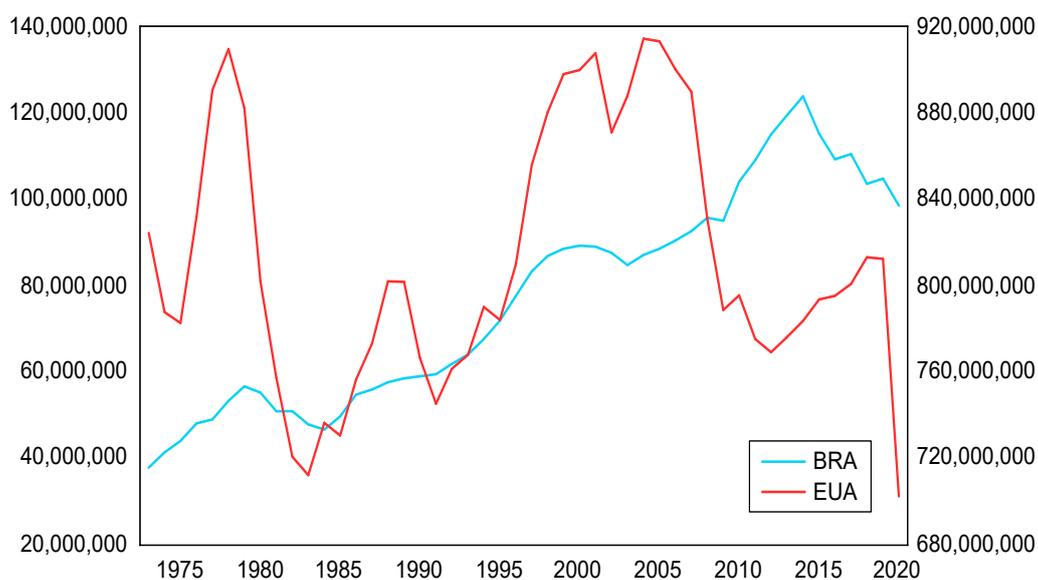
Fonte: Elaboração própria.

[1] Nota: O eixo vertical à esquerda representa os valores do Brasil. O eixo vertical à direita é referente aos Estados Unidos.

Com relação ao consumo de petróleo, os Estados Unidos também lideram o ranking mundial, seguido dos seguintes países: China (2º), Índia (3º), Brasil (8º), dentre outros. EUA e China respondem por mais de um terço da demanda mundial por petróleo. Por sua vez, o Brasil demanda pouco mais de 2% dessa *commodity*. A Figura 1.2 mostra a trajetória de consumo por petróleo para EUA e Brasil. As estatísticas indicam que, em decorrência da diferença entre ofertantes e demandantes, o petróleo é bastante comercializado, seja óleo bruto ou quaisquer produtos derivados.

De modo particular, para o ano de 2020, constatamos que houve uma diminuição da produção de petróleo no mundo, sobretudo dos grandes produtores (EUA, países membros da OPEP, dentre outros). De acordo com dados dos relatórios da OPEP, em determinados meses de 2020, a oferta de petróleo esteve inferior ao consumo mundial, o que provocou aumento de preços de petróleo, conforme mostrado na Figura 1.3. Esse resultado é consequência da pandemia da Covid-19, notificada em dezembro de 2019.

Figura 1.2 – Consumo de Óleo Bruto dos EUA e Brasil (Toneladas)



Fonte: Elaboração própria.

[2] Nota: O eixo vertical à esquerda representa os valores do Brasil. O eixo vertical à direita é referente aos Estados Unidos.

Conforme relatórios da OPEP, a pandemia da Covid-19 e suas medidas de restrição à mobilidade urbana têm impactos sem precedentes na demanda global de petróleo. Por exemplo, ao longo de 2020, houve uma redução da demanda de petróleo mundial de 9 mb/d (milhões de barris/dia). Por outro lado, com forte retração da oferta de petróleo, os preços dessa *commodity* aumentaram e afetaram os indicadores de inflação da economia global. Conforme já demonstrado, como o setor de transporte é bastante dependente do óleo bruto, o resultado final é aumento de inflação no mundo inteiro, dado o encarecimento para transporte de mercadorias.

Com propósito de analisar o efeito do mercado de petróleo na macroeconomia,

Hamilton (1983) testa a causalidade entre preços de petróleo e variáveis macroeconômicas dos Estados Unidos. O estudo indica uma forte relação entre preços de petróleo e o PIB americano, sendo a *commodity* responsável por parte das recessões deste país desde a Segunda Guerra Mundial. Ainda, é destacado que a performance da economia está diretamente relacionada com o setor energético, principalmente após a criação da OPEP. A Figura 1.3 sinaliza uma forte relação entre aumentos de preços do petróleo e recessões norte-americanas, em uma série histórica de janeiro de 1973 a dezembro de 2020.

Figura 1.3 – Trajetória de preços reais do petróleo WTI (linha sólida) e recessões nos Estados Unidos (área sombreada) (Barril/US\$)



Fonte: Macrotrends.

A primeira área hachurada sinaliza uma recessão na economia norte-americana, consequência do primeiro choque de petróleo na década de 1970. Alguns anos depois, emerge o segundo choque de petróleo na economia mundial, o que leva a aumentos de preços dessa *commodity* e, conseqüentemente, períodos de recessão nos Estados Unidos. Após os choques petrolíferos, a grande crise vivenciada pelos Estados Unidos acontece após o boom das *commodities*, decorrência da crise do subprime, o que provoca forte redução no preço de petróleo. Por fim, temos os efeitos da Covid-19 durante o ano de 2020, que inicialmente reduziu o preço do óleo bruto, porém, acarreta um forte crescimento no preço

do óleo, em virtude da baixa oferta mundial.

1.2.1 Revisão da literatura empírica

A literatura de choques de petróleo em variáveis macroeconômicas é ampla, porém, ainda existem lacunas que podem ser exploradas, conforme apresentamos nesse estudo. Para um estudo durante a década de 1970, [Burbidge e Harrison \(1984\)](#) testam o efeito de um aumento nos preços de petróleo em variáveis macroeconômicas dos Estados Unidos, Japão, Canadá e Alemanha. Baseado em modelos de vetores auto-regressivos, os autores indicam que a crise do petróleo na década 1970 (1973-74) explica parte do desempenho da produção industrial em cada país analisado, o que corrobora o resultado encontrado por [Hamilton \(1983\)](#). Entretanto, a alta volatilidade do preço do petróleo no período de 1979-80 teve baixo impacto na produção industrial.

[Bernanke et al. \(1997\)](#) utilizam técnicas baseadas em VAR para identificação da resposta da política monetária a choques nos preços de petróleo. Os resultados indicam que a resposta endógena da política monetária em parte é a causa dos efeitos negativos dos choques petrolíferos sobre a atividade econômica. Outro destaque é que não somente a política monetária, mas também outros distúrbios não monetários e não ligados a petróleo desempenham funções relevantes, como choques em *commodities* agrícolas.

[Hamilton e Herrera \(2004\)](#) sugerem que a magnitude do efeito da política monetária para evitar um aumento da taxa de juros (consequência contracionista) a um choque adverso no preço de petróleo não é grande suficiente, como mostrado por [Bernanke et al. \(1997\)](#). Sendo assim, os choques de petróleo podem ter um impacto na economia superior ao que havia sido mostrado. A justificativa para esse resultado foi o número incorreto de defasagens do modelo auto-regressivo usado por [Bernanke et al. \(1997\)](#). Ainda, [Hamilton e Herrera \(2004\)](#) sinalizam que é duvidosa a eficácia de implementação da política monetária necessária para neutralizar até pequenos choques no preço de petróleo.

De outro modo, com a utilização de projeções de mínimos quadrados, [Kilian \(2008\)](#) sugere que somente uma pequena parcela dos aumentos de preços do petróleo durante a década de 1970 pode ser concedida à redução na produção de petróleo. De modo geral,

choques exógenos na produção de petróleo têm pouco efeito no comportamento da economia norte-americana desde os anos 70.

[Kilian \(2009\)](#) utiliza um modelo VAR estrutural com objetivo de identificar os choques contidos de demanda e oferta no mercado petrolífero mundial. O estudo sinaliza que um choque positivo no preço de petróleo pode ter diferentes efeitos no preço real dessa *commodity*, sendo a causa implícita da elevação de preços que determina a trajetória final da variável de resposta. Por um lado, uma elevação na demanda por petróleo provoca um aumento rápido, duradouro e grande no preço real do petróleo. De outro lado, uma redução na oferta de petróleo causa um crescimento menor e transitório no preço real do petróleo nos períodos iniciais.

A principal limitação de [Kilian \(2009\)](#) e dos demais estudos empíricos, na visão de [Baumeister e Peersman \(2013\)](#), é que eles se fundamentam em regressões invariantes no tempo, ou seja, os efeitos de choques do petróleo em indicadores macroeconômicos não mudam ao longo do tempo. Para tornar as estimativas mais robustas, [Baumeister e Peersman \(2013\)](#) utilizam uma abordagem bayesiana de vetores auto-regressivos com parâmetros variantes no tempo para analisar os efeitos de choques de oferta de petróleo em agregados macroeconômicos. Os resultados indicam uma redução significativa na elasticidade preço de curto prazo da demanda de petróleo a partir do início da década de 1980. De modo geral, os autores sinalizam que os impactos de choques de oferta de petróleo na atividade econômica norte-americana foram leves.

[Cavalcanti e Jalles \(2013\)](#) analisam os choques dos preços de petróleo em duas economias diferentes, Brasil e Estados Unidos. Os autores avaliam os choques petrolíferos em inflação e PIB para dois períodos distintos: 1975-1984 e 1985-2008. A partir de um modelo auto-regressivo multivariado, os resultados sugerem que a volatilidade da atividade econômica norte-americana teve redução, assim como houve uma diminuição do efeito de choques de petróleo nessa volatilidade. Para o Brasil, tais choques têm impacto incerto na atividade econômica e baixo efeito na inflação e volatilidade da taxa de crescimento do produto.

Sob a ótica da demanda, [Aastveit, Bjørnland e Thorsrud \(2015\)](#) contribuem com

a literatura no sentido de analisar a relevância da demanda das economias emergentes e desenvolvidas na condução do preço real de petróleo. Para isso, os autores estimam um Vetor Auto-Regressivo com Fator Aumentado (FAVAR) que envolve indicadores observáveis da produção de petróleo mundial e do preço real de petróleo. Os autores indicam que a demanda de países em desenvolvimento, sobretudo os países asiáticos, tem duas vezes mais importância no papel das flutuações no preço de petróleo real. Outro destaque é que regiões geográficas são impactadas de formas diferentes por um choque positivo no preço real do petróleo.

Com a perspectiva de estudar o impacto dos choques de preços do petróleo no produto real, na inflação e na taxa de câmbio de países asiáticos integrantes do ASEAN-5³, [Basnet e Upadhyaya \(2015\)](#) estimam um modelo de vetores auto-regressivos na sua forma estrutural. Os resultados indicam que as variáveis macroeconômicas analisadas possuem relação de longo prazo em comum, conforme apontado por testes de cointegração. Em contraste com [Aastveit, Bjørnland e Thorsrud \(2015\)](#), as funções de respostas ao impulso (IRF, sigla em inglês) mostram que a volatilidade do preço do petróleo não afeta os países asiáticos analisados no longo prazo. Esse estudo revela ainda que os preços de petróleo não têm relevância na explicação das variáveis macroeconômicas dos países asiáticos.

Já [Zhao et al. \(2016\)](#) avaliam o impacto de diferentes choques dos preços de petróleo no produto e na inflação na maior economia da Ásia, a China.⁴ Para isso, os autores estabelecem um modelo DSGE com duas economias: China e o resto do mundo. As simulações revelam que os choques políticos por membros da OPEP geram efeitos de curto prazo na economia chinesa; por outro lado, os demais choques produzem efeitos de médio e longo prazos. Ainda, a maior parcela das flutuações da atividade econômica e da inflação na China são afetadas por choques de demanda específicos do mercado petrolífero.

Na tentativa de avaliar a economia global, [Choi et al. \(2018\)](#) realizam um estudo sistemático do efeito de choques mundiais de preços de petróleo sobre a inflação de 72

³ ASEAN-5 significa Associação das Nações do Sudeste Asiático. Os cinco países membros dessa associação, são: Indonésia, Malásia, Filipinas, Cingapura e Tailândia.

⁴ Foram estudados quatro tipos de flutuações no preço de petróleo: choques de oferta de petróleo pelos membros da OPEP, outros choques de oferta, choques de demanda por *commodities* industriais, e choques de demanda específicos para o mercado petrolífero.

países emergentes e desenvolvidos, para o período de 1970 a 2015. A partir da estratégia de estimação via método de projeção local, proposto por [Jordà \(2005\)](#), os principais resultados sugerem que um aumento de 10% nos preços globais de petróleo, em média, provoca um aumento da inflação doméstica em aproximadamente 0,4 pontos percentuais, sendo o efeito não estatisticamente significativo dois anos após o choque. Ao longo do tempo, o choque de preços de petróleo sofre diminuição sobre o nível de preços dos países analisados é reduzido devido a duas características: política monetária com maior credibilidade e menor necessidade de importações de energia em virtude da maior produção doméstica.

Além de testar os impactos dos choques de preços de petróleo na taxa de desemprego dos Estados Unidos, [Kocaaslan \(2019\)](#) avalia os efeitos da incerteza de preços de petróleo no desemprego, a partir de uma abordagem de vetores auto-regressivos para um modelo ARCH generalizado (GARCH), que permite erros na média. O período avaliado varia do segundo trimestre de 1974 ao quarto trimestre de 2017. Os resultados sinalizam que a incerteza do preço de petróleo aumenta a taxa de desemprego dos Estados Unidos. Da mesma forma, um choque positivo nos preços de petróleo provoca maior desemprego. Por sua vez, [Kocaarslan, Soytaş e Soytaş \(2020\)](#) a partir de técnicas de co-integração, analisam a presença de assimetria entre preços de petróleo, incerteza do preço de petróleo, taxa de juros e taxa de desemprego. Diferente do estudo anterior, os resultados sugerem que um aumento da incerteza do preço do petróleo não tem efeito significativo sobre desemprego.

Diante do exposto, mais especificamente para estudos recentes que tratam do impacto da incerteza do mercado de petróleo sobre variáveis macroeconômicas, observamos que ainda não existe um consenso na literatura que indique a direção exata de resposta das variáveis macroeconômicas.

1.3 CONSTRUÇÃO DA VARIÁVEL DE SENTIMENTO PARA MERCADO DE PETRÓLEO

Com objetivo de contribuir com a literatura de choques do petróleo em variáveis macroeconômicas, construímos uma variável de sentimento para o mercado de petróleo mundial. Essa série temporal para o mercado petrolífero é construída a partir da captação

de sentimento dos relatórios mensais disponibilizados pela OPEP, que abrange o período de janeiro de 2001 até dezembro de 2020, o que totaliza 240 relatórios da OPEP, sendo um relatório para cada mês analisado. Devido os relatórios serem amplos, para esse caso, nossa análise será destinada à seção de movimentos dos preços de petróleo.⁵

Após a importação dos relatórios, a segunda etapa é a limpeza dos documentos no sentido que ocorra remoção de elementos que não possuem relevância para estimação do índice, por exemplo: espaços duplos, pontuação, quebras de linha, quebras de página, espaçamento de parágrafos, dentre outros. Essa etapa foi executada com uso do pacote 'tm', do software R. Por exemplo, utilizamos o argumento "Smart" para remover as palavras irrelevantes dos relatórios. A partir da limpeza realizada nos textos (relatórios), ilustramos as palavras mais frequentes nos 240 relatórios tratados através de uma nuvem de palavras, conforme mostrada na Figura 1.4.

Figura 1.4 – Nuvem de palavras do texto utilizado para construção do sentimento



Fonte: Elaboração própria.

Na terceira etapa, o indicador de sentimento é obtido a partir do algoritmo de Jockers (2017), que seleciona palavras de acordo com seus aspectos cognitivos positivo ou negativo. O dicionário utilizado pelo algoritmo, para classificação dos aspectos cognitivos das palavras, é especificado por um dicionário de palavras pré-estabelecido, conforme

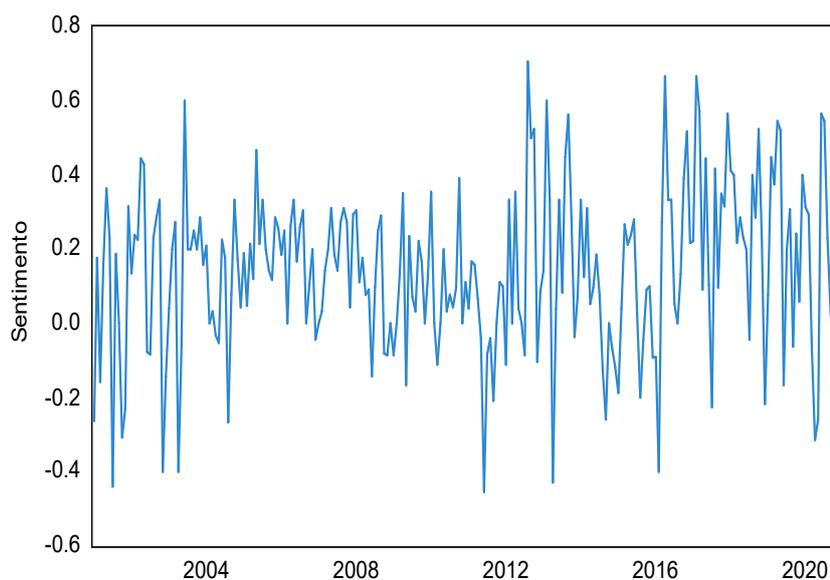
⁵ Os relatórios utilizados para construção da variável de sentimento podem ser consultados no link: https://www.opec.org/opec_web/en/.

Deeney et al. (2015). Em seguida, com base na contagem de palavras positivas e negativas presentes nos relatórios, construímos um índice que representa o tom ou sentimento dos relatórios ligados ao setor petrolífero, conforme apresentado na Equação 1.1:

$$\text{Sent}_t = \frac{\sum \text{Palavras Positivas} - \sum \text{Palavras Negativas}}{\sum \text{Palavras Positivas} + \sum \text{Palavras Negativas}} \quad (1.1)$$

O índice Sent_t varia no intervalo de $[-1, 1]$. O tom dos relatórios da OPEP é positivo caso o valor do índice seja maior que 0. Por outro lado, caso o valor do índice seja menor que 0, o tom é negativo. Por fim, quando o valor do índice é igual a 0, o resultado do tom é neutro, algo que não é esperado.

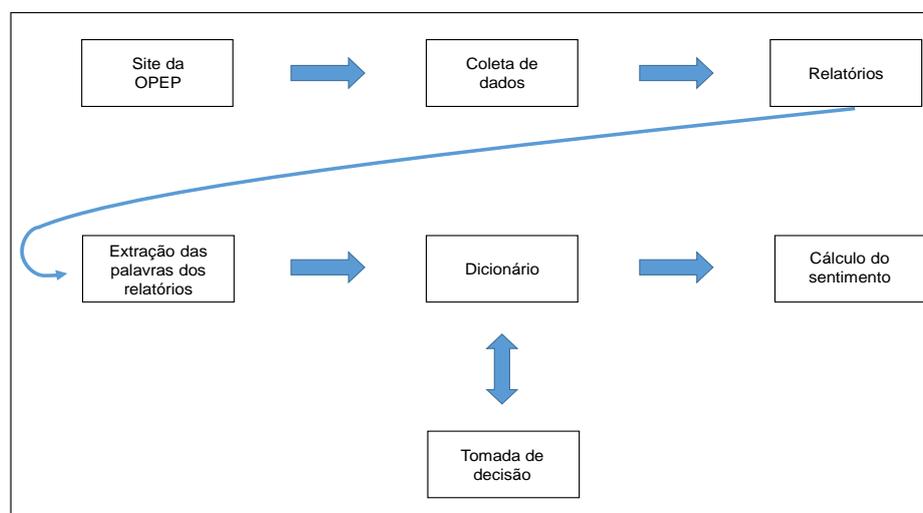
Figura 1.5 – Trajetória da variável de sentimento do mercado de petróleo (Sent)



Fonte: Elaboração própria.

A Figura 1.5 mostra a trajetória da variável Sent , construída a partir de técnicas de análise sentimenal textual, para o período de janeiro de 2001 a dezembro de 2020. O valor do sentimento mais positivo em relação ao mercado petrolífero é de 0.71. Por outro lado, o tom mais negativo atingiu -0.45. Ressaltamos que a média do tom das palavras é positiva, ou seja, existem mais palavras positivas nos relatórios analisados. Dessa forma, é esperado notícias mais positivas sobre petróleo, quando comparadas com notícias negativas acerca dessa *commodity*. A variável de sentimento desenvolvida será utilizada para identificação do choque de petróleo.

Figura 1.6 – Diagrama do processo de coleta e estimação do sentimento textual



Fonte: Elaboração própria.

A Figura 1.6 mostra um resumo do processo de coleta dos dados e estimação do sentimento textual para o mercado petrolífero. Sendo assim, observamos que a construção do sentimento possui três etapas, a saber: i) acessar o site da OPEP, coletar os dados dos relatórios; ii) extrair as palavras dos relatórios, utilizar o dicionário e calcular o sentimento; e, por fim, iii) a última etapa é a tomada de decisão a partir da série temporal gerada. O indicador de sentimento construído pode ser utilizado para aplicação econômica ou outra qualquer. Nesse caso, utilizamos o sentimento como choque em variáveis econômicas, para EUA e Brasil.

1.4 METODOLOGIA E DADOS

1.4.1 Estimação por Projeção Local

Nesse estudo, detalhamos a metodologia usada para avaliar os efeitos de choques do mercado de petróleo mundial em variáveis macroeconômicas dos EUA e Brasil, a saber: PIB, taxa de juros e inflação. Tais efeitos de choques foram avaliados a partir de uma investigação estocástica das variáveis. As variáveis que serão identificadas como choque, são: *Sent* (variável construída a partir de técnicas de análise de sentimento textual,

conforme descrito na Seção 1.3) e o Índice de Volatilidade S&P 500 do mercado de ações norte-americano (VIX).

Para testar os efeitos de choques do petróleo em variáveis macroeconômicas, utilizamos o método desenvolvido por Jordà (2005), que consiste da estimação de funções de resposta ao impulso a partir de projeções locais.⁶ Segundo aquele autor, esse método possui algumas vantagens em relação ao modelo convencional por vetores auto-regressivos, tais como: a) pode ser estimado via regressão linear por OLS; b) são mais robustos a erros de especificação; c) a realização de inferência analítica conjunta ou pontual é simples; e d) permite características fortemente não lineares e flexíveis que são complexas na forma multivariada. Semelhante a este estudo, Hamilton (2011) e Choi et al. (2018) testam os efeitos dos choques dos preços de petróleo em variáveis macroeconômicas.

Nesse estudo estimamos as impulso-respostas de duas maneiras: choque identificado como endógeno (Jordà (2005)) e choque identificado como exógeno (Ramey e Zubairy (2018)). A principal diferença entre ambos é que a estimação de forma exógena é considerada a partir de um vetor de exogeneidade, ou seja, um choque pré-estabelecido.⁷

De modo particular, a Equação 1.2 de forma reduzida é estimada para dados mensais⁸:

$$y_{t+h} = \alpha_h + \psi_h(L)x_{t-1} + \beta_h S_t + \epsilon_{t+h} \quad (1.2)$$

com $h=0, \dots, 10$ e em que y_t é a variável de interesse; α é uma constante; ψ é um polinômio de operador de defasagem; x representa um conjunto de variáveis defasadas utilizado como controle; β é a resposta de x no tempo $t+h$ dado o choque no tempo t ; S indica a variável de sentimento captada pelo mercado de petróleo; e ϵ um termo residual. Estimamos a Equação 1.2 de forma individual por OLS. Por fim, estimamos as impulso-respostas sem a correção dos erros-padrão de Newey e West (1986), proposta por Jordà (2005), já que Olea e Plagborg-Møller (2021) indicam que aquela correção não é necessária.

⁶ A identificação dos parâmetros é realizada através da decomposição de Cholesky.

⁷ A estimação por projeção local é realizada pelo software R, inspirada no estudo de Adämmer (2019).

⁸ Será incluído um coeficiente de tendência quadrática quando a variável de interesse for o PIB.

Em um comparativo com estimação por VAR, [Jordà \(2005\)](#) sugere que o método de projeção local é mais robusto na modelagem de erros de especificação quando comparado às IRF do VAR, já que a estimativa direta por projeção local é mais eficaz que o processo iterado por vetores auto-regressivos. A inferência para respostas ao impulso de VAR é complexa porque os coeficientes da IRF são funções não lineares de alta dimensão dos parâmetros estimados.

[Olea e Plagborg-Møller \(2021\)](#) indicam que a inferência por projeção local é robusta a duas características intrínsecas em aplicações macroeconômicas, a saber: i) dados persistentes e a estimativa de impulso-respostas para horizontes longos. Por sua vez, [Plagborg-Møller e Wolf \(2021\)](#) mostram que projeções locais lineares e VARs estimam as mesmas respostas ao impulso na população. Algumas implicações são consideradas: i) baseado em VAR a identificação estrutural, com inclusão de restrições de curto, longo prazo ou de sinal, pode ser executada de forma semelhante com uso de projeção local, sendo o contrário também válido; ii) a estimativa estrutural com uma *proxy* (instrumento) pode ser obtida a partir da ordenação do primeiro instrumento em um VAR recursivo, mesmo sob não-invertibilidade; iii) VARs lineares são tão robustos a não lineares quanto projeções locais lineares.

1.4.2 Banco de dados

A estimação da Equação 1.2 é baseada em dados com periodicidade mensal. A amostra de estudo inicia em janeiro de 2001 e termina em dezembro de 2020, com 240 observações disponíveis. O período analisado é justificado pela disponibilidade de dados, sobretudo para a construção da variável de choque.

A variável de sentimento utilizada para choque, construída a partir de técnicas de análise de sentimento textual, foi obtida através de relatórios mensais disponibilizados pela OPEP, conforme Seção 1.3. Além dessa variável, utilizamos três variáveis macroeconômicas de interesse para cada país analisado, nesse caso Brasil e Estados Unidos, são elas: produto da economia, inflação e taxa de juros. A escolha das variáveis de interesse foi baseada em [Bernanke et al. \(1997\)](#). Por fim, usamos o Índice de Volatilidade S&P 500 do mercado de

ações norte-americano (VIX) como variável a ser identificada como choque, como forma de comparar com os resultados obtidos a partir do choque identificado a partir da variável de sentimento construída. A Tabela A.1, do Apêndice A, apresenta um resumo das principais estatísticas descritivas das variáveis utilizadas no estudo.

Para os Estados Unidos, o conjunto de dados consiste em: índice da produção industrial ajustado sazonalmente como *proxy* para o produto interno bruto; taxa de juros efetiva ou fundos federais; e a taxa de inflação representada pela variação percentual do índice de preços ao consumidor para todos os consumidores urbanos. As variáveis foram obtidas no Banco Central de St. Louis.

Tabela 1.1 – Resumo das variáveis utilizadas

Variável	Descrição	Fonte
Estados Unidos		
Produção Industrial (proxy para o PIB)	Índice para a produção industrial ajustado sazonalmente	Banco Central de St. Louis
Juros	Taxa de juros efetiva ou fundos federais	Banco Central de St. Louis
Inflação	Índice de preços ao consumidor para todos os consumidores urbanos	Banco Central de St. Louis
Brasil		
PIB	Produto Interno Bruto acumulado em 12 meses	Banco Central do Brasil
Juros	Taxa de juros Selic anualizada	Banco Central do Brasil
Inflação	Índice de Preços ao Consumidor Amplo (IPCA) acumulado em 12 meses	Banco Central do Brasil
Variáveis de choque		
Sent	Variável construída a partir do sentimento do mercado de petróleo da OPEP	Autor
VIX	Índice de Volatilidade S&P 500 do mercado de ações norte-americano	Yahoo Finance

Fonte: Elaboração própria.

De modo semelhante, as variáveis para o Brasil, são: taxa de juros Selic anualizada; PIB acumulado em 12 meses; e a taxa de inflação representada pelo Índice Nacional de Preços ao Consumidor Amplo (IPCA) acumulado em 12 meses. Todas as variáveis foram obtidas no Sistema Gerenciador de Séries Temporais do banco central do Brasil. A Tabela 1.1 mostra um resumo das variáveis macroeconômicas utilizadas no estudo.

1.5 RESULTADOS E DISCUSSÃO

Exploramos o sistema de equações, do método proposto por Jordà (2005), a partir da análise das Funções de Respostas ao Impulso (doravante IRF's) para um choque de

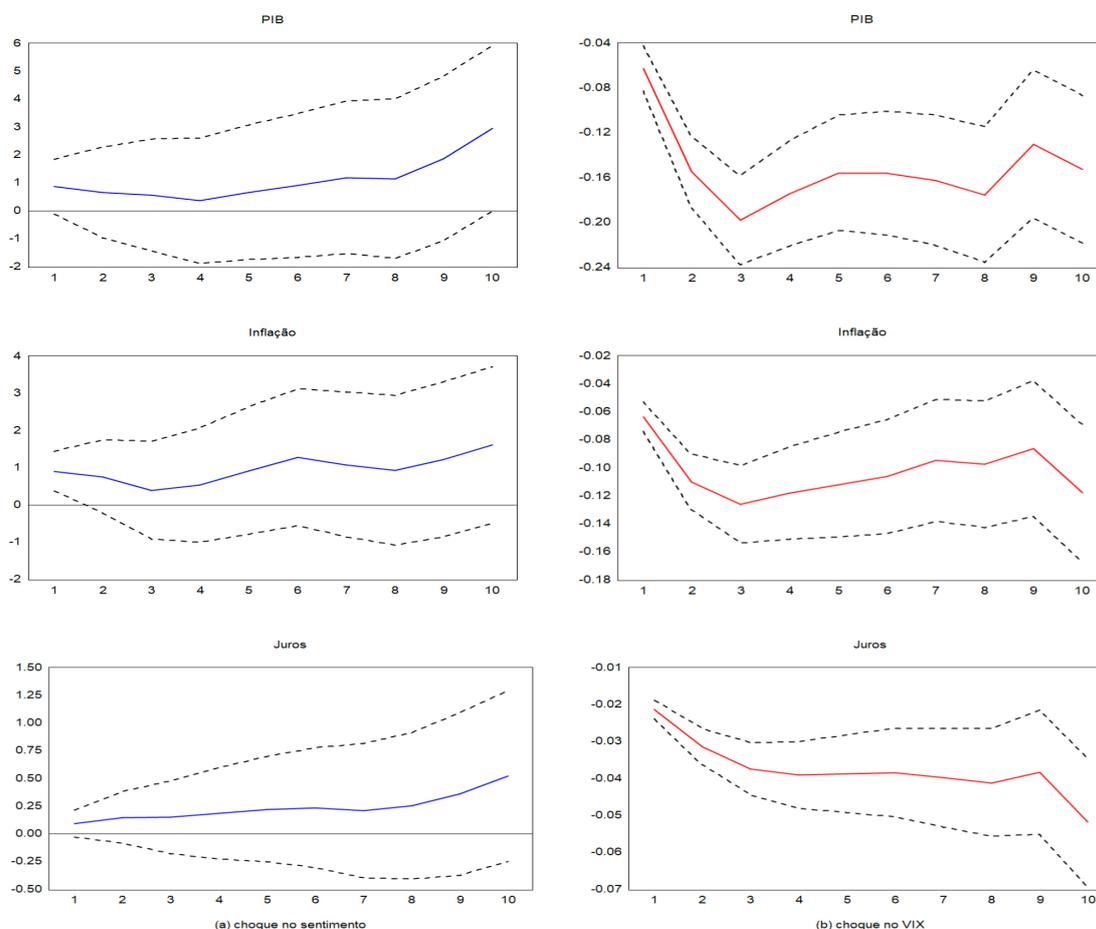
dois desvios-padrão nas variáveis de choques do petróleo, *Sent* e VIX. Uma IRF identifica o efeito de um choque residual único nos valores atuais e futuros das variáveis endógenas.

A obtenção da defasagem ótima das IRF's é escolhida com base em três critérios, são eles: Akaike (AIC), Hannan-Quin (HQ) e o bayesiano de Schwarz (1978). A escolha de defasagem ótima será aquela em que o critério apresenta o menor valor. Entretanto, quando houver divergência para defasagem entre os critérios, optamos pela escolha do menor *lag* ou quando a defasagem ótima for comum apenas a dois critérios.

A Figura 1.7 mostra as respostas das variáveis macroeconômicas (PIB, Inflação e Juros) a um choque de petróleo, para os Estados Unidos. Para estimação, consideramos o choque de petróleo como endógeno, sendo esse choque identificado pela variável de sentimento (*Sent*) e pelo índice VIX. A justificativa para isso é devido à alta representatividade do EUA na produção e consumo mundial de petróleo. Dado que identificação dos choques é pela decomposição de Cholesky, a estimação considera o seguinte ordenamento das variáveis: choque, PIB, Inflação e Juros.

A Figura 1.7 indica que um choque unitário na variável *Sent* provoca uma resposta positiva na inflação, no PIB e na taxa de juros, sendo estatisticamente significante nos períodos iniciais apenas para inflação. Este aumento da inflação pode ser justificado pelo aumento nos custos de transportes, já que grande parcela do petróleo é destinada para uso de transportes, o que provoca aumento de preços na cadeia produtiva. Destacamos que os Estados Unidos também é o maior consumidor de óleo bruto do mundo, o que encarece suas importações e, conseqüentemente, produtos derivados de petróleo. Este resultado é semelhante ao estudo de [Cavalcanti e Jalles \(2013\)](#), em que os autores utilizam choques nos próprios preços de petróleo. Por outro lado, constatamos que as variáveis macroeconômicas reagem de forma negativa, e estatisticamente significativa, a um choque de volatilidade na economia, identificado a partir da variável VIX.

Figura 1.7 – Respostas ao impulso em choques de petróleo - EUA

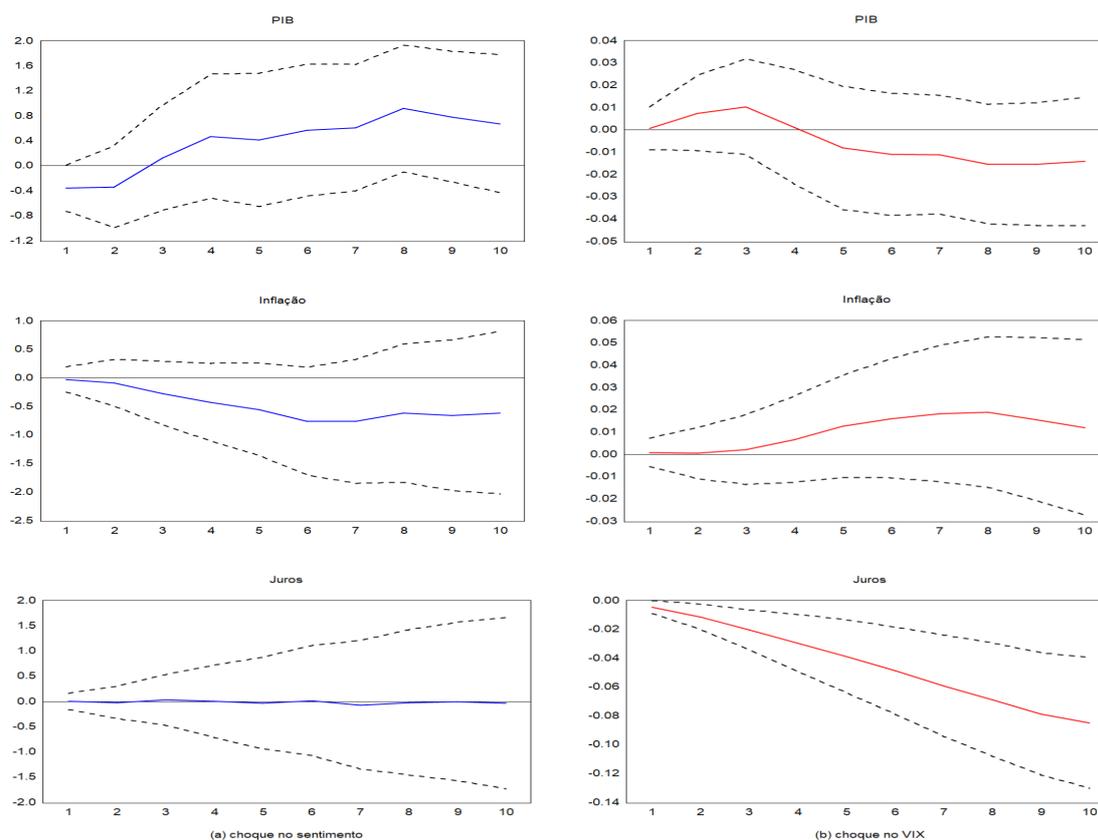


Fonte: Elaboração própria.

[3] Nota: A coluna do lado esquerdo (linha em azul) representa as respostas a um choque de petróleo, identificado a partir da variável de sentimento *Sent*. A coluna do lado direito (linha em vermelho) representa as respostas a um choque de petróleo, identificado a partir da variável VIX.
 [4] Nota: A defasagem ótima das variáveis endógenas utilizadas na estimação da coluna à esquerda está na Tabela A.2. A defasagem ótima das variáveis usadas para estimação das IRF's na coluna à direita é mostrada na Tabela A.3. As tabelas estão no Apêndice A.

A Figura 1.8 mostra as respostas das variáveis macroeconômicas do Brasil aos choques unitários do petróleo, mensurados pela variável de sentimento e o indicador de volatilidade VIX. Diferentemente da Figura 1.7, as IRF's da coluna esquerda (em azul) e da coluna direita (em vermelho) foram estimadas a partir de choques exógenos, visto que a variável de sentimento construída é baseada em relatórios da OPEP, e a literatura mostra que a economia brasileira tem pouco influência sobre o preço do petróleo mundial; e, por outro lado, a variável VIX é cotada na bolsa de Nova York.

Figura 1.8 – Respostas ao impulso em choques de petróleo - Brasil



Fonte: Elaboração própria.

[5] Nota: A coluna do lado esquerdo (linha em azul) representa as respostas a um choque de petróleo, captado pela variável de sentimento *Sent*. A coluna do lado direito (linha em vermelho) representa as respostas a um choque de petróleo, captado pelo índice VIX.

[6] Nota: A defasagem ótima da variável de choque utilizada nas IRF's na linha azul é mostrada na Tabela A.4. A defasagem ótima da variável de choque utilizada nas IRF's na linha vermelha é mostrada na Tabela A.5. As tabelas estão no Apêndice A.

[7] Nota: A defasagem das demais variáveis estimadas como endógenas está na Tabela A.6, do Apêndice A.

Com relação às respostas das variáveis macroeconômicas do Brasil, constatamos comportamento semelhante às respostas das variáveis dos Estados Unidos. Entretanto, verificamos um impacto menor para a economia brasileira, visto que a inflação não apresenta resposta estatisticamente significativa, ou seja, resultado incerto quanto ao choque unitário no sentimento do mercado de petróleo. Quanto à atividade econômica, aumentos no sentimento acerca dos preços de petróleo sinalizam uma queda no PIB real brasileiro, entretanto, a resposta sinaliza não ser estatisticamente diferente de zero.

Com exceção das respostas da taxa de juros ao choque exógeno na variável VIX,

os demais resultados foram estatisticamente não significativos para a economia brasileira, ou seja, incertos. Este resultado corrobora o estudo de [Cavalcanti e Jalles \(2013\)](#). Dessa forma, os resultados sinalizam que alterações no preço de petróleo mundial têm menor impacto na economia brasileira, comparado ao impacto na economia dos Estados Unidos.

1.6 RESULTADOS ADICIONAIS

Esta seção apresenta resultados adicionais, baseados em testes de robustez para respostas das variáveis macroeconômicas a diferentes abordagens. A primeira abordagem de robustez testada é a mudança do método de estimação das respostas ao impulso. Nesse caso, estimamos as respostas das variáveis dos EUA a partir de Vetores Auto-Regressivos, de forma estrutural (SVAR), em que consideramos todas as variáveis mensais endógenas, semelhante a estimação por projeção local mostrada na [Figura 1.7](#).

Representamos de forma resumida a metodologia SVAR utilizada como robustez. Consideramos o seguinte modelo de autoregressão vetorial de ordem p :

$$y_t = c + \sum_{i=1}^p \phi_i y_{t-1} + \epsilon_t \quad (1.3)$$

em que y_t corresponde a um vetor ($nx1$) de variáveis endógenas, $c = (c_1, \dots, c_n)'$ é um vetor de intercepto, ϕ_i é uma matriz (nxn) de coeficientes autoregressivos para $i = 1, 2, \dots, p$, e $\epsilon_t = (\epsilon_{1t}, \dots, \epsilon_{nt})'$ é a generalização (n) de um processo ruído branco. Utilizamos duas defasagens para as variáveis endógenas, conforme indicado pelos critérios de informação, mostrados na [Tabela A.7](#).

Utilizamos a decomposição de Choleski para identificação dos parâmetros do modelo SVAR. Para identificação de choque do petróleo, mensurado pela variável VIX, seguimos a literatura e consideramos que mudanças inesperadas no preço de petróleo são exógenas em relação aos valores das demais variáveis, ou seja, PIB, Inflação e Juros.

De acordo a [Figura A.1](#), constatamos que as respostas das variáveis macroeconômicas dos Estados Unidos mantém semelhança com as respostas estimadas na [Figura 1.7](#) (linha vermelha), por projeções locais, os resultados sendo significativos também para

todos os períodos.

Outro exercício de robustez consiste na agregação ou média das séries mensais para trimestrais, sendo a estimação por SVAR. Esse resultado está apresentado na Figura A.3, do Apêndice A. Para as variáveis VIX, PIB e Inflação, utilizamos a média para a transformação de mensal para trimestral, já que as variáveis estão em forma de índice. Por sua vez, acumulamos três meses para formação de um trimestre, para os Juros, já que seus valores mensais estão em termos absolutos. Conforme as impulso-respostas, verificamos que a mudança de frequência das variáveis não altera as trajetórias de resposta das variáveis macroeconômicas. Entretanto, destacamos que alguns períodos das respostas tornam-se não estatisticamente significativos.

Por fim, realizamos estimação por projeção local a partir dos dados utilizados em um SVAR, por [Cavalcanti e Jalles \(2013\)](#). Nesse caso, utilizamos os dados trimestrais da sub amostra referente ao primeiro trimestre de 1975 ao quarto trimestre de 1984, para as seguintes variáveis dos Estados Unidos: preço do petróleo internacional (WTI), PIB e taxa de inflação.

Mantemos o mesmo tratamento econométrico para as variáveis, antes da estimação, ou seja, consideramos todas as variáveis em logaritmo. Entretanto, não aplicamos a primeira diferença nas variáveis, e aplicamos apenas uma defasagem nas variáveis, e não quatro defasagens, devido ao número limitado de observações e a frequência trimestral. Os resultados indicam que a inflação responde de forma positiva a um choque unitário no sentimento de petróleo, sendo estatisticamente significativa até dez períodos à frente; por sua vez, o PIB responde também de forma positiva, entretanto, não possui significância estatística, conforme a Figura A.3. Portanto, com a redução de defasagem das variáveis e sem aplicação de primeira diferença, nossos resultados são mais robustos comparados ao estudo [Cavalcanti e Jalles \(2013\)](#), já que as respostas para inflação foram estatisticamente significativas.⁹

⁹ Mantemos o mesmo critério de identificação dos choques, por decomposição de Cholesky.

1.7 CONCLUSÕES

Em virtude da relevância do petróleo para a economia mundial, a investigação da reação de variáveis macroeconômicas a choques petrolíferos é fundamental para análise e tomadas de decisão por gestores públicos e privados. Neste estudo, buscamos avaliar as respostas da atividade econômica (PIB), inflação e juros, para dois países com economias distintas, porém, dependentes do petróleo, como Brasil e Estados Unidos.

A literatura de choques do petróleo em variáveis econômicas é bastante ampla. Entretanto, este estudo avança na literatura com desenvolvimento de um índice de sentimento do mercado de petróleo a partir de relatórios da OPEP, utilizado como *proxy* para identificação de choque. Aliado a essa inovação, propomos aplicação do método por projeções locais, como alternativa a outras metodologias bastante exploradas na literatura de choques, por exemplo, a metodologia de vetores auto-regressivos.

Nossos resultados sugerem: (i) alterações nos sentimentos do mercado de petróleo provocam impactos maiores na economia norte-americana, comparado aos efeitos na economia brasileira; (ii) em grande parte, os resultados foram não estatisticamente significativos para a economia brasileira, ou seja, incertos; (iii) as respostas das variáveis dos EUA e Brasil foram semelhantes quanto ao uso do indicador de sentimento ou o índice de volatilidade VIX; (iv) a análise a partir de choques por SVAR mantém similaridade com o método proposto no estudo, por Projeção Local; e, por fim, (v) a diminuição de frequência das séries temporais, juntamente com mudança de método, não altera as trajetórias de resposta das variáveis macroeconômicas.

Nossos resultados motivam pesquisas futuras que considerem mais variáveis de choques de petróleo, que podem ser desenvolvidas a partir de outras técnicas de análise textual, por exemplo, com uso de dicionário variante no tempo, com incorporação de palavras. Ademais, variáveis a serem identificadas como choques de petróleo podem ser construídas com base em relatórios de mercado futuro dessa *commodity*.

2 ESTRATÉGIAS DE PREVISÃO FISCAL: UM ESTUDO EMPÍRICO PARA A ECONOMIA BRASILEIRA

2.1 INTRODUÇÃO

O período de pandemia da Covid-19 trouxe um cenário de incerteza para a economia global, onde os países mais afetados são aqueles em processo de desenvolvimento. Por exemplo, o Brasil enfrenta dificuldades para retomada do crescimento pré-pandemia, acompanhada de elevados gastos governamentais com política de transferência de renda. Nesse sentido, a antecipação do comportamento de variáveis fiscais consiste em uma estratégia para formulação de políticas governamentais, com responsabilidade fiscal.

De modo geral, a literatura internacional para previsão sobre variáveis macroeconômicas é ampla, com estudos voltados para a atividade econômica, e também estratégias de previsão para variáveis fiscais. Com o uso de modelo econométrico univariado, [ARIMA](#), [Kamlet, Mowery e Su \(1987\)](#) comparam suas previsões de curto e longo prazos com as previsões elaboradas pelo Poder Executivo (OMB, sigla do órgão em inglês) e pelo Congresso (CBO, sigla em inglês) norte-americano, para taxa de crescimento real da economia, inflação e desemprego. Os autores indicam que para o período de um ano (curto prazo), as previsões são semelhantes. Por sua vez, [Plesko \(1988\)](#) avaliam as previsões do CBO e OMB para o PIB nominal, despesas correntes, receitas correntes e do déficit, e indicam resultados parecidos para as previsões de curto prazo. Enquanto, [Krause e Douglas \(2005\)](#) avaliam que as previsões do CBO e OMB para PIB real, inflação, desemprego, receitas do setor público, despesas fiscais e déficit orçamentário, de 1976 a 2001, são estatisticamente iguais, com exceção das variáveis desemprego e receitas. Outros estudos também comparam projeções macroeconômicas, por exemplo, [Frankel \(2011\)](#) e [Kliesen e Thornton \(2012\)](#)

Para a economia brasileira, alguns estudos realizam previsões para variáveis ma-

croeconômicas (por exemplo, Zuanazzi e Ziegelmann (2014), Caldeira, Moura e Santos (2015), Medeiros e Mendes (2016), Garcia, Medeiros e Vasconcelos (2017) e Barbosa, Ferreira e Silva (2020)), entretanto, abordagens para variáveis fiscais ainda são limitadas. Nesse contexto, Campos (2009) a partir de diversos modelos econométricos univariados e multivariados, desenvolve previsões para a arrecadação tributária brasileira, em virtude da importância dessa variável para construção do orçamento anual do governo brasileiro. O estudo indica a relevância do uso de métodos de séries temporais no desempenho da previsão da arrecadação tributária, quando comparados com método dos indicadores, até então utilizado pela Receita Federal Brasileira (RFB).

Além de poucos estudos sobre previsão de variáveis macroeconômicas brasileiras relacionadas à política fiscal, os métodos utilizados consideram poucas variáveis preditivas. Nesse sentido, este estudo adota e compara inúmeros algoritmos de aprendizado de máquina, em que estabelecemos o melhor modelo econométrico para previsão fiscal. Em um cenário macroeconômico diferente e para a economia norte-americana, Gu, Kelly e Xiu (2020) indicam que métodos de *machine learning*, juntamente com grande número de preditores, melhoram a acurácia de previsão. Por sua vez, Medeiros et al. (2021) indicam a superioridade de modelos da classe *machine learning* quando comparados com modelos univariados simples, por exemplo, passeio aleatório, arima, dentre outros.

Diante da importância do cenário fiscal após a pandemia de Covid-19 e escassez de estudos voltados para esse contexto, este estudo objetiva testar e comparar modelos de aprendizado de máquina na busca por um modelo de maior acurácia para previsão da principal variável orçamentária, a arrecadação tributária (preliminar na formação da responsabilidade fiscal do governo federal).¹

Este estudo contribui com a literatura de previsão fiscal em três sentidos: i) como modelos consolidados na literatura de previsão, estimamos e comparamos a acurácia de previsão dos modelos de aprendizado de máquina; ii) realizamos a previsão da arrecadação tributária mensal brasileira para um longo período amostral (2002 a 2021), algo jamais realizado para a economia brasileira; por fim, iii) aplicamos um exercício econométrico

¹ Ver Lei do Orçamento Anual (LOA) e Lei de Responsabilidade Fiscal (Lei Complementar nº 101, de 04/05/2000).

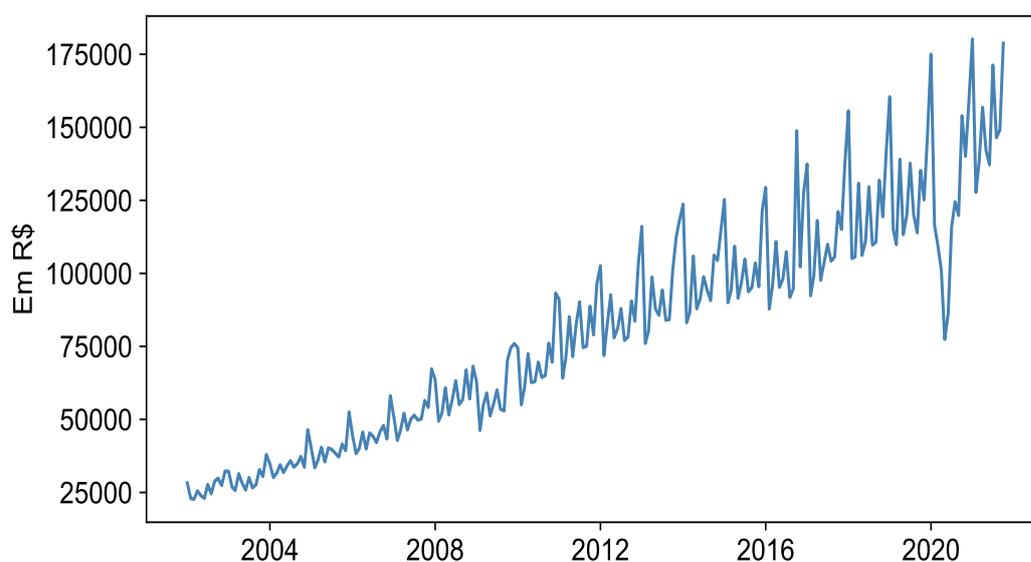
robusto para frequência trimestral, essencial para tomadas de decisão de gestores públicos. Portanto, essa pesquisa auxilia na formulação de políticas governamentais para antecipação e controle do orçamento público.

Além desta introdução, este estudo está estruturado em mais cinco seções. A Seção 2.2 é destinada para apresentação do conjunto de dados utilizados na previsão da variável fiscal. A Seção 2.3 apresenta a metodologia executada. A Seção 2.4 é destinada para os resultados e discussão da pesquisa. O exercício de robustez está na Seção 2.5. Por fim, a Seção 2.6 contém as considerações finais do estudo.

2.2 BANCO DE DADOS

Esta seção é reservada para apresentação do conjunto de dados utilizados para previsão da variável fiscal de interesse, arrecadação tributária, obtida na Receita Federal do Brasil. O período amostral utilizado na previsão varia de janeiro de 2002 a outubro de 2021, com frequência mensal, o que totaliza 238 observações. O período amostral é justificado pela limitação de observações para todas as variáveis usadas na estimação, ou seja, a arrecadação e as variáveis preditoras. A Figura 2.1 ilustra a trajetória da arrecadação tributária ao longo do tempo.

Figura 2.1 – Trajetória da arrecadação fiscal (Em Milhões de R\$)



Fonte: Elaboração própria.

De acordo com a trajetória da arrecadação, observamos uma forte oscilação de seus valores ao longo dos anos. Entretanto, em meados de 2020 houve uma queda brusca na arrecadação brasileira, o que pode ser justificado pelo início da pandemia de Covid-19. Em virtude da escolha de variáveis preditivas para a arrecadação, optamos por modelos de previsão de aprendizado de máquina, o que requer um grande número de preditores.

Para este estudo, além da variável de interesse (arrecadação), utilizamos 34 variáveis econômicas que possivelmente explicam o comportamento da arrecadação tributária. Em virtude do grande número de preditores, a técnica de *machine learning* sobressai na escolha, já que possui algoritmos de penalização com intuito de seleção de variáveis preditoras.

Baseado na técnica de [McCracken e Ng \(2016\)](#), classificamos as variáveis explicativas em cinco grupos, são eles: atividade econômica, finanças públicas, indicador de crédito, mercado financeiro e de capitais e setor externo.

Para o grupo de Atividade Econômica, foram utilizadas 21 variáveis. As variáveis papelão ondulado, produção industrial (produtos de borracha e material plástico) e produção industrial (celulose, papel e produtos de papel) foram extraídas do Instituto Brasileiro de Geografia e Estatística (IBGE). As demais 18 variáveis ligadas à atividade econômica foram coletadas no Banco Central do Brasil, são elas: produção geral, produção (indústria de transformação), produção de bens de capital, produção de bens intermediários, produção de bens de consumo, produção total de autoveículos, vendas totais de autoveículos, consumo de energia elétrica (comercial), consumo de energia elétrica (industrial), consumo de energia elétrica (total), volume de vendas no varejo (automóveis, motocicletas, partes e peças), Índice Geral de Preços do Mercado (IGP-M), Índice Geral de Preços - disponibilidades internas (IPA-M), Índice de Preços por Atacado-Mercado (IPA-M), Índice Nacional de Preços ao Consumidor (INPC), Índice Nacional de Preços ao Consumidor Amplo, itens livres, produção de caminhões e índice de volume de vendas no varejo.

Tabela 2.1 – Resumo das variáveis utilizadas na predição de arrecadação

Grupo	Fonte	Código
Atividade econômica		
Papelão ondulado	IBGE	-
Produção industrial (produtos de borracha e material plástico)	IBGE	-
Produção industrial (celulose, papel e produtos de papel)	IBGE	-
Produção (geral)	Banco Central	21859
Produção (indústria de transformação)	Banco Central	21862
Produção (bens de capital)	Banco Central	21863
Produção (bens intermediários)	Banco Central	21864
Produção (bens de consumo)	Banco Central	21865
Produção total de autoveículos	Banco Central	1373
Vendas de autoveículos (total)	Banco Central	1378
Consumo de energia elétrica (comercial)	Banco Central	1402
Consumo de energia elétrica (industrial)	Banco Central	1404
Consumo de energia elétrica (total)	Banco Central	1406
Volume vendas no varejo (automóveis, motocicletas, partes e peças)	Banco Central	1548
Índice Geral de Preços do Mercado (IGP-M) - Var. % mensal	Banco Central	189
Índice Geral de Preços - disponibilidades Interna (IPA-M) - Variação % mensal	Banco Central	190
Índice de Preços por Atacado-Mercado (IPA-M) - Variação % mensal	Banco Central	7450
Índice Nacional de Preços ao Consumidor - Amplo (INPC) - Itens livre - Var % mensal	Banco Central	188
Índice Nacional de Preços ao Consumidor Amplo (IPCA) - Itens livres - Variação % mensal	Banco Central	11428
Produção de caminhões	Banco Central	1375
Índice Volume de Vendas no Varejo	Banco Central	1455
Finanças públicas		
Contribuição - Finsocial / Cofins - Total - Receita bruta	Ministério da Economia \SRF	-
Imposto sobre a Renda - Pessoas Físicas	Ministério da Economia \SRF	-
Imposto sobre a Renda - Pessoas Jurídicas	Ministério da Economia \SRF	-
Dívida Líquida do Setor Público (% PIB) - Interna - Setor Público Consolidado	Banco Central	4524
Dívida Líquida do Setor Público (% PIB) - Total - Setor Público Consolidado	Banco Central	4513
Dívida Líquida do Setor Público (% PIB) - Externa - Setor Público Consolidado	Banco Central	4535
Dívida Líquida do Governo Geral (% PIB)	Banco Central	4536
Dívida Fiscal Líquida - Saldos (% PIB)	Banco Central	10831
NFSP sem desvalorização cambial (% PIB)	Banco Central	5717
Dívida mobiliária (saldos)	Banco Central	2213
Indicador de crédito		
Saldo da carteira de crédito - Total	Banco Central	20539
Mercado financeiro e de capitais		
Taxa de juros - Selic	Banco Central	4390
Setor Externo		
Taxa de câmbio	Banco Central	3697

Fonte: Elaboração própria.

O grupo Finanças Públicas é composto por dez indicadores. As variáveis Contribuição - Finsocial/Cofins, imposto sobre a renda (pessoas físicas) e imposto sobre a

renda (pessoas jurídicas) foram extraídas do Ministério da Economia/Secretária da Receita Federal. As demais sete variáveis foram extraídas do Banco Central, são elas: Dívida Líquida do Setor Público (DLSP) interna do setor público consolidado, DLSP total do setor público consolidado, DSLP externa do setor público consolidado, dívida líquida do governo geral, saldos da dívida fiscal líquida, Necessidade de Financiamento do Setor Público (NFSP) sem desvalorização cambial e dívida imobiliária.

As três variáveis restantes foram coletadas no banco central e são subdivididas em três grupos restantes. O grupo Indicador de Crédito considera a variável saldo da carteira de crédito total; o grupo Mercado Financeiro e de Capitais inclui a taxa de juros básica da economia brasileira (SELIC); por fim, o grupo Setor Externo considera a taxa de câmbio comercial (compra). A Tabela 2.1 resume o conjunto de variáveis usadas como possíveis preditoras da arrecadação tributária. As estatísticas descritivas das variáveis preditoras e da arrecadação (variável de interesse) estão descritas na Tabela B.1, do Apêndice B.

2.3 METODOLOGIA E MODELOS DE APRENDIZADO DE MÁQUINA

Esta seção é reservada para apresentação da metodologia utilizada para previsão de arrecadação tributária, bem como são apresentadas especificações de nove modelos de *machine learning* utilizados.

2.3.1 Metodologia

Para estimação dos inúmeros modelos de aprendizado de máquina, seguimos o seguinte modelo:

$$T_{t+h} = G_h(x_t) + u_{t+h}, \quad h = 1, \dots, H, \quad t = 1, \dots, T \quad (2.1)$$

em que T_{t+h} é a arrecadação no mês $t + h$; $x_t = (x_{1t}, \dots, x_{nt})'$ é um vetor n de co-variáveis que indicam prováveis defasagens de T_t e/ou fatores comuns, bem como um conjunto de preditores candidatos; $G(\cdot)$ é o mapeamento entre co-variáveis e arrecadação futura; e u_t é

um erro aleatório de média zero. A função alvo $G_h(\cdot)$ pode ser um único modelo ou um conjunto de diferentes especificações. Os mapeamentos são diferentes conforme varia o horizonte de previsão.

A equação de previsão objetiva é dada por:

$$\hat{T}_{t+h|t} = \hat{G}_{h,t-R_h+1:t}(x_t) \quad (2.2)$$

em que $\hat{G}_{h,t-R_h+1:t}$ é uma função objetivo estimada baseada em uma série temporal $t - R_h + 1$ até t e R_h é o tamanho da janela, que varia conforme o horizonte de previsão e o número de variáveis defasadas no modelo. Seguindo [Medeiros et al. \(2021\)](#), consideramos as previsões diretas já que não realizamos nenhuma tentativa de previsão das co-variáveis.

As previsões realizadas para a variável de interesse, arrecadação, será baseada em uma estrutura de janela móvel com tamanho determinado. Em outras palavras, para o período amostral de janeiro de 2002 até outubro de 2021, o número de observações é $R_h = 238 - h - p - 1$, em que p é o número de defasagens do modelo. A escolha de previsão por janela móvel é justificada a partir de dois pontos: i) diminuir os efeitos de possíveis quebras estruturais e *outliers*; e ii) evitar problemas de execução de testes de desempenho preditivos superiores entre modelos econométricos aninhados.²

Dentre os modelos de aprendizado de máquina, consideramos os métodos de *shrinkage* que estabelecem uma seleção de variáveis com objetivo de elevar a performance preditiva da variável alvo ou de interesse. O processo de estimação envolve a penalização de estimadores (preditores) irrelevantes como penalização. É comum os coeficientes estimados dos preditores (variáveis explicativas) irrelevantes serem reduzidos na previsão, em que podem tender a zero.

A partir de redução dos valores absolutos dos parâmetros, os modelos *shrinkage* selecionam variáveis com maior capacidade de predição da variável alvo, mecanismo conhecido como regularização. De modo geral, métodos de *shrinkage* executam o processo de penalização abaixo:

² Ver [Giacomini e White \(2006\)](#).

$$\min_b \|y_t - z'_t b\|^2$$

sujeito a $g(b) \leq s$.

em que $\|y_t - z'_t b\|^2$ representa a soma do quadrado dos resíduos, s denota um parâmetro de ajuste de penalização e $g(b)$ é uma função penalizadora dos coeficientes b da expressão.

Neste estudo, utilizamos cinco estimadores *shrinkage*, são eles: Lasso, Elastic-Net, Ridge, Lasso Adaptativo (LassoAdap) e Elastic-Net Adaptativo (LassoAdap). Também utilizamos modelos que envolve métodos de média em conjunto, em que optamos por Bagging e CSR. Além desses modelos, também usamos dois métodos baseados na extração por Fatores, nesse caso optamos pelo Fatores *Target*, baseado em Bai e Ng (2008), e o Boosting, baseado em Bai e Ng (2009). Além dos modelos destacados, também analisamos dois modelos alternativos: modelo com média simples e outro baseado na mediana, de todos os modelos de aprendizado estatísticos utilizados na predição da variável alvo. Os horizontes de previsões para todos os modelos especificados são para oito períodos à frente, previsão acumulada de quatro períodos à frente e, por fim, previsão acumulada com oito períodos à frente.

2.3.2 Shrinkage

Para estimadores *shrinkage* de modelos lineares executados neste estudo temos que:

$$G_h(x_t) = \beta'_h x_t \tag{2.3}$$

$$\hat{\beta}_h = \arg \min_{\beta_h} \left[\sum_{t=1}^{T-h} (y_{t+h} - \beta'_h x_t)^2 + \sum_{i=1}^n p(\beta_{h,i}; \lambda, \omega_i) \right]$$

em que $p(\beta_{h,i}; \lambda, \omega_i)$ é uma função de penalidade que depende do parâmetro de penalidade λ e em um peso $\omega_i > 0$. Consideramos tipos de escolhas diferentes para a função de penalidade.

2.3.2.1 Ridge

A Regressão Ridge foi proposta por Hoerl e Kennard (1970). A função de penalidade é dada por:

$$\sum_{i=1}^n p(\beta_{h,i}; \lambda, \omega_i) := \lambda \sum_{i=1}^n \beta_{h,i}^2 \quad (2.4)$$

A principal vantagem dessa regressão é a facilidade de cálculo de sua solução analítica, bem como reduz coeficientes ligados às variáveis menos relevantes na predição, para quase zero. Entretanto, os coeficientes quase nunca atinge zero para qualquer tamanho de λ .

2.3.2.2 LASSO

Este método foi proposto por [Tibshirani \(1996\)](#), em que a função de penalidade é dada por:

$$\sum_{i=1}^n p(\beta_{h,i}; \lambda, \omega_i) := \lambda \sum_{i=1}^n |\beta_{h,i}| \quad (2.5)$$

Este operador reduz as variáveis operadoras para zero, eliminando-as da estimação quando forem irrelevantes. Todavia, a consistência de seleção do modelo é obtida apenas sob hipóteses rigorosas. Quando seleciona as variáveis, o LASSO minimiza a soma dos quadrados dos resíduos, desde que a soma do valor absoluto dos coeficientes seja menor que uma constante. Além disso, com $\lambda = 0$ a estimação será por mínimos quadrados ordinários (OLS).

Conforme [Zou e Hastie \(2005\)](#), apesar de renomado para estudos de previsão, o LASSO apresenta algumas limitações em algumas situações. Esses autores listam alguns exemplos: i) quando o número de preditores é superior ao número de observações, aquele operador seleciona apenas a quantidade de preditores equivalente ao número de observações, em virtude do problema de otimização convexa inerente ao método; no caso de efeito de agrupamento, o LASSO quase sempre seleciona apenas uma variável por grupo; e, por fim, quando o número de observações é superior ao número de preditores e na presença de fortes correlações entre os preditores, é esperado que a regressão Ridge tenha performance superior ao LASSO.

2.3.2.3 LASSO Adaptativo

Este método foi proposto por [Zou \(2006\)](#) para melhora de desempenho do modelo LASSO convencional. Este método adaptativo utiliza a mesma penalidade do LASSO, com a inclusão de um parâmetro de ponderação. Agora, a função de penalidade é dada por:

$$\sum_{i=1}^n p(\beta_{h,i}; \lambda, \omega_i) := \lambda \sum_{i=1}^n \omega |\beta_{h,i}| \quad (2.6)$$

em que $\omega = |\beta_{h,i}^*|^{-1}$ e $\beta_{h,i}^*$ é o coeficiente do primeiro passo da estimação LASSO. O LASSO Adaptativo suporta mais variáveis do que observações e possui performance alta em processos não gaussianos e sob heteroscedasticidade.³

2.3.2.4 Elastic Net

Este método foi proposto por [Zou e Hastie \(2005\)](#). Elastic Net é uma generalização que incluem casos especiais como LASSO e Ridge. A função de penalidade é definida por:

$$\sum_{i=1}^n p(\beta_{h,i}; \lambda, \omega_i) := \alpha \lambda \sum_{i=1}^n \beta_{h,i}^2 + (1 - \alpha) \lambda \sum_{i=1}^n |\beta_{h,i}| \quad (2.7)$$

em que $\alpha \in [0, 1]$. Semelhante ao LASSO, Elastic Net realiza de forma simultânea seleção automática e encolhimento contínuo, em que é possível a seleção de grupos de variáveis correlacionadas.

Elastic Net é uma combinação convexa das normas ℓ_1 e ℓ_2 . Em outras palavras, esse método é vantajoso porque incorpora dois tipos de penalidade específicas: penalidade normada de ordem 2, estabelecida por $\sum_{i=1}^n \beta_{h,i}^2$, e penalidade normada de ordem 1, estabelecida por $\sum_{i=1}^n |\beta_{h,i}|$. No primeiro caso, se $\alpha = 1$, o método se reduz ao Ridge. Por outro lado, no segundo caso, com $\alpha = 0$, temos como resultado o operador proposto por [Tibshirani \(1996\)](#), ou seja, o LASSO.

A obtenção dos valores de α é através de validação cruzada. Entretanto, este método não é indicado para séries temporais em virtude da correlação entre as variáveis preditoras. Nesse sentido, [Bai e Ng \(2008\)](#) sugerem a utilização de critérios de escolha alternativos,

³ Ver [Medeiros e Mendes \(2016\)](#).

como Critério de Informação de Akaike (AIC) e Critério Bayesiano de Schwarz (BIC), para obtenção dos parâmetros de ajuste. Nesse estudo, optamos pelo critério BIC.

Seguindo Kock e Callot (2015) e Medeiros e Mendes (2016), para todos métodos da classe LASSO, o parâmetro de penalização λ é escolhido a partir do critério BIC. O parâmetro α para a penalização Elastic Net é definido em 0,5. Os pesos das extensões adaptativas de LASSO e Elastic Net são definidas por $\omega_i = \frac{1}{|\tilde{\beta}_i| + \frac{1}{\sqrt{T}}}$, em que $\tilde{\beta}_i$ é o estimador do método adaptativo.

2.3.3 Métodos de ponderação (*ensemble*)

As previsões de conjunto são resultados de uma ponderação das previsões de vetor de métodos.

2.3.3.1 Bagging

O método Bagging foi proposto por Breiman (1996). Esse método combina previsões de um conjunto de modelos instáveis estimados para diferentes amostras de *bootstrap*, ou seja, consiste de um método de agregação através de re-amostragem. Quase sempre a combinação é melhorada quando os modelos envolvidos são muito diferentes. As etapas do Bagging são:

Passo 1: para cada amostra do *bootstrap*, executa uma regressão OLS com as variáveis candidatas e seleciona aquelas com estatística t absoluta superior a um determinado limite c .

Passo 2: estima uma nova regressão apenas com variáveis selecionadas no passo 1.

Passo 3: os coeficientes da segunda regressão (passo 2) são usados para computar as previsões da amostra atual.

Passo 4: as três primeiras etapas para amostras de *bootstrap* B são repetidas e a previsão final como a média das previsões de B é calculada.

Em termos matemáticos, temos:

$$y_{t+h}^{bagg} = W_t \hat{\beta}^{OLS} + \sum_{j=1}^r \Gamma(t_j) \hat{\beta}_{F_j} \hat{F}_j \quad (2.8)$$

em que $\hat{\beta}^{OLS}$ é a estimação de OLS de y_{t+h} contra W_t ; $\hat{\beta}_{F_j}$ é a regressão OLS residual; $j = 1 \dots r$ estabelece a quantidade de fatores estimados; e t_j é a estatística *t-student* associada a $\hat{\beta}_{F_j}$.

$$\Gamma(t_j) = 1 - \psi(t+c) + t - 1[\phi(t-c) - \phi(t+c)] \quad (2.9)$$

em que ψ é a densidade normal padrão e ϕ representa uma função de distribuição acumulada da distribuição normal.

O número de replicações para o método Bagging é $B = 100$. O procedimento de pré-teste adotado é baseado em [Inoue e Kilian \(2008\)](#), em que o nível estabelecido é de 5%.

2.3.3.2 CSR

O método de Regressões de Subconjunto Completadas (CSR) foi proposto por [Elliott, Gargano e Timmermann \(2013\)](#) e [Elliott, Gargano e Timmermann \(2015\)](#). A seleção de subconjunto ótimo de x_t para previsão de T_{t+h} com testagem de todas as combinações de regressores é computacionalmente muito exigente e, quase sempre, inviável. Suponha que temos n variáveis candidatas. O objetivo é a seleção de um número $q \leq n$ variáveis e execução de regressões com uso de todas as combinações possíveis de q de n variáveis.

O CSR suporta um número pequeno de variáveis. Para grandes conjuntos de variáveis, o número de regressões a serem estimadas aumenta rápido. Dessa forma, iniciamos a regressão A_{t+h} em cada uma das possíveis variáveis candidatas (incluindo defasagens) e guardamos as estatísticas t de cada variável. As estatísticas t são classificadas por valor absoluto, e selecionamos as \tilde{n} variáveis mais relevantes na classificação. A previsão CSR é calculada sobre essas variáveis. Semelhante ao teste Bagging, o pré-teste executado é ao nível de 5%.

2.3.4 Modelos Fatoriais

A literatura sobre modelos fatoriais é bem representada no estudo de Bai (2003). O modelo fatorial reduz a dimensão do modelo completo a partir de extração de componentes comuns de todos preditores. Os fatores são calculados como componentes principais de um grande conjunto de variáveis z_t tal que $F_t = Az_t$, em que A é uma matriz de rotação e F_t representa o vetor de componentes principais. De acordo com a Equação 2.1, temos que x_t é dado por T_{t+h} , $j = 0, 1, 2, 3$ mais f_{t-j} , $j = 0, 1, 2, 3$, em que f_t é o vetor com os quatro primeiros componentes principais de z_t .

Para esse estudo, o número de fatores e defasagens dos modelos fatoriais é definido em quatro. Para os fatores *target*, o nível de significância é de 5% ($\alpha = 0.05$). O número máximo de iterações estabelecido é $M = 10 \times$ número de variáveis. Entretanto, o algoritmo será parado quando o critério de escolha BIC aumenta. Seguindo Bai e Ng (2008) e Medeiros et al. (2021), o parâmetro *shrinkage* é definido por $v = 0, 2$.

2.3.4.1 Fatores Target

Bai e Ng (2008) sugerem este procedimento para aumento da performance das previsões por fatores. A proposta dessa metodologia por fatores direcionados é o cálculo dos componentes principais somente das variáveis com alto poder de predição da arrecadação futura. Seja $z_{i,t}$, $i = 1, \dots, q$ variáveis candidatas e w_t um conjunto de controles. Conforme Bai e Ng (2008) e Medeiros et al. (2021), usamos os valores defasados da variável de interesse, nesse caso a arrecadação, como controles. O procedimento ocorre em quatro etapas:

Passo 1: para $i = 1, \dots, q$, regrida T_{t+h} em w_t e $z_{i,t}$ e calcule as estatísticas t para os coeficientes correspondentes em $z_{i,t}$.

Passo 2: escolha um nível de significância α e selecione as variáveis que são significativas com uso das estatísticas t calculadas.

Passo 3: deixe $z_t(\alpha)$ ser as variáveis selecionadas nos passos 1 e 2. Estime os fatores F_t de $z_t(\alpha)$ pelos componentes principais.

Passo 4: regrida T_{t+h} em w_t e $f_{t-j}, j = 0, 1, 2, 3$, em que $f_t \subset F_t$. Para esse caso, o número de fatores em f_t é selecionado a partir do critério de informação de escolha BIC.

2.3.4.2 Fatores Boosting

Seguindo Bai e Ng (2008), deixe $z_t \in \mathbb{R}$, o conjunto dos n fatores calculados a partir das n variáveis originais mais quatro defasagens de cada fator. Dessa forma, $q = 5n$. O algoritmo é definido a partir de três passos, a saber:

Passo 1: deixe $\Phi_{t,0} = \bar{T}$ para cada t , em que $\bar{T} = \frac{1}{t} \sum_{i=1}^t T_i$.

Passo 2: para $m = 1, \dots, M$: a) calcule $\hat{u}_t = T_t - \Phi_{t-h,m-1}$; b) para cada variável candidata $i = 1, \dots, q$, regrida o resíduo atual em $z_{i,t}$ para obter \hat{b}_i , e compute $\hat{e}_{t,i} = \hat{u}_t - z_{i,t} \hat{b}_i$, e calcule $SSR_i = \hat{e}'_i \hat{e}_i$; c) Selecione i_m^* como índice da variável que fornece menor SSR e defina $\hat{\phi}_{t,m} = \hat{\phi}_{t,m-1} + v \phi_{t,m}$, em que v é o tamanho da etapa.

Passo 3: pare o algoritmo após a iteração M th ou quando o critério BIC começa aumentar.

A Tabela 2.2 exibe um resumo dos modelos de aprendizado de máquina utilizados para predição da arrecadação fiscal brasileira.

Tabela 2.2 – Modelos de Previsão

Modelo	Descrição	Categoria de aprendizagem
Ridge	Regressão Ridge	Shrinkage
Lasso	Método LASSO	Shrinkage
LassoAdap	Método LASSO Adaptativo	Shrinkage
ElasticNet	Método Elastic Net	Shrinkage
ElasticNetAdap	Método Elastic Net Adaptativo	Shrinkage
Bagging	Método de ponderação Bagging	Ponderação
CSR	Regressões de Subconjunto Completadas	Ponderação
Target	Modelo de fatores Target	Fatorial
Boosting	Modelo de fatores Boosting	Fatorial

Fonte: Elaboração própria.

2.3.5 Acurácia dos modelos

O desempenho dos modelos será mensurado com base na Raiz do Erro Quadrático Médio (RMSE), fora da amostra. O RMSE é expressado pela Equação 2.10:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - y'_i)^2} \quad (2.10)$$

em que y' denota o valor predito, y denota o valor efetivo e n representa o número de observações.

2.4 RESULTADOS E DISCUSSÃO

A periodicidade dos dados utilizados para previsão da arrecadação tributária varia de janeiro de 2002 a outubro de 2021, com frequência mensal. A Tabela 2.3 mostra as previsões da arrecadação para diferentes horizontes. Nesse caso, as previsões são para oito períodos à frente, para um período acumulado de quatro meses à frente e, por fim, para um período acumulado de oito meses à frente. A amostra de previsão ou de teste corresponde a 20% da amostra completa, ou seja, 80% da amostra foi destinada ao teste na estimação. Estratégia de previsão baseada em [Medeiros et al. \(2021\)](#).

Além das previsões pelos métodos de aprendizado estatístico, já detalhados na Seção 2.3, também realizamos dois tipos de previsão alternativos: o primeiro é realizado a partir da média das previsões de todos os modelos de aprendizado estatístico, para cada horizonte de previsão considerado, incluindo os horizontes cumulativos; por sua vez, o segundo tipo de previsão considerado é através da mediana de todos os modelos.

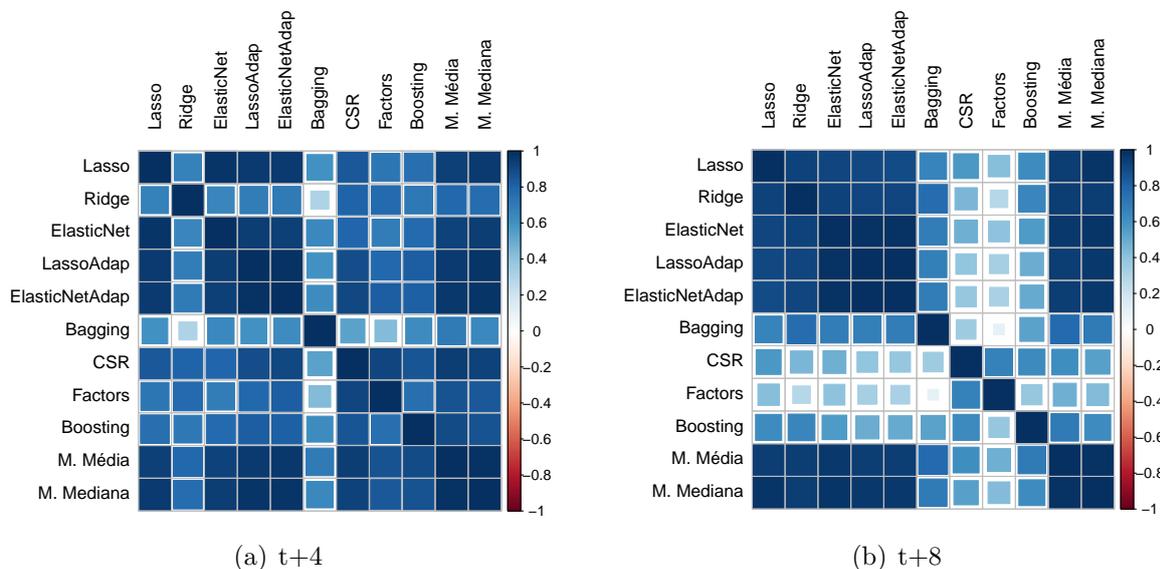
Tabela 2.3 – RMSE para fora da amostra - Previsão de arrecadação - Novembro/2017 a outubro/2021

Modelo	Horizonte									
	1	2	3	4	5	6	7	8	4m	8m
Lasso	0.1393	0.1400	0.1350	0.1444	0.1592	0.1744	0.1694	0.1903	0.1099	0.1147
Ridge	0.1264	0.1274	0.1536	0.1683	0.1743	0.1812	0.1561	0.1523	0.1281	0.1251
ElasticNet	0.1089	0.1176	0.1241	0.1291	0.1584	0.1514	0.1560	0.1753	0.1060	0.1106
Target	0.1433	0.1826	0.1717	0.1921	0.1915	0.1928	0.1955	0.1963	0.1461	0.1533
Baggins	0.1749	0.1817	0.1734	0.1484	0.1984	0.1963	0.1957	0.2162	0.1302	0.1262
Boosting	0.1448	0.1472	0.1535	0.1498	0.1583	0.1564	0.1623	0.1587	0.1382	0.1358
ElasticNetAdap	0.1264	0.1354	0.1347	0.1465	0.1600	0.1632	0.1622	0.1914	0.1182	0.1188
LassoAdap	0.1393	0.1400	0.1350	0.1444	0.1592	0.1744	0.1694	0.1903	0.1186	0.1189
CSR	0.1161	0.1364	0.1410	0.1474	0.1521	0.1486	0.1427	0.1414	0.1187	0.1203
M.Média	0.1127	0.1213	0.1277	0.1347	0.1486	0.1506	0.1398	0.1482	0.1110	0.1162
M.Mediana	0.1155	0.1236	0.1319	0.1383	0.1522	0.1498	0.1499	0.1573	0.1137	0.1166

Fonte: Elaboração própria.

[1] Nota: Para essa previsão, utilizamos 80% da amostra para treino e 20% da amostra para teste, ou seja, o período usado para previsão, novembro de 2017 a outubro de 2021.

Figura 2.2 – Correlação das previsões para Arrecadação de novembro de 2017 a outubro de 2021



Fonte: Elaboração própria.

[2] Nota: A figura (a) indica as correlações entre as previsões de cada modelo para o quarto período à frente, ou seja, o quarto mês. A figura (b) indica as correlações entre as previsões para o oitavo período à frente.

[3] Nota: Essas correlações foram construídas considerando 80% da amostra de treino e 20% para amostra de teste.

De acordo com a Tabela 2.3, o modelo Elastic Net possui a melhor a acurácia comparado aos demais modelos de aprendizado, para os quatro primeiros horizontes de previsão, com destaque também para o modelo CSR, com alta acurácia para as janelas de previsão iniciais. Garcia, Medeiros e Vasconcelos (2017) também destacam o forte poder de previsão do método de ponderação CSR, no caso para variável macroeconômica brasileira. Para os períodos cumulativos, o modelo Elastic Net também supera os demais modelos. Com relação aos modelos alternativos, constatamos que o modelo médio supera o modelo que leva em consideração sua mediana, para todos os horizontes de previsão, sendo muito próximos para o período cumulativo de oito períodos à frente.

A Figura 2.2 mostra as correlações entre as raízes do erro quadrático médio das previsões dos modelos utilizados, para quatro e oito períodos à frente. Conforme esperado, constatamos que todas as correlações são positivas, isso reforça a capacidade de previsão semelhante entre os modelos, apesar de categorias de aprendizagem distintas (shrinkage, ponderação e fatorial). Também observamos forte correlação do melhor modelo Elastic

Net com outros destaques, como CSR, M. Média e M. Mediana. O modelo com menor correlação com os demais é o Bagging, o que reforça a baixa performance daquele modelo de ponderação mostrada na Tabela 2.3. Uma possível explicação para esse resultado ruim do Bagging é sua baixa capacidade de predição para um conjunto de variáveis e número de observações não tão grande, baseada em sua metodologia de estimação.

Tabela 2.4 – RMSE para fora da amostra - Previsão de arrecadação - Dezembro/2015 a outubro/2021

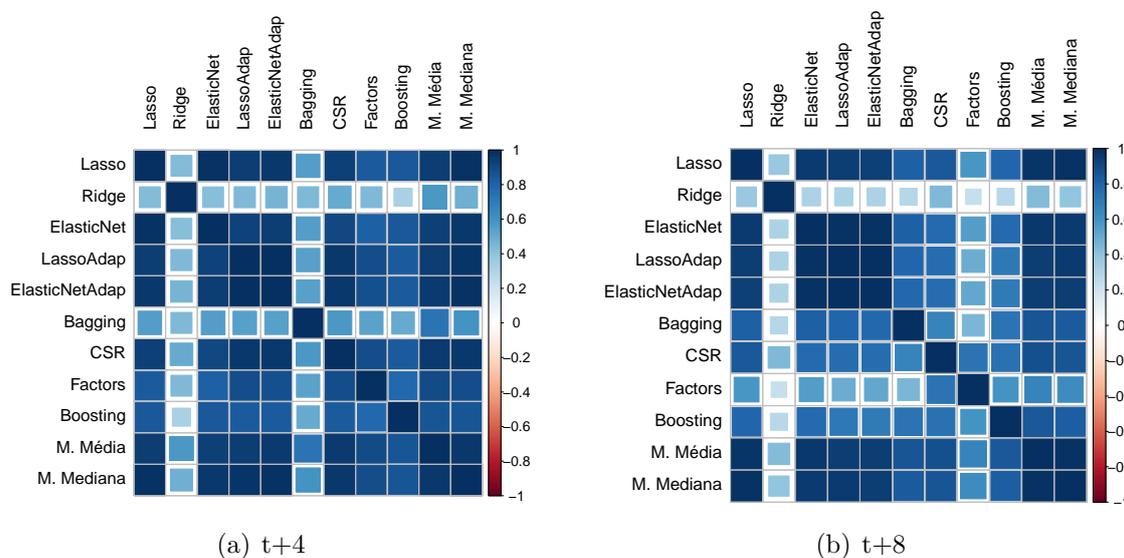
Modelo	Horizonte									
	1	2	3	4	5	6	7	8	4m	8m
Lasso	0.1141	0.1153	0.1276	0.1343	0.1430	0.1425	0.1423	0.1366	0.1093	0.1044
Ridge	0.1774	0.2701	0.2707	0.2699	0.2698	0.2689	0.2586	0.2554	0.2369	0.2481
ElasticNet	0.1129	0.1162	0.1250	0.1328	0.1399	0.1422	0.1383	0.1415	0.1094	0.1045
Factors	0.1403	0.1556	0.1624	0.1767	0.1758	0.1836	0.1733	0.1702	0.1359	0.1422
Baggins	0.1561	0.1872	0.1517	0.2692	0.1915	0.1769	0.6494	0.1543	0.1368	0.1452
Boosting	0.1369	0.1335	0.1359	0.1456	0.1436	0.1444	0.1496	0.1412	0.1274	0.1195
ElasticNetAdap	0.1257	0.1335	0.1385	0.1432	0.1457	0.1422	0.1401	0.1435	0.1215	0.1134
LassoAdap	0.1297	0.1258	0.1397	0.1479	0.1462	0.1471	0.1461	0.1467	0.1221	0.1143
CSR	0.1149	0.1408	0.1323	0.1448	0.1452	0.1336	0.1327	0.1292	0.1214	0.1166
M.Média	0.1090	0.1188	0.1275	0.1395	0.1385	0.1364	0.1427	0.1239	0.1142	0.1130
M.Mediana	0.1095	0.1152	0.1279	0.1359	0.1404	0.1375	0.1312	0.1307	0.1116	0.1090

Fonte: Elaboração própria.

[4] Nota: Para essa previsão, utilizamos 70% da amostra para treino e 30% da amostra para teste, ou seja, o período usado para previsão, dezembro de 2015 a outubro de 2021.

Para realização de um período mais longo de previsão com dados mensais, utilizamos 30% da amostra para teste e diminuimos a amostra de treino para 70%, estratégia semelhante a utilizada por [Garcia, Medeiros e Vasconcelos \(2017\)](#). Conforme mostra a Tabela 2.4, observamos que o modelo Elastic Net manteve sua alta performance de previsão, comparado aos demais modelos, para os períodos iniciais de previsão. Entretanto, os modelos M. Média e M. Mediana foram aqueles com maior acurácia para um período à frente. Para as previsões acumuladas, observamos que o modelo LASSO é altamente competitivo.

Figura 2.3 – Correlação das previsões para Arrecadação de dezembro de 2015 a outubro de 2021



Fonte: Elaboração própria.

[5] Nota: A figura (a) indica as correlações entre as previsões de cada modelo para o quarto período à frente, ou seja, o quarto mês. A figura (b) indica as correlações entre as previsões para o oitavo período à frente.

[6] Nota: Essas correlações foram construídas considerando 70% da amostra de treino e 30% para amostra de teste.

A Figura 2.3 apresenta as correlações das previsões dos modelos de aprendizado estatístico e dos modelos alternativos, M. Média e M. Mediana. Como esperado, todas as correlações são positivas, e a maioria com valores próximos a um. De acordo com a figura (a), o modelo Bagging apresenta maiores correlações com os demais modelos, comparado ao período de previsão da Figura 2.2. Com relação a figura (b), percebemos que as correlações das previsões do modelo Ridge com os demais são próximas de zero.

2.5 TESTE DE ROBUSTEZ

Esta seção apresenta resultados adicionais, baseados em um teste de robustez para a previsão da arrecadação tributária do Brasil. Para esse caso, utilizamos 20% da amostra como teste e 80% da amostra para treino. O exercício de robustez consiste na mudança de frequência de todas as variáveis utilizadas no estudo, inclusive a variável de interesse, arrecadação. Para isso, como os dados originais são mensais, acumulamos os dados em frequência trimestral. Com isso, reduzimos nossa amostra completa de 238 observações para 79 observações. Optamos por excluir a última observação mensal, mês de outubro de

2021, já que essa faz parte do quarto trimestre daquele ano, algo que não é possível para esse estudo.

Tabela 2.5 – RMSE para fora da amostra - Previsão de arrecadação para fora da amostra - Primeiro trimestre de 2018 ao terceiro trimestre de 2021

Modelo	Horizonte									
	1	2	3	4	5	6	7	8	4T	8T
Lasso	0.1481	0.1216	0.1430	0.1289	0.1289	0.1193	0.1241	0.1292	0.1325	0.1258
Ridge	0.1247	0.1372	0.1418	0.1533	0.1625	0.1688	0.1779	0.1837	0.1372	0.1537
ElasticNet	0.1201	0.1249	0.1188	0.1287	0.1283	0.1260	0.1291	0.1325	0.1198	0.1210
Factors	0.1227	0.2473	0.1920	0.1563	0.1223	0.1308	0.1419	0.1270	0.1581	0.1312
Boosting	0.1577	0.1553	0.1442	0.1440	0.1493	0.1625	0.1748	0.1553	0.1392	0.1349
ElasticNetAdap	0.1419	0.1363	0.1283	0.1476	0.1288	0.1220	0.1229	0.1311	0.1360	0.1270
LassoAdap	0.1579	0.1594	0.1604	0.1624	0.1285	0.1189	0.1254	0.1200	0.1599	0.1358
CSR	0.1519	0.1628	0.1538	0.1588	0.1249	0.1223	0.1264	0.1251	0.1539	0.1352
M.Média	0.1230	0.1407	0.1374	0.1374	0.1226	0.1208	0.1255	0.1232	0.1323	0.1254
M.Mediana	0.1224	0.1315	0.1359	0.1380	0.1257	0.1181	0.1240	0.1256	0.1293	0.1244

Fonte: Elaboração própria.

[7] Nota: Para essa previsão, utilizamos 80% da amostra para treino e 20% da amostra para teste, ou seja, o período usado para previsão, primeiro trimestre de 2018 ao terceiro trimestre de 2021.

A Tabela 2.5 apresenta os RMSEs das previsões de arrecadação, com dados trimestrais, para oito períodos à frente, acumulado de quatro períodos à frente (nesse caso, quatro trimestres ou um ano) e acumulado de oito períodos à frente (nesse caso, oito trimestres ou dois anos). Constatamos que o modelo Elastic Net possui a maior acurácia de previsão para os horizontes de previsão iniciais, que supera os modelos M. Média e M. Mediana, que corrobora os resultados para dados mensais apresentados na Tabela 2.3. Esse resultado justifica a alta performance do modelo de aprendizado Shrinkage, Elastic Net, para previsão da variável fiscal brasileira analisada.

2.6 CONCLUSÕES

Dada a relevância de antecipação e controle da arrecadação tributária brasileira, ainda mais após o período da pandemia de Covid-19, a busca por modelos de previsão de alta acurácia é essencial para tomada de decisão com responsabilidade fiscal, por parte de gestores públicos. Na literatura empírica, existem inúmeros estudos que testam métodos estatísticos para previsão de variáveis macroeconômicas, entretanto, existe uma carência de trabalhos voltados para previsão fiscal, sobretudo para o Brasil.

Este ensaio avalia diversas metodologias de alta dimensão, estabelecida na literatura por técnicas de aprendizado estatístico ou *machine learning*, para previsão da arrecadação

brasileira para um longo período. Consideramos três categorias de aprendizagem estatística: shrinkage, ponderação e fatorial. Além disso, estabelecemos modelos médio e mediano como parâmetros de comparação para os demais modelos, algo jamais abordado na literatura de previsão fiscal.

Nossos resultados sugerem: (i) conforme esperado, todas as previsões dos modelos de aprendizado estatístico, apesar de categorias distintas, possuem correlações positivas e altas; (ii) dentre os modelos de aprendizado estatístico, o modelo Elastic Net supera todos para as previsões mensal com uso de 20% e 30% da amostra para teste, para períodos curtos; (iii) o modelo CSR é altamente competitivo para previsões curtas; (iv) para previsões de períodos acumulados, o modelo LASSO tem alta performance; (v) e, por fim, constatamos que o modelo Bagging se torna limitado com aumento dos horizontes de previsão e, sobretudo, diminuição de frequência das variáveis.

Este estudo contribui para gestores e analistas públicos responsáveis por formulação de políticas e tomadas de decisão, que precisam prever receitas fiscais para manutenção e controle de gastos com responsabilidade fiscal. Nossos resultados motivam pesquisas futuras que considerem mais indicadores fiscais para previsão da arrecadação tributária, que podem ser construídos a partir de análise de sentimento textual.

REFERÊNCIAS

- AASTVEIT, K. A.; BJØRNLAND, H. C.; THORSRUD, L. A. What drives oil prices? emerging versus developed economies. *Journal of Applied Econometrics*, Wiley Online Library, v. 30, n. 7, p. 1013–1028, 2015.
- ADÄMMER, P. lpirfs: An r package to estimate impulse response functions by local projections. *The R Journal* (2019), v. 11, n. 2, p. 421–438, 2019.
- ALGABA, A.; ARDIA, D.; BLUTEAU, K.; BORMS, S.; BOUDT, K. Econometrics meets sentiment: An overview of methodology and applications. *Journal of Economic Surveys*, Wiley Online Library, v. 34, n. 3, p. 512–547, 2020.
- BAI, J. Inferential theory for factor models of large dimensions. *Econometrica*, Wiley Online Library, v. 71, n. 1, p. 135–171, 2003.
- BAI, J.; NG, S. Forecasting economic time series using targeted predictors. *Journal of Econometrics*, Elsevier, v. 146, n. 2, p. 304–317, 2008.
- BAI, J.; NG, S. Boosting diffusion indices. *Journal of Applied Econometrics*, Wiley Online Library, v. 24, n. 4, p. 607–629, 2009.
- 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.
- BASNET, H. C.; UPADHYAYA, K. P. Impact of oil price shocks on output, inflation and the real exchange rate: evidence from selected asean countries. *Applied Economics*, Taylor & Francis, v. 47, n. 29, p. 3078–3091, 2015.
- BAUMEISTER, C.; PEERSMAN, G. Time-varying effects of oil supply shocks on the us economy. *American Economic Journal: Macroeconomics*, v. 5, n. 4, p. 1–28, 2013.
- BERNANKE, B. S.; GERTLER, M.; WATSON, M.; SIMS, C. A.; FRIEDMAN, B. M. Systematic monetary policy and the effects of oil price shocks. *Brookings papers on economic activity*, JSTOR, v. 1997, n. 1, p. 91–157, 1997.
- BREIMAN, L. Bagging predictors. *Machine learning*, Springer, v. 24, n. 2, p. 123–140, 1996.
- BURBIDGE, J.; HARRISON, A. Testing for the effects of oil-price rises using vector autoregressions. *International Economic Review*, JSTOR, p. 459–484, 1984.
- CALDEIRA, J. F.; MOURA, G. V.; SANTOS, A. A. Previsões macroeconômicas baseadas em modelos tvp-var: evidências para o brasil. *Revista Brasileira de Economia*, SciELO Brasil, v. 69, p. 407–428, 2015.
- CAMPOS, C. Métodos econométricos aplicados à previsão da arrecadação de receitas federais. *Receita Federal do Brasil. ESAF*, 2009.
- CAVALCANTI, T.; JALLES, J. T. Macroeconomic effects of oil price shocks in brazil and in the united states. *Applied Energy*, Elsevier, v. 104, p. 475–486, 2013.

- CHOI, S.; FURCERI, D.; LOUNGANI, P.; MISHRA, S.; POPLAWSKI-RIBEIRO, M. Oil prices and inflation dynamics: Evidence from advanced and developing economies. *Journal of International Money and Finance*, Elsevier, v. 82, p. 71–96, 2018.
- CORREA, R.; GARUD, K.; LONDONO, J. M.; MISLANG, N. Sentiment in central banks' financial stability reports. *Review of Finance*, Oxford University Press, v. 25, n. 1, p. 85–120, 2021.
- DEENEY, P.; CUMMINS, M.; DOWLING, M.; BERMINGHAM, A. Sentiment in oil markets. *International Review of Financial Analysis*, Elsevier, v. 39, p. 179–185, 2015.
- ELLIOTT, G.; GARGANO, A.; TIMMERMANN, A. Complete subset regressions. *Journal of Econometrics*, Elsevier, v. 177, n. 2, p. 357–373, 2013.
- ELLIOTT, G.; GARGANO, A.; TIMMERMANN, A. Complete subset regressions with large-dimensional sets of predictors. *Journal of Economic Dynamics and Control*, Elsevier, v. 54, p. 86–110, 2015.
- FÈVE, P.; GUAY, A. Sentiments in svars. *The Economic Journal*, Oxford University Press, v. 129, n. 618, p. 877–896, 2019.
- FRANKEL, J. Over-optimism in forecasts by official budget agencies and its implications. *Oxford Review of Economic Policy*, Oxford University Press, v. 27, n. 4, p. 536–562, 2011.
- 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.
- GIACOMINI, R.; WHITE, H. Tests of conditional predictive ability. *Econometrica*, Wiley Online Library, v. 74, n. 6, p. 1545–1578, 2006.
- GU, S.; KELLY, B.; XIU, D. Empirical asset pricing via machine learning. *The Review of Financial Studies*, Oxford University Press, v. 33, n. 5, p. 2223–2273, 2020.
- HAMILTON, J. D. Oil and the macroeconomy since world war ii. *Journal of political economy*, The University of Chicago Press, v. 91, n. 2, p. 228–248, 1983.
- HAMILTON, J. D. This is what happened to the oil price-macroeconomy relationship. *Journal of monetary economics*, Elsevier, v. 38, n. 2, p. 215–220, 1996.
- HAMILTON, J. D. Nonlinearities and the macroeconomic effects of oil prices. *Macroeconomic dynamics*, Cambridge University Press, v. 15, n. S3, p. 364–378, 2011.
- HAMILTON, J. D.; HERRERA, A. M. Comment: oil shocks and aggregate macroeconomic behavior: the role of monetary policy. *Journal of Money, credit and Banking*, JSTOR, p. 265–286, 2004.
- HOERL, A. E.; KENNARD, R. W. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, Taylor & Francis, v. 12, n. 1, p. 55–67, 1970.
- IEA. *IEA Atlas of Energy*. [S.l.]: Report, 2020.
- IEA. *World energy balances and statistics*. [S.l.]: Report, 2020.

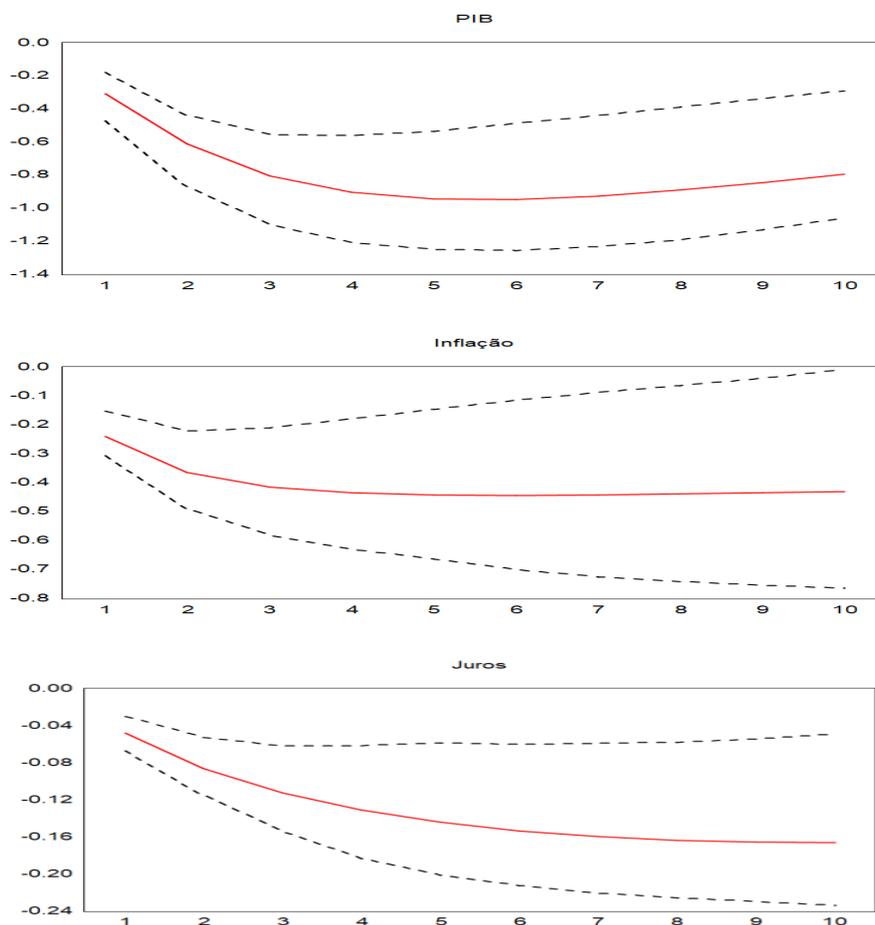
- INOUE, A.; KILIAN, L. How useful is bagging in forecasting economic time series? a case study of us consumer price inflation. *Journal of the American Statistical Association*, Taylor & Francis, v. 103, n. 482, p. 511–522, 2008.
- JOCKERS, M. Package ‘syuzhet’. URL: <https://cran.r-project.org/web/packages/syuzhet>, 2017.
- JORDÀ, Ò. Estimation and inference of impulse responses by local projections. *American economic review*, v. 95, n. 1, p. 161–182, 2005.
- KAMLET, M. S.; MOWERY, D. C.; SU, T.-T. Whom do you trust? an analysis of executive and congressional economic forecasts. *Journal of Policy Analysis and Management*, Wiley Online Library, v. 6, n. 3, p. 365–384, 1987.
- KILIAN, L. Exogenous oil supply shocks: how big are they and how much do they matter for the us economy? *The Review of Economics and Statistics*, MIT Press, v. 90, n. 2, p. 216–240, 2008.
- KILIAN, L. Not all oil price shocks are alike: Disentangling demand and supply shocks in the crude oil market. *American Economic Review*, v. 99, n. 3, p. 1053–69, 2009.
- KLIESEN, K. L.; THORNTON, D. L. How good are the government’s deficit and debt projections and should we care? *Federal Reserve Bank of St. Louis Review*, Citeseer, v. 94, n. 1, p. 21–39, 2012.
- KOCAARSLAN, B.; SOYTAS, M. A.; SOYTAS, U. The asymmetric impact of oil prices, interest rates and oil price uncertainty on unemployment in the us. *Energy Economics*, Elsevier, v. 86, p. 104625, 2020.
- KOCAASLAN, O. K. Oil price uncertainty and unemployment. *Energy Economics*, Elsevier, v. 81, p. 577–583, 2019.
- KOCK, A. B.; CALLOT, L. Oracle inequalities for high dimensional vector autoregressions. *Journal of Econometrics*, Elsevier, v. 186, n. 2, p. 325–344, 2015.
- KRAUSE, G. A.; DOUGLAS, J. W. Institutional design versus reputational effects on bureaucratic performance: Evidence from us government macroeconomic and fiscal projections. *Journal of Public Administration Research and Theory*, Public Management Research Association, v. 15, n. 2, p. 281–306, 2005.
- MCCRACKEN, M. W.; NG, S. Fred-md: A monthly database for macroeconomic research. *Journal of Business & Economic Statistics*, Taylor & Francis, v. 34, n. 4, p. 574–589, 2016.
- MEDEIROS, M. C.; MENDES, E. F. 1-regularization of high-dimensional time-series models with non-gaussian and heteroskedastic errors. *Journal of Econometrics*, Elsevier, v. 191, n. 1, p. 255–271, 2016.
- MEDEIROS, M. C.; VASCONCELOS, G. F.; VEIGA, Á.; ZILBERMAN, E. Forecasting inflation in a data-rich environment: the benefits of machine learning methods. *Journal of Business & Economic Statistics*, Taylor & Francis, v. 39, n. 1, p. 98–119, 2021.
- MEDEIROS, R. K. de; BESARRIA, C. da N.; JESUS, D. P. de; ALBUQUERQUEMELLO, V. P. de. Forecasting oil prices: New approaches. *Energy*, Elsevier, v. 238, p. 121968, 2022.

- NEWKEY, W. K.; WEST, K. D. A simple, positive semi-definite, heteroskedasticity and autocorrelation-consistent covariance matrix. National Bureau of Economic Research Cambridge, Mass., USA, 1986.
- OLEA, J. L. M.; PLAGBORG-MØLLER, M. Local projection inference is simpler and more robust than you think. *Econometrica*, Wiley Online Library, v. 89, n. 4, p. 1789–1823, 2021.
- PEERSMAN, G.; ROBAYS, I. V. Oil and the euro area economy. *Economic Policy*, Oxford University Press, v. 24, n. 60, p. 603–651, 2009.
- PLAGBORG-MØLLER, M.; WOLF, C. K. Local projections and vars estimate the same impulse responses. *Econometrica*, Wiley Online Library, v. 89, n. 2, p. 955–980, 2021.
- PLESKO, G. A. The accuracy of government forecasts and budget projections. *National Tax Journal*, The University of Chicago Press, v. 41, n. 4, p. 483–501, 1988.
- RAMEY, V. A.; ZUBAIRY, S. Government spending multipliers in good times and in bad: evidence from us historical data. *Journal of political economy*, University of Chicago Press Chicago, IL, v. 126, n. 2, p. 850–901, 2018.
- TIBSHIRANI, R. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, Wiley Online Library, v. 58, n. 1, p. 267–288, 1996.
- ZHAO, L.; ZHANG, X.; WANG, S.; XU, S. The effects of oil price shocks on output and inflation in china. *Energy Economics*, Elsevier, v. 53, p. 101–110, 2016.
- ZOU, H. The adaptive lasso and its oracle properties. *Journal of the American statistical association*, Taylor & Francis, v. 101, n. 476, p. 1418–1429, 2006.
- ZOU, H.; HASTIE, T. Regularization and variable selection via the elastic net. *Journal of the royal statistical society: series B (statistical methodology)*, Wiley Online Library, v. 67, n. 2, p. 301–320, 2005.
- ZUANAZZI, P. T.; ZIEGELMANN, F. A. Previsões para o crescimento do pib trimestral brasileiro com séries financeiras e econômicas mensais: uma aplicação de midas. *Economia Aplicada*, SciELO Brasil, v. 18, p. 295–318, 2014.

Apêndices

APÊNDICE A – CAPÍTULO 1

Figura A.1 – Respostas ao impulso em choques de petróleo com estimação por SVAR

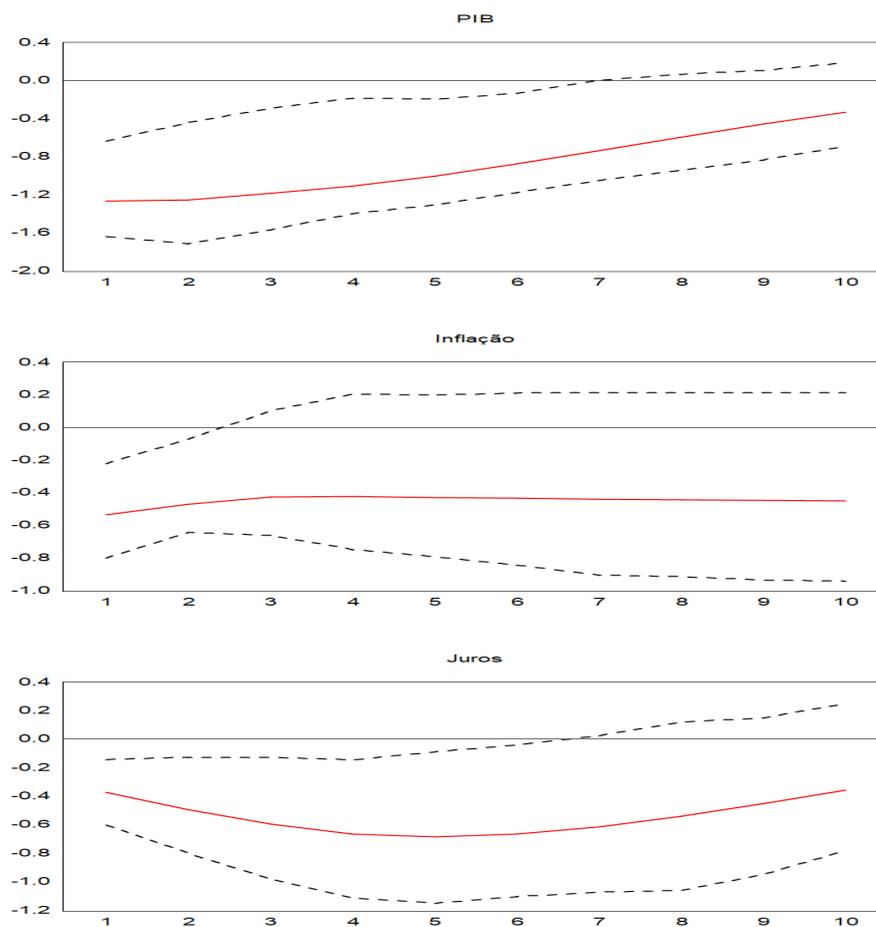


Fonte: Elaboração própria.

[8] Nota: Para estimação das IRF's, utilizamos a variável de impulso (choque) VIX.

[9] Nota: A defasagem ótima das variáveis utilizadas na estimação das IRF's (todas as variáveis endógenas) é apresentada na Tabela A.3.

Figura A.2 – Respostas ao impulso em choques de petróleo com frequência trimestral

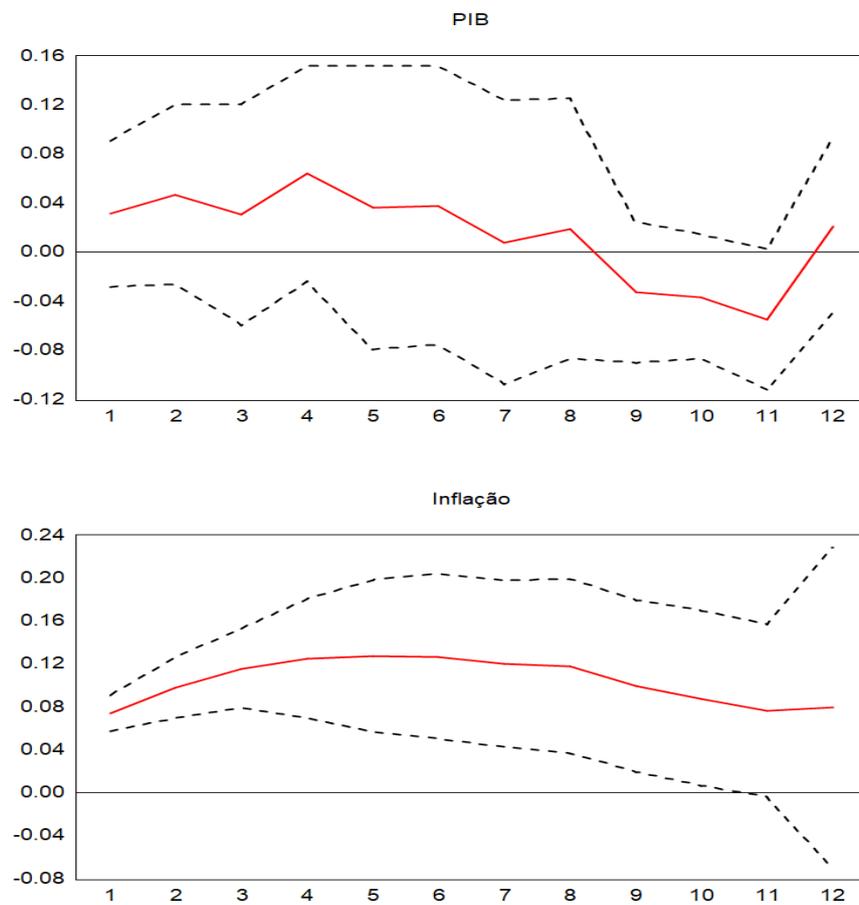


Fonte: Elaboração própria.

[10] Nota: Para estimação das IRF's, utilizamos a variável de impulso (choque) VIX.

[11] Nota: A defasagem ótima das variáveis utilizadas na estimação das IRF's (todas as variáveis endógenas) é apresentada na Tabela A.7.

Figura A.3 – Respostas ao impulso de petróleo WTI (Cavalcanti e Jalles (2013)) - Estimaco por Projeo Local



Fonte: Elaboraco prpria.

[12] Nota: Os intervalos de confiana possuem nvel de confiana de 95%.

[13] Nota: A estimaco  realizada com uma defasagem, j que o nmero de observaces  pequeno e a freqncia das variveis  trimestral.

Tabela A.1 – Estatísticas descritivas das variáveis utilizadas

Variável	Média	Mediana	Máximo	Mínimo	D. Padrão	Observações
PIB - EUA	96.62	98.00	104.18	84.73	4.96	240
Juros - EUA	1.49	1.01	5.98	0.05	1.64	240
Inflação - EUA	219.69	220.83	261.56	175.60	25.12	240
PIB - Brasil	4108924	3908022	7475380	1209046	2043350	240
Juros - Brasil	12.34	11.74	26.32	1.90	5.10	240
Inflação - Brasil	6.17	5.90	17.24	1.88	2.79	240
Sent	-0.14	-0.14	0.45	-0.71	0.22	240
VIX	19.84	17.30	59.89	9.51	8.43	240

Fonte: Elaboração própria.

Tabela A.2 – Defasagem ótima das variáveis endógenas com a inclusão da variável de choque Sent - Estimação por Projeção Local e SVAR - EUA

Defasagem	AIC	HQ	SC
1	-7.7672	-7.6467	-7.4683
2	-8.4658	-8.2488	-7.9277*
3	-8.6300	-8.3164*	-7.8526
4	-8.6458*	-8.2358	-7.6293
5	-8.5592	-8.0527	-7.3036
6	-8.5897	-7.9868	-7.0949
7	-8.5319	-7.8325	-6.7979
8	-8.4432	-7.6473	-6.4701
9	-8.4156	-7.5232	-6.2032
10	-8.3665	-7.3776	-5.9150

Fonte: Elaboração própria.

Tabela A.3 – Defasagem ótima das variáveis endógenas com a inclusão da variável de choque VIX - Estimação por Projeção Local e SVAR - EUA

Defasagem	AIC	HQ	SC
1	-1.9678	-1.8472	-1.6688
2	-2.4807	-2.2636	-1.9425*
3	-2.6352	-2.3216*	-1.8579
4	-2.6469*	-2.2369	-1.6305
5	-2.6077	-2.1012	-1.3521
6	-2.5518	-1.9488	-1.0570
7	-2.5250	-1.8255	-0.7910
8	-2.5072	-1.7113	-0.5341
9	-2.4589	-1.5665	-0.2465
10	-2.4639	-1.4750	-0.0124

Fonte: Elaboração própria.

Tabela A.4 – Optimal lag of the exogenous variable Sent

Lag	AIC	HQ	SC
1	-3.1366	-3.1246	-3.1068*
2	-3.1467*	-3.1286*	-3.1019
3	-3.1380	-3.1139	-3.0783
4	-3.1349	-3.1047	-3.0601
5	-3.1322	-3.0960	-3.0425
6	-3.1376	-3.0954	-3.0330
7	-3.1324	-3.0842	-3.0128
8	-3.1241	-3.0698	-2.9896
9	-3.1161	-3.0558	-2.9666
10	-3.1129	-3.0466	-2.9485

Source: Authors' elaboration.

Tabela A.5 – Defasagem ótima da variável exógena VIX

Defasagem	AIC	HQ	SC
1	3.1520*	3.1640*	3.1819*
2	3.1593	3.1774	3.2042
3	3.1574	3.1815	3.2172
4	3.1625	3.1927	3.2373
5	3.1712	3.2074	3.2609
6	3.1771	3.2193	3.2817
7	3.1765	3.2248	3.2961
8	3.1839	3.2382	3.3184
9	3.1916	3.2519	3.3410
10	3.1946	3.2609	3.3590

Fonte: Elaboração própria.

Tabela A.6 – Defasagem ótima das variáveis endógenas - Estimação por Projeção Local para variáveis mensais (sem a variável de choque) - Brasil

Defasagem	AIC	HQ	SC
1	23.4464	23.5188	23.6258
2	21.6696	21.7963*	21.9836*
3	21.6882	21.8691	22.1366
4	21.7020	21.9372	22.2850
5	21.5711	21.8605	22.2886
6	21.4945*	21.8382	22.3466
7	21.5309	21.9288	22.5174
8	21.5418	21.9940	22.6629
9	21.5821	22.0886	22.8377
10	21.5675	22.1282	22.9576

Fonte: Elaboração própria.

Tabela A.7 – Defasagem ótima das variáveis endógenas - Estimação por Projeção Local para variáveis trimestrais (com inclusão da variável de choque VIX) - EUA

Defasagem	AIC	HQ	SC
1	4.4677*	4.7229	5.1101
2	3.9360	4.3954*	5.0924*
3	3.8905	4.5540	5.5608
4	3.9070	4.7746	6.0913
5	3.9854	5.0571	6.6835
6	4.1535	5.4294	7.3657
7	4.2671	5.7471	7.9931
8	4.4801	6.1643	8.7201
9	4.3106	6.1989	9.0646
10	4.3360	6.4285	9.6039

Fonte: Elaboração própria.

APÊNDICE B – CAPÍTULO 2

Tabela B.1 – Estatísticas descritivas das variáveis utilizadas

Variável	Média	Mediana	Máximo	Mínimo	D. Padrão	Observações
Arrecadação	80444.95	79539.15	180221.24	22682.45	37847.86	238
Papelão ondulado	261277.91	269930.00	371179.00	157096.00	47699.40	238
Produção industrial (produtos de borracha e material plástico)	92.21	91.80	112.50	55.10	9.21	238
Produção industrial (celulose, papel e produtos em geral)	95.71	97.35	117.90	67.80	9.77	238
Produção (geral)	91.43	91.45	112.60	60.30	9.92	238
Produção (indústria de transformação)	91.25	91.15	113.70	57.70	10.28	238
Produção (bens de capital)	84.19	80.00	127.10	37.70	19.14	238
Produção (bens intermediários)	93.03	93.55	111.40	66.80	8.74	238
Produção (bens de consumo)	90.90	91.30	116.30	53.50	11.00	238
Produção total de autoveículos	223.69	219.15	352.33	1.85	58.65	238
Vendas de autoveículos (total)	200849.90	213283.00	383535.00	1323.00	82889.75	238
Consumo de energia elétrica (comercial)	6023.47	6254.50	8204.00	3450.00	1389.92	238
Consumo de energia elétrica (industrial)	13796.83	14075.00	15886.00	11.00	1858.12	238
Consumo de energia elétrica (total)	34622.08	36344.00	43416.00	22324.00	5359.67	238
Volume vendas no varejo (automóveis, motocicletas, partes e peças)	77.82	80.00	133.70	37.70	21.90	238
IGP-M	0.69	0.57	5.19	-1.10	0.92	238
Índice Geral de Preços - disponibilidade interna (IPA-M)	0.69	0.53	5.84	-1.24	0.92	238
Índice de Preços por Atacado - Mercado (IGP-M)	0.77	0.63	6.73	-1.77	1.28	238
INPC	0.52	0.44	3.39	-0.30	0.45	238
IPCA	0.48	0.44	2.53	-0.35	0.38	238
Produção de caminhões	10164.34	9384.50	22647.00	403.00	4099.02	238
Índice Volume de Vendas no Varejo	79.02	85.45	130.30	42.80	20.35	238
Contribuição - Finsocial / Cofins	13216.69	13435.90	26314.11	3495.75	5731.81	238
Imposto sobre a Renda - Pessoas Físicas	1876.11	1397.53	11663.13	156.30	1749.50	238
Imposto sobre a Renda - Pessoas Jurídicas	8689.31	6840.00	37839.48	1629.19	5942.01	238
Dívida Líquida do Setor Público - Interna	49.20	45.78	75.91	36.83	9.99	238
Dívida Líquida do Setor Público - Total	44.92	45.98	62.54	30.01	9.09	238
Dívida Líquida do Setor Público - Externa	-4.29	-6.76	17.20	-15.82	7.97	238
Dívida Líquida do Governo Geral	45.59	45.32	67.16	30.71	9.61	238
Dívida Fiscal Líquida	41.57	36.66	74.48	29.50	12.47	238
NFSP sem desvalorização cambial	4.04	2.85	13.75	0.00	3.03	238
Dívida mobiliária	2968422.49	2508667.55	7232573.36	699812.73	1842412.33	238
Saldo da carteira de crédito	1999196.59	2010899.00	4511956.00	337735.00	1220041.75	238
Selic	0.92	0.89	2.08	0.13	0.40	238
Taxa de câmbio	2.85	2.51	5.65	1.56	1.05	238

Fonte: Elaboração própria.

Anexos

ANEXO A – CAPÍTULO 2

anpec associação nacional dos centros de pós-graduação em economia
50º Encontro Nacional de Economia - ANPEC

Em 27/09/2022

Prezado(a) RENNAN KERTLLY DE MEDEIROS

Temos a satisfação de comunicar que o trabalho indicado abaixo foi aceito para apresentação no 50º Encontro Nacional de Economia:

Título do trabalho: "ESTRATÉGIAS DE PREVISÃO FISCAL: UM ESTUDO EMPÍRICO PARA A ECONOMIA BRASILEIRA"

Autor(es):

RENNAN KERTLLY DE MEDEIROS
EDILEAN KLEBER DA SILVA BEJARANO ARAGÓN
CÁSSIO DA NÓBREGA BESARRIA

O Encontro será realizado em Fortaleza/CE (Hotel Grand Mareiro), no período de 06 a 09 de dezembro de 2022.

Maiores informações a respeito do Encontro estarão disponíveis no site www.anpec.org.br. Lembramos que os apresentadores de trabalhos também deverão fazer sua inscrição no evento, pagando a respectiva taxa. A inscrição deverá ser feita por meio de formulário online que estará disponível entre 28/09 e 30/11 no site da ANPEC.

Atenção: Para trabalhos em co-autoria solicitamos a gentileza de informar a Secretaria da ANPEC (eliane@anpec.org.br), com a máxima brevidade possível, o nome do autor que será o apresentador do trabalho.

Na certeza de que a sua valiosa participação em muito contribuirá para o sucesso do nosso Encontro, agradecemos antecipadamente.

Atenciosamente,


Prof. Luciano Nakabashi
Presidente da ANPEC