

**UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL  
FACULDADE DE CIÊNCIAS ECONÔMICAS  
PROGRAMA DE PÓS-GRADUAÇÃO EM ECONOMIA**

**LUÍS EDUARDO DE PAULA SOUSA**

**PREVISÃO E *NOWCASTING* DO CRESCIMENTO TRIMESTRAL DO PIB DO RIO  
GRANDE DO SUL VIA MIDAS**

**Porto Alegre  
2022**

**LUÍS EDUARDO DE PAULA SOUSA**

**PREVISÃO E *NOWCASTING* DO CRESCIMENTO TRIMESTRAL DO PIB DO RIO  
GRANDE DO SUL VIA MIDAS**

Dissertação submetida ao Programa de Pós-Graduação em Economia da Faculdade de Ciências Econômicas da UFRGS, como requisito parcial para obtenção do título de Mestre em Economia, área de concentração: Economia Aplicada.

Orientador: Prof. Dr. Flávio A. Ziegelmann

**Porto Alegre  
2022**

### CIP - Catalogação na Publicação

Sousa, Luis Eduardo de Paula  
Previsão e nowcasting do crescimento trimestral do  
PIB do Rio Grande do Sul / Luis Eduardo de Paula  
Sousa. -- 2022.  
51 f.  
Orientador: Flávio Augusto Ziegelmann.

Dissertação (Mestrado) -- Universidade Federal do  
Rio Grande do Sul, Faculdade de Ciências Econômicas,  
Programa de Pós-Graduação em Economia, Porto Alegre,  
BR-RS, 2022.

1. PIB. 2. Previsão. 3. Rio Grande do Sul. 4.  
MIDAS. I. Ziegelmann, Flávio Augusto, orient. II.  
Título.

**LUÍS EDUARDO DE PAULA SOUSA**

**PREVISÃO E *NOWCASTING* DO CRESCIMENTO TRIMESTRAL DO PIB DO RIO  
GRANDE DO SUL VIA MIDAS**

Dissertação submetida ao Programa de Pós-Graduação em Economia da Faculdade de Ciências Econômicas da UFRGS, como requisito parcial para obtenção do título de Mestre em Economia, área de concentração: Economia Aplicada.

Orientador: Prof. Dr. Flávio A. Ziegelmann

Aprovada em Porto Alegre, 21 de dezembro de 2022.

**BANCA EXAMINADORA:**

Prof. Dr. Flávio Augusto Ziegelmann – Orientador  
PPGE/UFRGS

Prof. Dr. Sabino Porto Junior  
PPGE/UFRGS

Dr. Pedro Tonon Zuanazzi  
PPGE/UFRGS

Dr. Wallace Patrick Santos De Farias Souza  
PPGE/UFPB

## **AGRADECIMENTOS**

Gostaria de agradecer a minha esposa Carolina pelo apoio, carinho e paciência durante todo esse mestrado, que foram fundamentais para poder superar esta jornada. Meus sinceros agradecimentos!

Agradecimentos também para os meus filhos Gustavo e Marina (os Pinpinhos do papai) que compartilharam comigo momentos de alegria ao longo dessa trajetória. Um agradecimento a minha mãe Lucia, que sempre esteve do meu lado e me motivou por toda vida.

Agradeço em especial ao meu orientador, o professor Dr. Flávio A. Ziegelmann, que me encorajou na busca por conhecimentos, sempre presente, solícito, dedicado e atencioso. Além dos aspectos acadêmicos, um agradecimento particular à amizade, colaboração e por sempre acreditar em mim.

Por ultimo, agradeço ao Programa de Pós-Graduação em Economia Aplicada da UFRGS pela estrutura e o profissionalismo oferecidos durante o curso, contribuindo decisivamente para a minha formação.

## RESUMO

Tendo como princípio o fato de que previsão do PIB é essencial para a orientação das políticas públicas e, também, é um balizador para os agentes econômicos, o objetivo deste trabalho é realizar a previsão do PIB trimestral do Estado do Rio Grande do Sul para a distância temporal do 2º trimestre de 2003 ao 1º trimestre de 2022. Para realização deste estudo são aplicadas as abordagens tradicionais MIDAS (Mixed Data Sampling) e UMIDAS (Unrestricted Data Sampling), além de serem desenvolvidos dois novos métodos, que introduzem parâmetros variantes no tempo, denominados MIDAS Rolling Regression e a MIDAS Recursive Regression, na busca de maior flexibilidade aos parâmetros para tentar captar as abruptas quedas e a recuperação na economia durante o período mais agudo da pandemia do COVID-19. Em nossas análises empíricas, os métodos que adotam parâmetros variantes no tempo, que são o MIDAS Rolling Regression e o MIDAS Recursive Regression, têm desempenhos superiores àqueles obtidos através dos tradicionais MIDAS, AR-MIDAS, UMIDAS e AR-UMIDAS, com destaque para o MIDAS Rolling Regression que apresenta o melhor desempenho entre todos.

**Palavras-chave:** PIB. modelo de previsão. Rio Grande do Sul. MIDAS

## **ABSTRACT**

The GDP forecast is essential to guide public policies and a guide for economic agents. The objective of this work is to forecast the quarterly GDP of the State of Rio Grande do Sul for the periods covering the 2nd quarter of 2003 to the 1st quarter of 2022. To carry out this study, the traditional MIDAS (Mixed Data Sampling) and UMIDAS approaches are applied. (Unrestricted Data Sampling), in addition to the development of two new models in this work, called MIDAS Rolling Regression and MIDAS Recursive Regression, in the search for greater flexibility in the parameters to try to capture the abrupt declines and recovery in the economy during the most acute period of COVID-19 pandemic. These results are compared with the ARMA benchmark. The models that adopt time-varying parameters, which are the MIDAS Rolling Regression and the MIDAS Recursive Regression, have superior performances than the MIDAS, AR-MIDAS, UMIDAS and AR-UMIDAS models, with emphasis on the MIDAS Rolling Regression that presents the best performance among all.

**Keywords:** GDP. forecasting. MIDAS. Rio Grande do Sul.

## LISTA DE FIGURAS

Figura 1 – Taxa de Crescimento do PIB do RS . . . . .	45
Figura 2 – Vendas Internas . . . . .	45
Figura 3 – Vendas para outros Estados . . . . .	46
Figura 4 – Exportação . . . . .	46
Figura 5 – Compras Internas . . . . .	47
Figura 6 – Compras de Outros Estados . . . . .	47
Figura 7 – Importação . . . . .	48
Figura 8 – Carga Tributária . . . . .	48
Figura 9 – Valor Adicionado . . . . .	49
Figura 10 – Arrecadação de ICMS . . . . .	49
Figura 11 – Tabela de Correlação . . . . .	50
Figura 12 – Gráfico de Correlação . . . . .	51

## LISTA DE TABELAS

Tabela 1 – Teste de raiz unitária . . . . .	26
Tabela 2 – Teste de Raiz Unitária com Quebras Estruturais de Zivot e Andrews (1992) .	27
Tabela 3 – Séries temporais . . . . .	27
Tabela 4 – Um regressor - resultados A . . . . .	30
Tabela 5 – Um regressor - resultados B (continuação...) . . . . .	31
Tabela 6 – Resultados - Modelos com múltiplos regressores . . . . .	34
Tabela 7 – Comparativo - Modelos com múltiplos regressores . . . . .	35
Tabela 8 – Model Confidence set (H=1) . . . . .	37
Tabela 9 – Model Confidence set (H=2/3) . . . . .	37
Tabela 10 – Model Confidence set (H=1/3) . . . . .	38
Tabela 11 – Model Confidence set (H=0) . . . . .	38

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO . . . . .</b>	<b>9</b>
<b>2</b>	<b>METODOLOGIA . . . . .</b>	<b>14</b>
2.1	ABORDAGEM MIDAS e UMIDAS . . . . .	14
2.1.1	<b>Midas Básico . . . . .</b>	<b>14</b>
2.1.2	<b>Midas autoregressivo: AR - MIDAS . . . . .</b>	<b>15</b>
2.1.3	<b>Modelo ar-midas com múltiplo regressores . . . . .</b>	<b>16</b>
2.1.4	<b>Unconstrained MIDAS:UMIDAS . . . . .</b>	<b>16</b>
2.2	ESTIMAÇÃO . . . . .	17
2.3	REGRESSORES VARIANTES NO TEMPO . . . . .	18
2.3.1	<b>Matriz de covariáveis . . . . .</b>	<b>18</b>
2.3.2	<b>Regressão por janela . . . . .</b>	<b>20</b>
2.3.3	<b>Regressão recursiva . . . . .</b>	<b>21</b>
2.3.4	<b>Regressores fixos no tempo . . . . .</b>	<b>24</b>
2.4	PREVISÃO COM INFORMAÇÕES DO TRIMESTRE CORRENTE . . . . .	24
<b>3</b>	<b>ANÁLISE EMPÍRICA . . . . .</b>	<b>26</b>
3.1	TESTE DE RAIZ UNITÁRIA . . . . .	26
3.2	MODELOS INDIVIDUAIS . . . . .	28
3.3	MODELOS COM MÚLTIPLOS REGRESSORES . . . . .	33
3.4	COMPARATIVO: MODELO COM MÚLTIPLOS REGRESSORES . . . . .	35
3.5	MODEL CONFIDENCE SET . . . . .	36
<b>4</b>	<b>CONCLUSÃO . . . . .</b>	<b>39</b>
	<b>REFERÊNCIAS . . . . .</b>	<b>41</b>
	<b>APÊNDICE A - GRÁFICOS DAS SÉRIES TEMPORAIS . . . . .</b>	<b>45</b>
	<b>APÊNDICE B - MATRIZ DE CORRELAÇÃO DAS VARIÁVEIS . . . . .</b>	<b>50</b>

## 1 INTRODUÇÃO

O Produto Interno Bruto (PIB) representa a soma de todos os bens e serviços finais produzidos em uma determinada região durante um certo período de tempo. Cada país, estado ou município calculam o seu PIB considerando apenas os bens e serviços finais, excetuando os bens intermediários utilizados. Pode-se afirmar, então, que o PIB representa o fluxo de novos bens e serviços finais produzidos durante um período. O PIB tem sido o indicador mais utilizado nos institutos de pesquisa e estatística econômica para mensurar os níveis de atividade da economia de um país.

Ao contrário de outras variáveis macroeconômicas, que tendem a ser extraídas e publicadas em uma frequência maior no tempo, a apuração do PIB é realizada trimestralmente e ainda algumas semanas depois do período de referência. Essa lacuna temporal implica que Governos e Instituições – tais como Bancos Centrais, Ministérios da Fazenda, Planejamento ou Secretarias da Fazenda – tomem suas decisões sobre leis orçamentarias (receita e despesas), fixação da taxa básica de juros, arrolamento de dívidas públicas, arrecadação e investimentos públicos baseadas frequentemente em dados econômicos incompletos, pois algumas estatísticas sobre a economia são divulgadas com certo atraso, ou seja, sem o devido conhecimento do estado atual da economia.

Isso tem levado ao crescimento na literatura do conceito de *nowcasting*, usando conjuntos de dados de alta dimensão, que envolvem a previsão do estado atual da economia antes dos números oficiais serem liberados (ZUANAZZI; ZIELGELMANN, 2014; CEPNI *et al.*, 2019). A questão essencial deste trabalho é, então, a construção de um modelo de previsão e *nowcasting* da taxa de crescimento do PIB trimestral do Rio Grande do Sul (RS), por meio de variáveis econômicas apuradas mensalmente. Atualmente o RS possui o 4º (quarto) maior PIB do país, sendo uma das federações mais industrializadas, além de possuir umas das maiores rendas per captas do Brasil (IBGE, 2022), contribuindo positivamente para geração de riqueza do Brasil (IBGE, 2022).

A metodologia de apuração do PIB inviabiliza que sua realização seja feita com periodicidade mensal, sendo mais comuns ciclos trimestrais ou anuais. Já alguns indicadores macroeconômicos capazes de indicar os ciclos econômicos – tais como vendas do varejo, produção industrial, compras da indústria, exportação, importação e etc – são divulgados com frequência mensal. Como eles são publicados de maneira prévia ao PIB, possuem informações relevantes sobre a atividade econômica corrente e futura. Historicamente, sabe-se que vários modelos econométricos de predição do PIB ajustam suas variáveis explicativas para a mesma frequência da variável resposta, ou seja, as séries mensais são convertidas para um frequência trimestral, seja por meio de uma agregação (soma dos meses do trimestre), seja por meio da escolha de apenas o último mês do trimestre. Isso faz com que o grau de informação acabe sendo perdido

durante a adaptação das variáveis, para que se tenha a mesma frequência da variável resposta (GALVÃO; CLEMENTS, 2008). Há também modelos que adotam simultaneamente todos os regressores em diferentes frequências, como no caso dos modelos VAR com Multi-frequência (MF-VAR)– para maiores detalhes, ver Kuzin *et al.* (2009) e Borges e Portugal (2018).

Neste trabalho é utilizada a regressão MIDAS (*Mixed Data Sampling*) para a construção de um modelo preditivo para o crescimento do PIB trimestral do Estado do Rio Grande do Sul. Nesse sentido, essa abordagem, introduzida por Ghysels *et al.* (2004, 2005), permite que regressores e regressandos sejam relacionados em diferentes frequências, sem a necessidade de desconsiderar parte da informação das séries de maiores frequências (exemplo: caso em que é feita a média dos três meses ou são descartados alguns deles), ou seja, uma regressão MIDAS permite que a variável de baixa frequência seja explicada por variáveis de alta frequência e suas defasagens. Os dados que possuem uma maior frequência são ponderados por uma função polinomial parcimoniosa, que é capaz de adquirir diversos formatos com poucos parâmetros, não havendo, assim, prejuízos com o número de graus de liberdade (ver, por exemplo, GALVÃO; CLEMENTS, 2009).

A regressão MIDAS já foi objeto de estudo para construção de modelos de previsão do PIB dos Estados Unidos (EUA), da Europa e do Brasil. Tay (2007) utilizou-se do MIDAS para prever o crescimento do PIB dos EUA utilizando como variáveis explicativas o índice de ações. Os resultados obtidos por ele indicam que a inclusão dos retornos do mercado de ações é um importante indicador para acompanhar o ciclo da economia, e que o uso dos retornos diários das ações são úteis para atualizar o comportamento do crescimento do PIB. Andreou *et al.* (2013) estudaram o modelo MIDAS para identificar se ativos financeiros produzem informações relevantes para previsões do crescimento do PIB dos EUA. Para isso, valeram-se de mais de 1000 preditores, subdivididos em cinco classes de ativos, conforme a seguir: risco corporativos, ações, renda fixa, commodities e câmbio, cujos dados possuíam frequência diária. Como resultado disso, eles identificaram que a capacidade preditiva de modelos que fazem uso de dados macroeconômicos e financeiro agregados são inferiores aos que usam dados financeiros diários.

Ferrara e Marsilli (2013) utilizaram-se do MIDAS para realizar previsões para o PIB da Alemanha, França, Itália e Espanha. Como variáveis explicativas do modelo utilizaram índices financeiros e econômicos. Eles, mais precisamente, concluíram que índices de preços de ações possuem melhores resultados preditivos para o crescimento do PIB do que indicadores econômicos, tais como o preço do barril de petróleo ou do *spread* bancário. Barsoum (2011) aplicou o MIDAS para prever o PIB do Reino Unido, usando 242 variáveis financeiras, divididas em 5 grandes grupos: Commodities (44 variáveis), *corporate risk* – CDI e os certificado de depósitos – que totalizaram 10 variáveis, *equities* – índices de preço, ações e etc – com 60 variáveis, títulos do governo (87 variáveis) e câmbio (a cotação da libra esterlina em 41 moedas diferentes).

Conclui-se, a partir disso, que o modelo ADL-MIDAS apresentou melhor desempenho preditivo em relação aos outros modelos benchmark (modelos RW, AR, FAR, ADL, FADL) e que os indicadores financeiros apresentam melhor desempenho sobre o comportamento do crescimento do PIB britânico, quando comparados aos outros indicadores.

No Brasil, Zuanazzi e Ziegelmann (2014) foram os primeiros a aplicarem o MIDAS para previsões macroeconômicas, utilizando 16 séries financeiras e econômicas como preditores – seus resultados foram comparados com o benchmark ARMA. Conclui-se que os diversos regressores analisados individualmente tiveram um bom desempenho fora da amostra, com destaque para: Ibovespa, valor da exportação, produção das indústrias gerais e de transformação, principalmente quando utilizados dados dentro do mesmo trimestre. Ademais, quando foram utilizados múltiplos regressores, os resultados de previsão foram melhores do que quando utilizados os regressores individualmente.

Em um estudo mais recente, o de Ghysels *et al.* (2020), ampliou-se a discussão sobre a regressão MIDAS, introduzindo novos modelos a partir de diferentes funções parcimoniosas de defasagens. Essas novas derivações foram divididas em três categorias, sendo elas: a) os modelos lineares e quase-lineares, que são os modelos MIDAS e UMIDAS, estudados nesta dissertação; b) os modelos semiparamétricos não-lineares, que fazem o uso de polinômios locais como função parcimoniosa de defasagens; c) modelos paramétricos não-lineares, que adotam como função parcimoniosa de defasagens equações de transição logística suave ou equações de mínimo, médio, médio ponderado ou máximo para os diferentes regressores e suas defasagens.

No processo de desenvolvimento de modelos econométricos de previsão do crescimento do PIB, há eventos externos que dificultam a sua criação e que estão fora do controle dos agentes públicos, como as pandemias ou os desastres naturais. Isso aconteceu mais recentemente com a pandemia do Coronavírus, que foi o mais importante evento histórico, social e econômico das últimas sete décadas e exigiu dos governos de todo o mundo medidas de saúde pública para conter o avanço do vírus, gerando impactos econômicos diversos ao redor do mundo, afetando, com isso, além do cotidiano das famílias, a existência da atividade econômica ao redor do mundo (GROMLING, *et al.*, 2021).

O período analisado neste trabalho é impactado pelas consequências econômicas da pandemia do Coronavírus. Em períodos de instabilidade econômica – tais como inflações, recessões global, guerras ou pandemias – podem ocorrer mudanças estruturais que afetam o comportamento de firmas e consumidores em relação às suas expectativas, prejudicando o desempenho dos modelos de previsão econométricos. Uma solução, segundo Alptekin *et al.* (2018), seriam modelos econométricos preditivos que possuem parâmetros flexíveis ao longo do tempo, como uma alternativa para capturar dinamicamente essas variações estruturais na economia.

O Modelo Midas tradicional, assim como os demais modelos econométricos padrões,

tem como base parâmetros fixos por toda amostra analisada (ver GHYSELS *et al.*, 2007). Em Li *et al.* (2022), buscando maior flexibilidade dos parâmetros nos modelos de previsão econométrico, foi desenvolvido um modelo que combinava a regressão MIDAS, LASSO e Mudança de Regime Markoviana, chamado de MS-LASSO-MIDAS, para realizar previsões do índice S&P 500 durante o período da pandemia de COVID-19, concluindo que a regressão com regimes markovianos distintos combinada com a LASSO apresentou melhores resultados que os modelos MIDAS tradicionais (MIDAS E AR-MIDAS).

Nesse estudo, portanto, são introduzidos dois novos modelos que trazem o conceito de parâmetros variantes no tempo, chamados de *MIDAS-Rolling Regression*, este que trabalha com sub-amostras (regressão por janelas), e *MIDAS-Recursive Regression*, este que utiliza regressão recursiva, os quais são aqui propostos com base na metodologia utilizada por Guerin e Marcelino (2013) e Marsilli (2014), que introduziram, respectivamente, os modelos MS-MIDAS (Markov Switching Midas) e MIDAS-LASSO. Nesses trabalhos, eles consideram a regressão MIDAS como um modelo de regressão linear padrão que possui um núcleo não-linear (função parcimoniosa de defasagens), o que possibilitou agregar com alto grau de flexibilidade às variáveis explicativas de maior frequência e às suas defasagens em um vetor de variáveis explicativas. Ambos autores constroem uma matriz de covariáveis, resultado do produto da matriz de regressores e das suas defasagens com o vetor de coeficientes das defasagens obtido através de seu núcleo não-linear, realizando posteriormente a regressão Markov Switching e a LASSO.

Além de mensurar o desempenho dos resultados de previsão (*forecasting*) do crescimento do PIB trimestral do Estado do Rio Grande do Sul, analisa-se aqui o desempenho dos resultados de *nowcasting* que os modelos MIDAS e suas derivações fornecem. Nota-se, assim, que o conceito de *nowcasting*, segundo Kagohara (2015), é definido como a previsão do presente, do futuro muito próximo e do passado recente.

Castle *et al.* (2009) apresentam quatro principais razões para o "*nowcasting*" do estado corrente da economia, que são os seguintes:

- a) os dados desagregados (no caso os dados mensais) estão disponíveis no momento da construção do modelo de *nowcasting* do PIB daquele período em análise;
- b) muitas séries econômicas temporais são apenas estimativas preliminares, ou "instantâneas", estando sujeitas a substanciais revisões ao longo do tempo à medida que mais informações se acumulam, não sendo necessariamente um guia confiável das condições atuais;
- c) muitos dados mensais, que entram no modelo, acabam indisponíveis em determinado período, não sendo, portanto, um banco de dados único e sempre disponível;
- d) quando valores reais das variáveis explicativas relevantes estão disponíveis, o *nowcasting* deve ser realizado – isso permite uma avaliação imediata, servindo como sinal de alerta antecipado quando as previsões se afastem muito da série medida.

Para a construção do nosso modelo de previsão da taxa de crescimento do PIB do Rio Grande do Sul, são utilizados como regressores as seguintes séries mensais: total de vendas

internas, total de vendas pra outros Estados, exportações, total compras internas, total compras de outros Estados, importação, arrecadação de ICMS, taxa de arrecadação efetiva e valor adicionado das operações . Essas informações estão disponíveis no banco de dados da Receita Estadual do Rio Grande do Sul (com exceção do PIB do RS que foi obtido do Departamento de Economia da Secretaria de Planejamento do Estado), a motivação da escolha dessas variáveis se deve, principalmente, pela velocidade com essas informações estão disponíveis, pois sua origem está na emissão das notas fiscais eletrônicas pelos contribuintes do ICMS <sup>1</sup>, suas obtenções são possíveis em até 3 dias após encerramento do mês, diferente de outras séries temporais como o PIM, CAGED ou o IBC que estão disponíveis em aproximadamente dois meses após o encerramento do mês analisado, o que dificulta a implementação do conceito de *nowcasting*.

Para esse trabalho são consideradas para as variáveis respostas dados referentes à taxa (%) de crescimento do PIB trimestral entre o 2º trimestre de 2003 e ao 1º trimestre de 2019. As previsões para os períodos fora da amostra são comparadas com aquelas do benchmark ARMA e abrangem o período do 2º trimestre de 2019 ao 1º trimestre de 2022, tanto para os modelos com apenas um regressor quanto para os modelos com múltiplos regressores.

Nossos principais resultados indicam que os modelos que adotam parâmetros variantes no tempo (midas *rolling-regression* e midas *recursive-regression*) apresentam desempenho superiores aos modelos que adotam parâmetros fixos no tempo (midas, ar-midas, umidas e ar-midas), principalmente neste contexto da pandemia Covid-19, com destaque para o midas *rolling-regression* (sem o componente autoregressivo) que teve o melhor desempenho entre todos os modelos . De maneira geral, em relação à regressão MIDAS, os resultados deste trabalho estão de acordo com a literatura, demonstrando a superioridade dos modelos midas e ar-midas em relação aos modelos umidas e ar-umidas, bem como uma melhor performance em relação benchmark ARMA (GALVÃO; CLEMENTS, 2008, 2009; GHYSELS *et al.*, 2004, 2007). Para os resultados de *nowcasting*, todos os modelos estudados apresentam ganhos, com exceção dos modelos umidas e ar-umidas, conforme são acrescentados meses do trimestre corrente na previsão, embora esse ganho tenha sido encontrado apenas para os regressores de vendas internas, vendas para outros estados, compras internas e compras para outros estados. Este ganho com o *nowcasting* foi obtido tanto para os modelos com um ou múltiplos regressores.

Para além desta introdução, este trabalho apresenta, na seção 2, a metodologia utilizada nos diferentes modelos descrita de maneira detalhada; na seção 3, a base de dados, os testes de raiz unitária e quebras estruturais, as análises empíricas e os testes de *Model Confidence Set*. Por fim, na seção 4, são feitas as conclusões e as considerações finais.

---

<sup>1</sup> De maneira geral o ICMS não considera serviços em sua base de cálculo, razão pela qual a variável serviço não foi considerada na composição do modelo. A obtenção dessa variável de uma fonte externa (municípios), não atendia o quesito de celeridade na obtenção dessas informações buscado nesse trabalho. Além disso, a variável serviço acaba sendo considerada implicitamente quando levamos em consideração o fator autoregressivo do PIB nos modelos, já que o serviço é uma das variáveis que compõe o PIB.

## 2 METODOLOGIA

Neste t3pico 3 desenvolvimento o modelo MIDAS cl3ssico com suas deriva33es, com destaque para as duas inova33es desenvolvidas ao longo desse trabalho, que s3o: o *MIDAS-Rolling Regression* que trabalha com sub-amostras (regress3o por janelas) para uma regress3o de m3nimos quadrados e o Modelo *MIDAS-Recursive Regression*.

### 2.1 ABORDAGEM MIDAS e UMIDAS

Nos dois pr3ximos t3picos deste trabalho s3o explanados os modelos MIDAS e UMIDAS, considerados as abordagens tradicionais desenvolvidas por Ghysels *et al.* (2004).

#### 2.1.1 Midas B3sico

Ghysels *et al.* (2004,2007) proporam uma nova classe de modelos denominada MIDAS (*Mixed Data Sampling*), que permite que vari3veis dependentes e independentes sejam relacionadas em diferentes frequ3ncias. Nesse modelo, as vari3veis explicativas s3o ponderadas por fun33es que utilizam um pequeno n3mero de par3metros para serem estimados, isso garante uma maior flexibilidade ao modelo (ZUANAZZI; ZIEGELMANN, 2014; GALV3O; CLEMENTS, 2008).

Segundo Mariano e Ozmucur (2020), as fun33es parcimoniosas mais utilizadas que garantem essa flexibilidade s3o as fun33es exponencial de almon, exponencial beta, *Step Function* (cujos pesos s3o iguais para as defasagens do mesmo trimestre), almon exponencial e Irrestritos.

3 adotada neste trabalho a not33o de Galv3o e Clements (2008) para a regress3o MIDAS com uma vari3vel explicativa e  $h$  passos 3 frente para previs3o, sendo dada por

$$y_t = \beta_0 + \beta_1 B(L^{1/m}; \theta) x_{t-h}^{(m)} + \epsilon_t, \quad (2.1)$$

onde  $B(L^{1/m}; \theta) = \sum_{k=1}^K b(k; \theta) L^{(k-1)/m}$ , sendo  $L^{s/m} x_t^{(m)} = x_{t-1-s/m}^{(m)}$ . Nesse modelo,  $b(j; \theta)$  representa uma fun33o parcimoniosa de pesos, sendo  $\theta$  o vetor de par3metros dessa fun33o. Logo, o 3ndice  $t$  apresentado refere-se 3 unidade de tempo b3sico da vari3vel de baixa frequ3ncia (vari3vel resposta), sendo  $m$  o n3mero de observa33es em  $t$ .  $B(L^{1/m})$  3 um polin3mio de grau  $K$ , sendo  $K$  o n3mero de defasagens de  $x_t^{(m)}$  que impactam na previs3o da vari3vel resposta ( $y_t$ ). J3 os par3metros  $\beta_0$ ,  $\beta_1$  e  $\theta$  s3o estimados por m3nimos quadrados n3o lineares,  $\epsilon_t$  3 um ru3do branco (ZUANAZZI; ZIEGELMANN, 2014).

Para melhor compreens3o da equa33o 2.1, suponha que tenhamos os dados do PIB trimestral do Rio Grande do Sul e as vari3veis explicativas tenham frequ3ncia mensal, ent3o temos  $m = 3$ , e que o n3mero de defasagens de nosso modelo seja de quatro meses, logo  $k = 4$ .

Caso a previsão seja um passo a frente, tomamos que  $h = 1$ . Neste caso, a equação 2.1 pode ser expressa como

$$y_t = \beta_0 + \beta_1 B(L^{1/3}; \theta) x_{t-1}^{(3)} + \epsilon_t. \quad (2.2)$$

Desse modo, temos que  $B(L^{1/3}; \theta) = \sum_{k=1}^4 b(k; \theta) L^{(k-1)/3}$  e  $L^{(s/3)} x_{t-1}^{(3)} = x_{t-1-s/3}^{(3)}$  e como resultado encontramos

$$y_t = \beta_0 + \beta_1 [b(1; \theta) x_{t-1}^{(3)} + b(2; \theta) x_{t-1-1/3}^{(3)} + b(3; \theta) x_{t-1-2/3}^{(3)} + b(4; \theta) x_{t-2}^{(3)} + \epsilon_t, \quad (2.3)$$

podemos interpretar a equação 2.3 da seguinte forma, se nosso objetivo é a previsão do PIB do 2º trimestre de 2021 e temos informações até março/2021, então  $x_{t-1}^{(3)}$  representa o mês de março de 2021,  $x_{t-1-1/3}^{(3)}$  é fevereiro de 2021,  $x_{t-1-2/3}^{(3)}$  é janeiro de 2021 e, por último, como temos quatro defasagens,  $x_{t-2}^{(3)}$  é dezembro de 2020.

A parametrização dos coeficientes de defasagens  $B(L^{1/m}; \theta)$  em uma função polinomial parcimoniosa é uma das principais características da regressão MIDAS. Ghysel *et al.* (2007) detalham várias alternativas para especificação deste polinômio, sendo a função de defasagens exponencias de almon, introduzida por Almon (1965), a mais comum, a qual é representada por

$$b(k; \theta) = \frac{e^{\theta_1 k + \dots + \theta_Q k^Q}}{\sum_{k=1}^K e^{\theta_1 k + \dots + \theta_Q k^Q}}, \quad (2.4)$$

onde ela é a escolha mais comum para especificação do polinômio  $B(L^{1/m}; \theta)$ , já que com apenas dois parâmetros ( $\theta_1$  e  $\theta_2$ , quando  $Q = 2$ ) é possível obter diversos formatos para o vetor de peso das defasagens, não impactando negativamente o número de graus de liberdade do modelo. Dependendo dos valores dos parâmetros escolhidos, é possível obter uma queda mais acentuada ou lenta dos pesos conforme as defasagens aumentam (ZUANAZZI; ZIEGELMANN, 2014; GUERIN; MARCELLINO, 2013).

### 2.1.2 Midas autoregressivo: AR - MIDAS

Segundo Ghysels *et al.* (2007), Galvão e Clement (2008, 2009) e Zuanazzi e Ziegelmann (2014), modelos de previsão do crescimento do PIB geralmente possuem um componente autorregressivo junto com suas variáveis explicativas, mas a simples inclusão de uma defasagem trimestral ao modelo, por exemplo a  $y_{t-1}$ , não é adequada para um modelo de previsão, já que o modelo ficaria com a seguinte estrutura:

$$y_t = \beta_0 + \lambda y_{t-1} + \beta_1 B(L^{1/3}; \theta) x_{t-1}^{(3)} + \epsilon_t, \quad (2.5)$$

portanto, essa equação, a 2.5, não é a mais adequada, visto, também, que o modelo poderia ser reescrito da seguinte forma

$$y_t = \beta_0(1 - \lambda)^{-1} + \beta_1(1 - \lambda)^{-1}B(L^{1/3}; \theta)x_{t-1}^{(3)} + \tilde{\epsilon}_t, \quad (2.6)$$

em que  $\tilde{\epsilon}_t = (1 - \lambda L)^{-1}\epsilon_t$ . É evidente que o polinomial em  $x_{t-1}^{(3)}$  é produto de um polinomial em  $L^{1/3}$  e um polinomial em  $L$ ,  $\sum \lambda^j L^j$ , que gera uma resposta "sazonal" de  $y$  para  $x^{(3)}$ , independentemente de  $x^{(3)}$  apresentar um padrão de sazonalidade.

Uma solução simples apresentada por Galvão e Clements (2008) foi a introdução de um fator comum para eliminar o padrão de sazonalidade, de tal forma que:

$$y_t = \beta_0 + \lambda y_{t-1} + \beta_1 B(L^{1/3}; \theta)(1 - \lambda L)x_{t-1}^{(3)} + \epsilon_t. \quad (2.7)$$

Um modelo para várias defasagens seria dado por

$$y_t = \beta_0 + \lambda y_{t-h} + \beta_1 B(L^{1/m}; \theta)(1 - \lambda L^h)x_{t-h}^{(m)} + \epsilon_t. \quad (2.8)$$

### 2.1.3 Modelo ar-midas com múltiplo regressores

Modelos preditivos com apenas um regressor geram desconfiças em seus resultados, enquanto modelos com múltiplos regressores apresentam, em geral, melhores resultados preditivos. O modelo MIDAS, especificado na equação 2.1, possui a vantagem de que a inclusão de múltiplos regressores não aumenta o número de parâmetros a serem estimados (GALVÃO; CLEMENTS, 2008).

O modelo M-MIDAS-AR, que combina a informação de múltiplos regressores para  $h$  passos à frente da previsão do PIB, pode ser escrito da seguinte forma:

$$y_t = \beta_0 + \lambda y_{t-h} + \sum_{i=1}^n \beta_i B(L^{1/m}; \theta_i)(1 - \lambda L^h)x_{i,t-h}^{(m)} + \epsilon_t, \quad (2.9)$$

nesta equação os regressores são indexados por  $i$  e  $m = 3$  (pois está sendo analisado o PIB trimestral gaúcho). Cada variável explicativa deste modelo exige apenas a estimação de apenas dois parâmetros para suas estruturas de defasagens ( $\theta_i$ ) e um parâmetro apenas para ponderar seu impacto no PIB gaúcho ( $\beta_i$ ). Como o número de parâmetros exigidos para estimação é pequeno, com a inclusão de regressores adicionais, pode-se admitir, previamente, que o modelo MIDAS possui uma melhor performance preditiva quando comparado a outros modelos que consideram as defasagens de maneiras individual, visto que estas são consideradas individualmente, o que leva a inclusão de um grande número de parâmetros para acomodar as variáveis explicativas dentro do modelo (GALVÃO; CLEMENTS, 2008, 2009).

### 2.1.4 Unconstrained MIDAS:UMIDAS

Conforme Zuanazzi e Ziegelmann (2014), o modelo UMIDAS é um alternativa à aplicação MIDAS, pois, além de levar em consideração todas as informações de  $x_t^{(m)}$  para prever

$y_t$ , ele não restringe que os pesos para as defasagens do regressor estejam ligados a uma função polinomial, que pode ser flexível o suficiente para representar os verdadeiros pesos gerados pelos dados aleatórios. Um dos prejuízos desse modelo, todavia, é a proliferação do número de parâmetros para os regressores, o que acaba diminuindo o número de graus de liberdade para estimação do modelo, principalmente quando a diferença nas frequências de  $x^{(m)}$  e  $y_t$  é elevada, ou seja,  $m$  é grande.

Foroni *et al.* (2011) aplicou a regressão *unconstrained mixed data sampling* (UMIDAS) para previsões de crescimento do PIB dos EUA e identificou que este modelo apresentou melhores resultados que o MIDAS. Entretanto, isso acontece apenas se não houver uma maior parametrização dos regressores, gerada por dados com maior frequência e com amostragem muito grande, isto é, apenas se eles não forem muito parametrizados e, em particular, apenas, também, se as diferenças nas frequências de amostragem não forem muito elevadas.

Adotando a notação de Foroni *et al.* (), o modelo U-MIDAS pode ser escrito como:

$$y_t = \beta_0 + \sum_{k=1}^K \beta_k L^{\frac{(k-1)}{m}} x_{t-1}^{(m)} + \epsilon_t, \quad (2.10)$$

nesta equação temos que cada parâmetro é alocado para cada observação de  $x_t^{(m)}$  dentro de cada período  $t$ , de maneira que as observações de  $x_t^{(m)}$  não tenham todas a mesma ponderação, dependendo de sua defasagem. Desse modo, serão estimados  $K + 1$  parâmetros para o modelo com  $K$  de defasagens.

Uma derivação do U-MIDAS denominada AR-UMIDAS, representada de maneira mais genérica, com a inclusão do fator auto-regressivo como mais uma variável explicativa e com suas defasagens, pode ser escrita da seguinte maneira:

$$y_t = \beta_0 + \sum_{d=1}^D \lambda_d y_{t-d} + \sum_{v=1}^V \sum_{k=1}^K \beta_{v,k} L^{\frac{(k-1)}{m}} x_{v,t-1}^{(m)} + \epsilon_t. \quad (2.11)$$

Assim, nota-se que  $V$  é o número de variáveis independentes no modelo e  $\lambda_d$  é o parâmetro autorregressivo da  $d$ -ésima defasagem da variável dependente  $y_t$ .

## 2.2 ESTIMAÇÃO

Ghysels *et al.* (2004) demonstrou o modelo MIDAS pode ser estimado por máximo verossimilhança (MLE), por mínimos quadrados não lineares (MQNL), pelo métodos dos momentos não generalizados (MMG), sendo o MQNL um método de estimação consistente para o padrão MIDAS (GALVÃO; CLEMENTS, 2008). Neste trabalho será utilizado o MQNL e também serão utilizados os seguintes algoritmos para otimização numérica dos parâmetros: algoritmo de

Broyden–Fletcher–Goldfarb–Shanno (BFGS) e nlm <sup>1</sup>.

Para estimar o modelo AR-MIDAS (equação 2.8), é utilizado o algoritmo proposto por Galvão e Clements (2008): primeiro, será tomado o resíduo ( $\hat{\epsilon}_t$ ) do modelo MIDAS e, então, será estimado um valor inicial para  $\lambda$  (chamado de  $\lambda_0$ ) de  $\lambda_0 = (\sum \hat{\epsilon}_{t-h})^{-1} \sum \hat{\epsilon}_t \hat{\epsilon}_{t-h}$ . Depois será construída a seguinte equação  $y^* = y_t - \hat{\lambda}_0 y_{t-d}$  e  $x_{t-h}^{*(3)} = x_{t-h}^{(3)} - \hat{\lambda}_0 x_{t-h-d}^{(3)}$ , obtendo, por conseguinte, o estimador  $\hat{\theta}_1$ , pela aplicação do MQNL para:

$$y_t^* = \beta_1 + \beta_2 B(L^{1/3}, \theta) x_{t-h}^{*(3)} + \epsilon_t, \quad (2.12)$$

obtém-se, dessa forma, um novo valor para  $\lambda$ ,  $\hat{\lambda}_1$  a partir dos resíduos da equação 3.12. Assim, usando os valores obtidos de  $\hat{\lambda}_1$  e  $\hat{\theta}_1$  como valores iniciais e por meio dos algoritmos de otimização (BFGS e nlm), serão obtidos os estimadores  $\hat{\lambda}$  e  $\hat{\theta}$ , que tendem a minimizar a soma dos quadrados dos resíduos.

Já para os modelos UMIDAS, segundo Ghysels (2020), tem-se a seguinte representação:

$$y_t = \beta_0 + \sum_{i=0}^K b_i x_{s(t)-i} + \epsilon_t. \quad (2.13)$$

Nela temos  $k + 2$  parâmetros irrestritos  $\beta_u(\beta_0, b_0, b_1, b_2 \dots b_k)'$  a serem estimados na função de regressão. Eles, ademais, podem – sob a lineariedade da função de regressão do vetor de parâmetros – ser estimados pelo método de mínimos quadrados ordinários (OLS)

$$\hat{\beta}_u = \underset{\beta_0, b_0, \dots, b_k}{\operatorname{argmin}} \sum_t (y_t - \beta_0 - \sum_{i=0}^k b_i x_{s(t)-i})^2 = \left( \sum_{t=1}^T x_{s(t)} x_{s(t)}' \right)^{-1} \sum_{t=1}^T x_{s(t)} y_t, \quad (2.14)$$

onde  $x_{s(t)} = (1, x_{s(t)-1}, \dots, x_{s(t)-k})$ .

## 2.3 REGRESSORES VARIANTES NO TEMPO

Neste trabalho são introduzidos dois novos métodos derivados da abordagem MIDAS e AR-MIDAS, denominados *MIDAS-Rolling Regression* e *MIDAS-Recursive regresison*, que são modelos que trabalham com parâmetros variantes ao longo do tempo. A primeira trabalha com regressão por janelas; a segunda, com regressão recursiva.

### 2.3.1 Matriz de covariáveis

Para esses dois novos modelos, é necessária a criação de uma matriz de covariáveis. Por isso, são tomados como ponto de partida os relevantes trabalhos de Guerin e Marcelino (2013) e Marsilli (2014), que desenvolveram, respectivamente, as abordagens MS-MIDAS (Markov

<sup>1</sup> Esses são os dois algoritmos de otimização numérica disponíveis para utilização pelo pacote *midasrs*, desenvolvido por Kvedaras e Balevicius (2021), que será a base deste trabalho

Switching Midas) e MIDAS-LASSO. Esses autores tratam a abordagem midas como um modelo de Regressão Linear padrão que possui um núcleo não-linear (função parcimoniosa de defasagens), que possibilita agregar com alto grau de flexibilidade as variáveis explicativas de maior frequência e suas defasagens – isso em um vetor de variáveis explicativas na mesma frequência da variável resposta. Ambos trabalhos fizeram a estimação dos parâmetros  $\theta_i$  da função de defasagens de almon conforme a abordagens tradicionais desenvolvidas nos trabalhos de Ghysels *et al.* (2005,2004), que já foram tratados aqui, no tópico 2.2, com o parâmetro  $\theta_i$  mantido fixo e, a partir disso, sabe-se que buscaram melhorar a performance do modelo por meio de mudanças na estimativa dos parâmetros  $\beta_i$ .

Deste modo, em um primeiro momento, são estimados os parâmetros  $\theta_i$  da regressão midas, para depois os substituir na função exponencial de almon (equação 2.4) e obter os coeficientes de cada uma das defasagens. Em um segundo momento, para obter a matriz de covariáveis de baixa frequência  $X(x, \theta_i)$ , é realizada a multiplicação da matriz de regressores de alta frequência (com suas defasagens) pelo vetor dos coeficientes das defasagens obtidos por meio da função de defasagens de almon, de forma que se tem

$$X(x, \theta) = \begin{bmatrix} x_{1,1}^{k1} & \cdots & x_{1,n}^{kn} \\ \vdots & \ddots & \vdots \\ x_{T,1}^{k1} & \cdots & x_{T,n}^{kn} \end{bmatrix}_{(n \times k)} \times \begin{bmatrix} b(k^1, \theta_1, \theta_2) \\ \vdots \\ b(k^n, \theta_1, \theta_2) \end{bmatrix}_{k \times 1}, \quad (2.15)$$

e, por conseguinte, tem-se

$$X(x, \theta) = \begin{bmatrix} x_{1,1}^{k1} * b(k^1, \theta_1, \theta_2) + \dots + x_{1,n}^{kn} * b(k^n, \theta_1, \theta_2) \\ \vdots \\ \vdots \\ x_{T,1}^{k1} * b(k^1, \theta_1, \theta_2) + \dots + x_{T,n}^{kn} * b(k^n, \theta_1, \theta_2) \end{bmatrix}_{(n \times 1)}. \quad (2.16)$$

Assim, a primeira matriz à direita da equação acima representa a matriz de variáveis explicativas de alta frequência de dimensões  $n \times k$ ,  $n$  é o número de observações e  $k$  é o número de defasagens. Ademais, o segundo vetor a direita é os coeficientes de defasagens, obtidos por meio da equação 2.4, de dimensões  $k \times 1$ .

Após essa transformação, a fim de que se encontre a estimação do crescimento trimestral do PIB do Rio Grande do Sul, tem-se a seguinte equação, que é utilizada para desenvolver a regressão recursiva e a regressão por janelas:

$$y_t = \beta_{0,t} + \beta_{i,t} X(x, \theta_i)_t + \epsilon_t. \quad (2.17)$$

### 2.3.2 Regressão por janela

As regressões por janelas usam a regressão linear, mas permitem que o conjunto de dados usado mude ao longo do tempo, já que sempre são realizadas regressões com subamostras (ou janelas). Dessa forma, tal flexibilidade temporal garante melhores respostas em períodos de intensas instabilidades.

Cai e Juhl (2021) estudaram modelos de persistências da inflação mensal dos Estados Unidos no período de fevereiro de 1947 até maio de 2020, por meio de modelos de regressão lineares, e adotaram a regressão por janelas para reduzir o viés dos parâmetros. Em Rubio-Herro e Wang (2021), eles desenvolveram um modelo linear reflexível, por meio de regressão por janelas, para acompanhar e monitorar os casos de infecção, internação, recuperação e mortes no Estado de Minnesota, encontrando resultados consistentes com as mudanças na política governamental, com a adoção de medidas de distanciamento e proteção individual, e com a melhoria da capacidade dos serviços de saúde, sistema para lutar contra esse novo vírus (covid-19).

São chamados, neste trabalho, de M-MIDAS-ROLL os modelos em que houve a regressão por janelas realizada a partir da equação 2.17 e com múltiplos regressores, e de AR-M-MIDAS-ROLL os modelos em que houve a regressão por janelas realizada a partir da equação 2.17, com múltiplos regressores e um componente autorregressivo.

Para desenvolver os modelos M-MIDAS-ROLL e AR-M-MIDAS-ROLL, é usada como base a metodologia adotada por Nakajima *et al.* (2011), em que, para que ocorra a regressão por janelas, repete-se uma regressão de mínimos quadrados lineares comuns, mas adotando uma janela que altera seus pontos fixos de começo e fim ao longo de toda a amostra. Dessa forma, tomando como base a equação 2.17 para obtermos o modelo M-MIDAS-ROLL, tem-se, então:

$$y_t = \beta_{i,t(h)} X(x, \theta_i)_{t(h)} + \epsilon_{t(h)}, \quad (2.18)$$

portanto,  $y$  é o PIB taxa de crescimento do Rio Grande do Sul,  $X(x, \theta_i)_{t(h)}$  é a matriz de covariáveis estimadas em 2.15 e 2.16,  $\beta_{i,t(h)}$  é o parâmetro e  $\epsilon_t(h)$  é o termo de erro e, além disso,  $h = 1, \dots, n$ . Como estamos tratando de modelos com parâmetros variantes no tempo, logo  $\beta_{i,t(h)}$  sofre alterações conforme a janela (ou intervalo de observações) que está sendo utilizada. A estimação do parâmetro  $\beta_{i,t(h)}$  é obtido por meio de:

$$\beta_{i,t(h)} = \min_{\beta_{i,t(h)}} \sum (y_t - \beta_{i,t(h)} X(x, \theta_i)_{t(h)})^2, \quad (2.19)$$

por conseguinte,  $n$  é o número de regressões e parâmetros possíveis de se obter para um número total de observações, que é chamado de  $N$ , e o tamanho da janela será dado por  $w$ . Nesse sentido,  $d$  é o conjunto de observações que ficaram fora da janela, então  $n$  pode ser obtido por meio da seguinte equação:

$$n = \left\lceil \frac{N-w}{d} \right\rceil + 1. \quad (2.20)$$

Na maioria dos casos, as regressões por janelas são conduzidas aumentando os pontos de começo e fim das subamostras, mantendo suas janelas fixas. Desse modo, temos que  $t(h) \in [h, h + w - 1]$ . Por exemplo, considerando que o número total de observações seja de 500 e colocando uma janela de tamanho 100, a regressão seria feita inicialmente usando os períodos de 1 a 100, seguidos de 2-201, 3-102,..., até 401-500. O número total de regressões realizados para cada parâmetro seria de um total de 401. Desse modo, é possível obter parâmetros variantes no tempo.

Segundo Su e Hwang (2009), os parâmetros de uma regressão por janela dependem, de maneira geral, do intervalo de cada regressão e do tamanho da janela. Quanto o maior intervalo utilizado, mais ampla deve ser a mudança nos parâmetros ao longo da regressão. É recomendado que pequenos intervalos – por exemplo, um – sejam escolhidos de acordo com a habilidade de processamento do seu computador. Já para os tamanhos da janela, com uma janela menor, mais detalhado será o movimento de mudança dos parâmetros ao longo do tempo, embora a variância do parâmetro seja maior.

Para obtermos o modelo AR-M-MIDAS-ROLL, combinamos a equação 2.9 e 2.17 para inclusão do termo autorregressivo:

$$y_t = \lambda y_{t-h} + \beta_{i,t(h)} X(x, \theta_i)_{t(h)} + \epsilon_{t(h)}. \quad (2.21)$$

Neste trabalho, são utilizadas as janelas que têm sempre o menor RFSME para os períodos de previsão fora da amostra. Nos testes empíricos as janelas utilizadas foram de 30 meses para variáveis de alta frequência ou 10 meses para as variáveis de baixa frequência.

### 2.3.3 Regressão recursiva

Os mínimos quadrados recursivos são um algoritmo de filtro adaptativo que recursivamente encontra os coeficientes que minimizam uma função de custo de mínimo quadrado linear ponderada em relação aos sinais de entrada. Em Silveira e Pereira Junior (2021), foi desenvolvido um algoritmo de mínimos quadrados recursivos para acompanhar e estudar o grau de infecciosidade dos indivíduos afetados pela COVID-19, considerando os impactos causados ao longo da pandemia em oito países diferentes: Argentina, Bélgica, Brasil, Alemanha, Itália, Nova Zelândia, Espanha e Estados Unidos. A partir disso, concluiu-se que os modelos de Mínimos Quadrados Recursivos performaram melhor do que os modelos de Mínimos Quadrados Ordinários, já que o processamento recursivo ou sequencial desses exige menos memória em uma única etapa de interação, possibilitando capturar a dinâmica de sistemas não-lineares e variantes no tempo.

São chamados, neste trabalho, de M-MIDAS-REC os modelos em que houve a regressão recursiva, realizada a partir da equação 2.17 e com múltiplos regressores, e de AR-M-MIDAS-

REC os modelos em que houve a regressão recursiva, realizada a partir da combinação da equação 2.9 e 2.17, com múltiplos regressores e um componente autorregressivo.

Para os modelos de regressão recursiva, é adotada a metodologia de Coelho (2004) e Rocha () para a construção do estimador de mínimos quadrados recursivos a partir da equação 2.17 para obtermos os modelos M-MIDAS-REC, e a combinação da 2.9 e 2.17 para obtermos o modelo AR-M-MIDAS-REC, respectivamente, tendo como intuito comparar a estimativa baseada nas medidas num determinado instante de tempo que varia de 1 até  $n$ , com as estimativas baseadas nas médias nos instantes de 1 até  $n + 1$ . Essa metodologia permite que o estimador do parâmetro  $\beta$  da equação 2.17 se adapte em tempo real, à medida em que as séries sofram mudanças (ou a economia apresente mudanças).

Tomemos que  $\beta$  seja obtido por meio da seguinte expressão:

$$\beta_n = [X(x, \theta_i)_n^T X(x, \theta_i)_n]^{-1} X(x, \theta_i)_n^T Y_n, \quad (2.22)$$

dispondo de uma nova observação  $(x_{n+1}, y_{n+1})$ , é possível derivar  $\beta_{n+1}$ , por meio da atualização de  $\beta_n$ , partindo das  $n + 1$  observações. Quando adicionamos uma nova observação  $(x_{n+1}, y_{n+1})$  ao conjunto de dados, a matriz  $X$  tem uma nova linha, o vetor  $Y$  uma nova componente, e podemos escrever da seguinte forma:

$$\beta_{n+1} = \left( \begin{bmatrix} X(x, \theta_i)_n \\ x(x, \theta_i)_{n+1} \end{bmatrix}^T \begin{bmatrix} X(x, \theta_i)_n \\ x(x, \theta_i)_{n+1} \end{bmatrix} \right)^{-1} \begin{bmatrix} X(x, \theta_i)_n \\ x(x, \theta_i)_{n+1} \end{bmatrix}^T \begin{bmatrix} Y_n \\ y_{n+1} \end{bmatrix}. \quad (2.23)$$

Definindo  $S_N$  como

$$S_N = (X_N^T X_N), \quad (2.24)$$

tem-se, portanto, que:

$$S_{n+1} = (X(x, \theta_i)_{n+1}^T X(x, \theta_i)_{n+1}) = \left( (X(x, \theta_i)_n^T x(x, \theta_i)_n \begin{bmatrix} X(x, \theta_i)_n \\ x(x, \theta_i)_{n+1} \end{bmatrix} \right), \quad (2.25)$$

simplificando, sabe-se que:

$$S_{n+1} = X(x, \theta_i)_n^T X(x, \theta_i)_n + x(x, \theta_i)_{n+1}^T x(x, \theta_i)_{n+1}, \quad (2.26)$$

então,

$$\begin{bmatrix} X(x, \theta_i)_n \\ x(x, \theta_i)_{n+1} \end{bmatrix}^T \begin{bmatrix} Y_n \\ y_{n+1} \end{bmatrix} = X(x, \theta_i)_n^T Y_n + x(x, \theta_i)_{n+1} y_{n+1}, \quad (2.27)$$

e,

$$S_n \beta_n = X(x, \theta_i)_n^T Y_n, \quad (2.28)$$

e, ainda:

$$\begin{bmatrix} X(x, \theta_i)_n \\ x(x, \theta_i)_{n+1} \end{bmatrix}^T \begin{bmatrix} Y_n \\ y_{n+1} y_{n+1} \end{bmatrix} = S_n \beta_n + x(x, \theta_i)_{n+1}, \quad (2.29)$$

simplificando, tem-se 2.29,

$$= (S_{n+1} - x(x, \theta_i)_{n+1}^T x(x, \theta_i)_{n+1}) \beta_n + x(x, \theta_i)_{n+1} y_{n+1}. \quad (2.30)$$

Ao valer-se das igualdades 2.20, 2.21, 2.23, 2.24 e 2.25, pode-se escrever:

$$\begin{aligned} \beta_{n+1} &= S_{n+1}^{-1} (X(x, \theta_i)_n^T Y_n + x(x, \theta_i)_{n+1} y_{n+1}) = S_{n+1}^{-1} (S_n \beta_n + x(x, \theta_i)_{n+1} y_{n+1}) = \\ &= S_{n+1}^{-1} [(S_{n+1} - x(x, \theta_i)_{n+1}^T x(x, \theta_i)_{n+1}) \beta_n + x(x, \theta_i)_{n+1} y_{n+1}] = \\ &= \beta_n + S_{n+1}^{-1} x(x, \theta_i)_{n+1}^T x(x, \theta_i)_{n+1} \beta_n + S_{n+1}^{-1} x(x, \theta_i)_{n+1} y_{n+1} = \\ &= \beta_n + S_{n+1}^{-1} x(x, \theta_i)_{n+1}^T (y_{n+1}) - x(x, \theta_i)_{n+1} \beta_n \end{aligned}$$

Chega-se à seguinte formulação recursiva:

$$\begin{cases} S_{n+1} = S_n + x(x, \theta_i)_{n+1}^T x(x, \theta_i)_{n+1} \\ \gamma_{n+1} = S_{n+1}^{-1} x(x, \theta_i)_{n+1} \\ e = y_{n+1} - x_{n+1} \beta_n \\ \beta_{n+1} = \beta_n + \gamma_{n+1} e \end{cases}$$

Conseguimos, desse modo, escrever  $\beta_{n+1}$ , em função de  $\beta_n$  e da nova observação  $(x_{n+1}, y_{n+1})$ . Nota-se, assim, que a inversão da matriz  $S(n+1)$  seria uma operação custosa a qual computacionalmente se pode utilizar o lema dado anteriormente para solucionar esse problema. Ao designar-se o Designando  $S_n^{-1} = (X_n^T X_n)^{-1}$  por  $V(n)$ , através do lema de inversão matricial de Sherman-Morrinson-Woodbury, obtém-se:

$$\begin{aligned} V_{n+1} &= V_n - V_n x(x, \theta_i)_{n+1} (1 + x(x, \theta_i)_{n+1}^T V_n x(x, \theta_i)_{n+1})^{-1} x(x, \theta_i)_{n+1}^T V_n = \\ &= V_n - \frac{V_n x(x, \theta_i)_{n+1} x(x, \theta_i)_{n+1}^T V_n}{1 + x(x, \theta_i)_{n+1}^T V_n x(x, \theta_i)_{n+1}} \end{aligned}$$

Chega-se, logo, ao seguinte sistema de equações:

$$\begin{cases} V_{n+1} = V_n - \frac{V_n x(x, \theta_i)_{n+1} x(x, \theta_i)_{n+1}^T V_n}{1 + x(x, \theta_i)_{n+1}^T V_n x(x, \theta_i)_{n+1}} \\ \gamma_{n+1} = S_{n+1}^{-1} x(x, \theta_i)_{n+1} \\ e = y_{n+1} - x_{n+1} \beta_n \\ \beta_{n+1} = \beta_n + \gamma_{n+1} e \end{cases}$$

O modelo para obtenção do algoritmo de Regressão Recursiva é um caso especial do Filtro de Kalman sobre um modelo Estado-Espaço descrito no conjunto de equações descrito no espaço.

Por meio da solução do sistema, obtemos dois resultados importantes,  $V_{n+1}$ , que obtém o resultado da função de ganho do estimador, e, também, o próprio parâmetro variante no tempo dado por  $\beta_{n+1} = \beta_n + \gamma_{n+1}e$ .

### 2.3.4 Regressores fixos no tempo

Para os regressores que não variam no tempo, utilizamos a denominação TRA, a fim diferenciá-los dos demais modelos. Esses, por sua vez, foram estimados por meio da regressão MIDAS tradicional.

## 2.4 PREVISÃO COM INFORMAÇÕES DO TRIMESTRE CORRENTE

As previsões do PIB são produzidas com frequência trimestral, sendo também realizadas previsões quando os dados mensais relativos àquele trimestre estão disponíveis. Nos métodos Midas, Ar-midas, Umidas, Ar-umidas, Midas *rolling regression* e Midas *recursive regression*, consideramos quatro estimativas diferentes para as observações de  $x_t^{(m)}$ , conforme as informações estão disponíveis para os horizontes ( $h = 1, h = 2/3, h = 1/3$  e  $h = 0$ ) dentro do respectivo trimestre a ser previsto. Neste trabalho, chamamos de  $h = 1$  o período em que são utilizadas informações até o último mês do trimestre anterior ao que iremos fazer a previsão. Além disso, é em  $h = 2/3$  quando utilizamos o primeiro mês do trimestre em que estamos realizando a previsão. Quando falamos de  $h = 1/3$ , são utilizadas informações dos dois primeiros meses do trimestre em que estamos realizando a previsão. Em  $h = 0$ , estamos utilizando todos os meses do trimestre em que estamos realizando a previsão.

A diferença existente entre cada um dos horizontes se dá na maneira em que a matriz de regressores é constituída – por exemplo, quando falamos em  $h = 1/3$ , são utilizados dois meses dentro do trimestre a ser previsto, sendo que o segundo mês desse trimestre é considerado a primeira defasagem, e o primeiro mês desse trimestre é a segunda defasagem. Agora, quando falamos em  $h = 0$ , o terceiro mês do trimestre a ser previsto é considerado a primeira defasagem, enquanto o segundo mês do trimestre é considerado a segunda defasagem. Em suma, conforme são acrescentadas as informações (meses) do trimestre corrente, eles ocupam a primeira posição na matriz de regressores.

Os parâmetros dos modelos com regressores individuais ou múltiplos são estimados conforme cada uma das novas informações (meses) chegam para o trimestre em que é realizada a previsão para os métodos apresentados nas equações. Para definição do número de defasagens dos modelos individuais ou com múltiplos regressores, foram escolhidos os de menores AIC e,

além disso, os intervalos de defasagem considerados variam do mínimo de 12 meses ao máximo de 23 meses.

### 3 ANÁLISE EMPÍRICA

Nos próximos tópicos deste trabalho são apresentados os testes de raiz unitária, os resultados para os modelos com um ou múltiplos regressores, bem como o teste Model Confidence Set.

#### 3.1 TESTE DE RAIZ UNITÁRIA

As séries econômicas utilizadas aqui são as vendas internas (*venda\_i*), vendas para outros estados (*venda\_o*), exportação, compras internas (*compras\_i*), compras de outros estados (*compra\_o*), importação, arrecadação de ICMS (*ICMS*) no estado, valor adicionado (*va*) da cadeia produtiva – que representa a diferença entre as compras e vendas – e taxa de arrecadação (*taxa*) – que se refere à arrecadação sobre as vendas. Todos esses valores (séries) são extraídos do banco de dados da Receita Estadual do Rio Grande do Sul. A representação gráfica das séries temporais estão disponíveis no Apêndice A deste trabalho.

Para a especificação dos modelos que são analisados ao longo deste trabalho, faz-se necessário, a princípio, verificar a existência de raízes unitárias e quebras estruturais. Para isso, foram utilizados o teste de Dickey-Fuller Aumentado (ADF) com várias defasagens, o teste de Phillips-Perron (PP), Teste KPSS (Kwiatkowski–Phillips–Schmidt–Shin) e o teste de Zivot e Andrews para raízes unitárias com quebras estruturais. A Tabela 01 apresenta os resultados dos testes ADF, PP e KPSS, enquanto, na Tabela 02, há os resultados do teste de Zivot e Andrews.

No teste ADF, para o nível de significância de 5%, foi rejeitada a hipótese nula de raiz unitária para as séries exportação, ICMS e taxa. Para as demais séries, não foi rejeitada a hipótese nula de raiz unitária. Nos testes PP, foi rejeitada a hipótese nula de raiz unitária para todas as séries, sendo todas consideradas estacionárias. No teste KPSS, a hipótese nula de tendência estacionária foi rejeitada para todas as séries analisadas.

Tabela 1 – Teste de raiz unitária

Testes de Raiz Unitária			
	Teste de ADF (p-valor)	TESTE PP (p-valor)	TESTE KPSS (p-valor)
<b>venda_i</b>	0,08*	0,01***	0,01***
<b>venda_o</b>	0,33	0,01***	0,01***
<b>exportação</b>	0,01***	0,01***	0,01***
<b>compra_i</b>	0,79	0,01***	0,01***
<b>compra_o</b>	0,24	0,01***	0,01***
<b>importação</b>	0,10*	0,01***	0,01***
<b>icms</b>	0,01***	0,01***	0,01***
<b>va</b>	0,29	0,01***	0,02**
<b>taxa</b>	0,01***	0,01***	0,04**

Fonte: Elaboração própria

**Nota:** I) a hipótese nula ( $H_0$ ) do teste ADF é que a série é não-estacionária, e a hipótese alternativa ( $H_1$ ) é que a série é estacionária; II) para o teste PP, temos que  $H_0$ : a série é estacionária,  $H_1$ : a série é não-estacionária; III) Para o teste KPSS  $H_0$ : a série é estacionária em torno de uma tendência determinística,  $H_1$ : possui raiz unitária; IV) Rejeita a 1% (\*\*\*) , a 5% (\*\*) ou a 10% (\*)

Análise dos gráficos contantes no apêndice A deste trabalho, demonstram quebras estruturais nos seus gráficos, pelo fato da natureza de nossas séries temporais serem econômicas e financeiras, elas estão sujeitas a possibilidade de quebras estruturais e, é por conta disso, que o poder dos testes explicativo de raiz unitária pode ser prejudicado. Ademais, na presença de quebras estruturais, os testes ADF, PP e KPSS são viesados em relação à estacionariedade ou não das séries (MARGARIDO, 2001; MORAIS; PORTUGAL, 2007; MATTOS, 2018). Dessa forma, a metodologia desenvolvida por Zivot-Andrews (1992), para o teste de raiz unitária, supera essa deficiência, levando em consideração a existência de quebra estrutural na identificação da tendência estacionária das séries (ZHOUE PENG, 2020).

As estatísticas calculadas para os testes de Zivot e Andrews (com "tendência", "constante" e "tendência e constante") rejeitaram a hipótese nula de presença de raiz unitária a 5% de significância para todas as séries analisadas.

Tabela 2 – Teste de Raiz Unitária com Quebras Estruturais de Zivot e Andrews (1992)

Série	Drift	Trend	Both
venda_i	-6,99***	-6,23***	-6,88***
venda_o	-6,64***	-5,12**	-6,59***
exportação	-6,68***	-6,32***	-6,68***
compra_i	-10,27***	-7,71***	-10,54***
compra_o	-6,26***	-6,15***	-6,51***
importação	-8,29***	-8,15***	-8,53***
taxa	-12,81***	-9,64***	-13,58***
va	-6,71***	-7,64***	-8,06***
icms	-11,49***	-10,78***	-12,25***

Fonte: Elaboração própria

Nota: Rejeita a 1% (\*\*\*), a 5% (\*\*) ou a 10% (\*)

Diante do exposto, as séries vendas internas, vendas para outros estados, exportação, compras internas, compras de outros estados, importação, arrecadação de ICMS, valor adicionado e taxa de arrecadação foram trabalhadas em nível. Todas as séries foram trabalhadas em sua forma logarítmica, com exceção da taxa de arrecadação que está em porcentagem.

Tabela 3 – Séries temporais

SÉRIES TEMPORAIS	UTILIZAÇÃO
venda_i	Em nível I(0)
vendas_o	Em nível I(0)
exportação	Em nível I(0)
Compra_i	Em nível I(0)
compra_o	Em nível I(0)
importação	Em nível I(0)
va	Em nível I(0)
ICMS	Em nível I(0)
taxa	Em nível I(0)

Fonte: Elaboração própria

Em relação a correlação entre as variáveis escolhidas, temos que elas todas elas possuem autocorrelação positiva, com exceção da série taxa de arrecadação. A série valor adicionado, embora positiva, tem uma baixa correlação (ou nula) em relação as demais variáveis. As maiores correlações estão entre as variáveis vendas internas e compras de outros estados, que apresentou um grau de correlação de 0,95 e vendas para outros estados com compras de outros estados, com um grau de correlação de 0,95. Maiores detalhes estão no apêndice B deste trabalho.

### 3.2 MODELOS INDIVIDUAIS

Os resultados empíricos para as equações (2.1), (2.2), (2.10), (2.11) com frequências de três meses e previsão um passo à frente ( $h+1$ ), com, além disso, um regressor individual  $e$ , também, com os testes de *now e forecasting* foram calculados para os horizontes  $h = 1, 2/3, 1/3$  e 0. Para isso é utilizado o Pacote Midassr (Packages 'midassr') no software estatístico R, desenvolvidos por Kvedaras e Balevicius (2021), que permite a estimação, seleção e revisão do modelo para uma regressão MIDAS.

Cada um dos nove regressores é utilizado individualmente para as diferentes equações acima citadas. As Tabelas 04 e 05 apresentam, em sua primeira coluna, o RMSFE de cada um dos regressores; na segunda coluna, temos a razão entre o RMSFE, encontrado em cada regressor, e o RMSFE do benchmark ARMA; já, na terceira coluna, temos os resultados dos valores-p do teste de Diebold-Mariano (DM)<sup>1</sup>. Os resultados das estimações  $\theta_i$  da função parcimoniosa de defasagens (equação 2.4) não são apresentadas no trabalho porque, a princípio, eles não têm nenhuma interpretação econômica, embora possuam grande importância, pois determinam a forma da função parcimoniosa de defasagens dada por  $B(k; \theta)$  e geram desdobramentos em sua interpretação econômica.

É importante destacar que os resultados para todos os modelos individuais foram bastante impactados nos *II, III e IV* trimestres de 2020. Em virtude da crise sanitária causada pela Covid-19, esse impacto negativo nos resultados gerais foi resultado da incapacidade das séries temporais de conseguirem captar com antecedência momentos disruptivos na economia. Nesse sentido, esse fenômeno foi estudado por Bernanke (2018) na Crise do Sub-Prime Americana de 2008. Em seu trabalho, ele testou a capacidade de mais 75 séries temporais, divididas em quatro grandes grupos – ("*housing and mortgage*", "*short-term funding*", "*non-mortgage-credit*", "*Bank solvency*") –, de antecipar o colapso do sistema financeiro. Com isso, Bernanke concluiu que 37 delas tiveram resultados satisfatórios. Neste trabalho, os regressores vendas internas, vendas para outros estados, compras internas e compras para outros estados foram os que apresentaram o melhor desempenho de *now e forecasting* durante o período da pandemia.

De maneira geral, todos os regressores apresentam bons resultados preditivos quando analisados individualmente. Dos 136 RMSFE obtidos por via dos métodos MIDAS, AR-MIDAS,

<sup>1</sup> Os testes DM foram considerados significantes para o nível de 10%

UMIDAS e AR-UMIDAS, com os quatro horizontes estudados, apenas 8 (5,5%) têm desempenho inferior ao Benchmark ARMA. Esses resultados abaixo do benchmark ocorreram somente para os modelos UMIDAS e AR-UMIDAS com os regressores ICMS e taxa. Embora haja essa superioridade em relação ao ARMA, quando adotamos o teste DM para determinar se esses métodos têm a mesma acurácia nas predições, somente em quatro resultados temos a rejeição da hipótese nula de que eles possuem a mesma acurácia preditiva, para um nível de significância de 10%.

Em relação ao MIDAS e ao AR-MIDAS, os dois métodos com melhores resultados, eles não apresentam diferenças significativas de performance. Para  $h = 1$ , temos que 2/3 dos resultados são melhores para o método AR-MIDAS. Já, para o horizonte  $h = 2/3$ , temos que o modelo MIDAS teve menor RMSFE em 55,5% dos resultados. Em  $h = 1/3$ , temos uma melhor performance para o modelo MIDAS. Por fim, no horizonte  $h = 0$ , temos que 55,5% dos menores RMSFE são obtidos no modelo AR-MIDAS. Nos próximos parágrafos, iremos tratar dos regressores individualmente.

Para a variável “vendas internas” temos os melhores desempenhos para as abordagens MIDAS em geral, sendo o AR-MIDAS o de melhor desempenho. O AR-MIDAS também foi o único modelo que apresentou ganhos na predição conforme foram acrescentadas as informações dos meses dos trimestres correntes, uma melhora de 11%. Em relação ao teste DM, não são encontrados nenhum resultado em que se rejeita a hipótese nula, ao nível de significância de 10%, que as séries possuam a mesma acurácia ou desempenho preditivo.

Em relação à variável de vendas para outros estados, tem-se que o melhor desempenho é para o modelo MIDAS em todos os horizontes de tempo, com uma única exceção em  $h=0$  no modelo AR-MIDAS. Aquele, além de ter sido o único modelo que apresenta ganhos conforme são acrescentados meses ao longo dos trimestres correntes, tem seu ganho de predição na ordem de 8,3%. Por fim, os modelos UMIDAS e AR-UMIDAS não apresentam resultados econômicos relevantes. Para o teste de DM, nenhuma das séries apresenta resultados significativos, isto é, não é rejeitada a hipótese nula de que os modelos possuem a mesma performance preditiva.

Sobre as transações com o exterior, tanto exportações quanto importações, os valores encontrados não apresentam grande relevância em termos de resultados, nem tão pouco foi encontrado uma interpretação econômica importante para nenhum dos modelos, ou seja, o núcleo não-linear, função exponencial de almon, não se comportou de maneira lógica, dando um peso expressivo para períodos distantes dos meses atuais analisados (grande peso ao passado distantes ou defasagens), conseqüentemente não houve ganhos de *nowcasting*. O teste DM também não demonstra diferenças entre eles.

Tabela 4 – Um regressor - resultados A

Séries	horizonte	AR-MIDAS			MIDAS		
		RMSFE	Razão Arma	Teste DM	RMSFE (%)	Razão Arma	Teste DM
<b>vendas-i</b>	h = 1	6,610	0,832	0,262	6,950	0,875	0,289
	h = 2/3	6,130	0,772	0,251	6,390	0,805	0,229
	h = 1/3	6,160	0,776	0,258	6,420	0,809	0,240
	h = 0	5,890	0,742	0,173	6,450	0,812	0,245
<b>vendas-o</b>	horizonte	RMSFE (%)	Razão Arma	Teste DM	RMSFE (%)	Razão Arma	Teste DM
	h = 1	6,820	0,859	0,130	6,950	0,875	0,240
	h = 2/3	6,680	0,841	0,127	6,930	0,873	0,239
	h = 1/3	6,760	0,851	0,128	6,890	0,868	0,221
<b>vendas-e</b>	horizonte	RMSFE (%)	Razão Arma	Teste DM	RMSFE (%)	Razão Arma	Teste DM
	h = 1	7,630	0,961	0,328	7,140	0,899	0,270
	h = 2/3	7,540	0,950	0,366	7,262	0,915	0,332
	h = 1/3	7,800	0,982	0,452	7,210	0,908	0,305
<b>compra-i</b>	horizonte	RMSFE (%)	Razão Arma	Teste DM	RMSFE (%)	Razão Arma	Teste DM
	h = 1	7,760	0,977	0,369	7,162	0,902	0,316
	h = 2/3	7,680	0,967	0,266	7,182	0,905	0,310
	h = 1/3	7,649	0,963	0,238	7,181	0,904	0,275
<b>compra-o</b>	horizonte	RMSFE (%)	Razão Arma	Teste DM	RMSFE (%)	Razão Arma	Teste DM
	h = 1	7,050	0,888	0,043*	7,074	0,891	0,269
	h = 2/3	6,640	0,836	0,100*	6,410	0,807	0,113
	h = 1/3	6,590	0,830	0,096*	6,422	0,809	0,120
<b>compra-e</b>	horizonte	RMSFE (%)	Razão Arma	Teste DM	RMSFE (%)	Razão Arma	Teste DM
	h = 1	7,660	0,965	0,380	6,910	0,870	0,192
	h = 2/3	7,070	0,890	0,268	7,140	0,899	0,332
	h = 1/3	7,252	0,913	0,153	6,870	0,865	0,183
<b>va</b>	horizonte	RMSFE (%)	Razão Arma	Teste DM	RMSFE (%)	Razão Arma	Teste DM
	h = 1	7,482	0,942	0,176	7,370	0,928	0,361
	h = 2/3	7,538	0,949	0,203	7,148	0,900	0,237
	h = 1/3	7,540	0,950	0,204	7,123	0,897	0,232
<b>icms</b>	horizonte	RMSFE (%)	Razão Arma	Teste DM	RMSFE (%)	Razão Arma	Teste DM
	h = 1	7,610	0,958	0,349	7,120	0,897	0,258
	h = 2/3	7,710	0,971	0,475	7,260	0,914	0,340
	h = 1/3	6,680	0,841	0,075*	7,280	0,917	0,349
<b>taxa</b>	horizonte	RMSFE (%)	Razão Arma	Teste DM	RMSFE (%)	Razão Arma	Teste DM
	h = 1	7,550	0,951	0,196	7,100	0,894	0,230
	h = 2/3	7,550	0,951	0,200	7,090	0,894	0,210
	h = 1/3	7,540	0,950	0,210	7,120	0,897	0,234
	horizonte	RMSFE (%)	Razão Arma	Teste DM	RMSFE (%)	Razão Arma	Teste DM
	h = 1	7,500	0,945	0,186	7,130	0,897	0,240
	h = 2/3	7,550	0,951	0,200	7,090	0,894	0,210
	h = 1/3	7,540	0,950	0,210	7,120	0,897	0,234

Fonte: Elaboração própria

**Nota:** OS RMFSE foram calculados através dos erros de previsão para 1,2/3, 1/3 e 0 passos a frente, estimando novamente os parâmetros para cada nova observação da variação do PIB trimestral do Rio Grande do Sul. Todos os regressores foram utilizadas em nível e trabalhados individualmente. Está destacado em asterisco os testes de Dibold- Mariano significantes ao nível de 10%

Tabela 5 – Um regressor - resultados B (continuação...)

Séries	horizonte	UMIDAS			AR-UMIDAS		
		RMSFE (%)	Razão Arma	Teste Mariano	RMSFE (%)	Razão Arma	Teste Mariano
vendas-i	h = 1	6,580	0,829	0,214	6,573	0,828	0,211
	h = 2/3	6,550	0,825	0,345	6,600	0,831	0,352
	h = 1/3	6,800	0,856	0,457	6,810	0,858	0,453
	h = 0	6,470	0,815	0,382	6,579	0,829	0,405
vendas-o	horizonte	RMSFE (%)	Razão Arma	Teste Mariano	RMSFE (%)	Razão Arma	Teste Mariano
	h = 1	7,460	0,943	0,625	7,321	0,922	0,512
	h = 2/3	6,430	0,810	0,123	6,570	0,827	0,141
	h = 1/3	6,470	0,815	0,135	6,690	0,843	0,159
vendas-e	horizonte	RMSFE (%)	Razão Arma	Teste Mariano	RMSFE (%)	Razão Arma	Teste Mariano
	h = 1	7,290	0,918	0,383	7,460	0,940	0,529
	h = 2/3	7,320	0,922	0,424	7,476	0,942	0,559
	h = 1/3	7,310	0,921	0,427	7,470	0,941	0,566
compra-i	horizonte	RMSFE (%)	Razão Arma	Teste Mariano	RMSFE (%)	Razão Arma	Teste Mariano
	h = 1	7,620	0,960	0,622	7,840	0,987	0,882
	h = 2/3	7,340	0,924	0,428	7,520	0,947	0,604
	h = 1/3	7,250	0,913	0,342	7,430	0,936	0,513
compra-o	horizonte	RMSFE (%)	Razão Arma	Teste Mariano	RMSFE (%)	Razão Arma	Teste Mariano
	h = 1	7,463	0,940	0,561	7,644	0,963	0,709
	h = 2/3	6,452	0,816	0,197	6,930	0,872	0,312
	h = 1/3	6,810	0,858	0,288	7,250	0,913	0,463
compra-e	horizonte	RMSFE (%)	Razão Arma	Teste Mariano	RMSFE (%)	Razão Arma	Teste Mariano
	h = 1	7,088	0,893	0,327	7,168	0,903	0,378
	h = 2/3	6,940	0,874	0,277	7,003	0,882	0,305
	h = 1/3	6,700	0,844	0,175	6,820	0,886	0,208
va	horizonte	RMSFE (%)	Razão Arma	Teste Mariano	RMSFE (%)	Razão Arma	Teste Mariano
	h = 1	7,270	0,916	0,357	7,473	0,935	0,465
	h = 2/3	7,290	0,918	0,351	7,472	0,941	0,502
	h = 1/3	7,290	0,918	0,364	7,443	0,937	0,485
icms	horizonte	RMSFE (%)	Razão Arma	Teste Mariano	RMSFE (%)	Razão Arma	Teste Mariano
	h = 1	8,750	1,102	0,682	9,084	11,441	0,559
	h = 2/3	7,690	0,914	0,867	8,061	1,015	0,942
	h = 1/3	7,830	0,986	0,942	8,133	1,024	0,905
taxa	horizonte	RMSFE (%)	Razão Arma	Teste Mariano	RMSFE (%)	Razão Arma	Teste Mariano
	h = 1	9,040	1,139	0,539	9,230	1,037	0,472
	h = 2/3	9,040	0,928	0,604	7,141	0,899	0,535
	h = 1/3	7,080	0,892	0,340	7,462	0,940	0,571
	h = 0	7,270	0,916	0,325	7,406	0,933	0,446

Fonte: Elaboração própria

**Nota:** OS RMFSE foram calculados através dos erros de previsão para 1,2/3, 1/3 e 0 passos a frente, estimando novamente os parâmetros para cada nova observação da variação do PIB trimestral do Rio Grande do Sul. Todos os regressores foram utilizadas em nível e trabalhados individualmente. Está destacado em asterisco os testes de Dibold- Mariano significantes ao nível de 10%

O regressor compras internas apresenta resultados relevantes para a interpretação dos modelos. O modelo MIDAS apresenta o menor RMSFE para os horizontes h=1, h=2/3 e h=1/3, com exceção de h=0. Em relação ao *nowcasting*, o acréscimo de meses do trimestre dá ganhos

preditivos de 2,5% no modelo AR-MIDAS, de 6,3% para o UMIDAS e 7,5% para AR-UMIDAS. O teste DM desses modelos não rejeita a hipótese nula de que eles possuem o mesmo desempenho preditivo.

A variável compras de outros estados apresenta os menores RMSFE para os modelos MIDAS e AR-MIDAS. No geral, os modelos UMIDAS e AR-UMIDAS não apresentam nenhum resultado ou interpretação econômica relevante. O modelo AR-MIDAS é o único que apresenta ganhos preditivos de 7,2% conforme são realizados os testes de *nowcasting*, além disso, o teste de DM rejeita a hipótese nula de que os modelos possuem o mesmo desempenho preditivo para os horizontes  $h = 2/3$  e  $h = 0$ .

Para os regressores valor adicionado, ICMS e taxa de arrecadação, temos que o modelo MIDAS e AR-MIDAS tem um desempenho superior ao UMIDAS e AR-UMIDAS. Esses regressores não apresentaram ganhos com *nowcasting* em nenhum período. Além disso, os testes de DM não encontram diferenças significativas no poder preditivo dos modelos em relação ao Benchmark ARMA.

Os resultados empíricos obtidos neste trabalho estão de acordo com os resultados obtidos por Li *et al.* (2022), que aplicou a abordagem MIDAS para um período fora da amostra que incluía a Pandemia do COVID-19, com resultados apenas satisfatórios para o MIDAS e AR-MIDAS. Mesmo sendo impactado negativamente, os modelos MIDAS e a AR-MIDAS aqui estudados têm desempenho superior ao Benchmark ARMA, bem como foram superior à abordagem UMIDAS (UMIDAS e AR-UMIDAS) para todos os horizontes de tempo analisado. Ademais, houve ganhos na predição conforme foram acrescentados os meses dos trimestres correntes. Todos esses resultados são semelhantes aos encontrados nos trabalhos de Galvão e Clement (2008, 2009) e Ghysels (2005, 2004).

Em relação ao *nowcasting*, para os horizontes  $h = 2/3$ ,  $h = 1/3$  e  $h = 0$ , há redução do RMSFE com o modelo AR-MIDAS para as séries vendas internas, compras internas e compras de outros estados. Em se tratando do modelo MIDAS, nota-se que ele apresenta ganhos na predição para as séries vendas internas e vendas para outros estados – esses ganhos, ademais, estão em linha com os resultados de Mariano e Ozmuçur (2020) e Ghysels *et al.* (2021). A única diferença dos resultados aqui encontrados em relação aos resultados da literatura é que o modelo MIDAS tem melhor desempenho do que o modelo AR-MIDAS. Uma explicação econômica para esse resultado seria a de que, com as medidas de restrição impostas pelos governantes devido a covid-19, houve quedas bruscas na atividade econômica seguidas de um trimestre de bom desempenho, e a inclusão do componente autorregressivo no modelo acabou por prejudicar os resultados empíricos.

Em relação ao modelo UMIDAS e AR-UMIDAS, com exceção aos regressores ICMS e Taxa, seus resultados são melhores do que o Benchmark ARMA. Nesse sentido, resultados semelhantes são encontrados em Foroni *et al.* (2011), e, novamente, a inclusão de um componente

autorregressivo ao modelo fez com que houvesse um pior resultado de RMSFE nos testes de previsão realizados para fora da amostra.

### 3.3 MODELOS COM MÚLTIPLOS REGRESSORES

Para a escolha dos modelos com múltiplos regressores, são levadas em consideração as séries que têm bons resultados considerando o menor RMSFE, além das que conseguem obter ganhos de predição ao se acrescentar meses dos trimestres correntes. As séries escolhidas são compras internas, compras para outros estados, vendas internas e vendas para outros estados. Na Tabela 06, é seguida a mesma estrutura das tabelas 04 e 05, diferindo apenas pela a inclusão do Teste DM monocausal de inferioridade.

São analisados sete modelos com múltiplos regressores (MR), que são: o MEAN-AR-M-MIDAS: média simples dos resultados de predição obtidos nos testes individuais para o modelo AR-MIDAS com as séries compras internas, vendas internas, compras de outros estados e vendas de outros estados; M-MIDAS-TRA: midas (MR) sem o componente autorregressivo e uma regressão MIDAS tradicional; M-MIDAS-REC: modelo midas com a regressão recursiva; M-MIDAS-ROLL: modelo com regressão em janelas; AR-M-MIDAS-TRA: midas com o componente autorregressivo e uma regressão midas tradicional; AR-M-MIDAS-REC: modelo midas com regressão recursiva e um componente autorregressivo; AR-M-MIDAS-ROLL: modelo com regressão em janelas e o componente autorregressivo;

Os modelos M-MIDAS-ROLL (MMR) e AR-M-MIDAS-REC (AMMR) obtêm, respectivamente, os menores resultados de RMSFE entre todos os modelos utilizados neste trabalho, levando em consideração tanto os resultados com regressores individuais quanto os resultados com múltiplos regressores. Para o horizonte  $h = 1$ , tem-se que MMR é 42,3% superior ao Benchmark ARMA, enquanto o AMMR foi de 30,9% superior ao ARMA. Para o horizonte  $h = 2/3$ , há um ganho preditivo de 46,5% para o MMR, enquanto, para o AMMR, há um ganho de 32,9% – ambos em relação ao Benchmark ARMA. Em  $h = 1/3$ , para os modelos MMR e AMMR, temos ganhos de 48,2% e 33,1%, respectivamente. Finalmente, em  $h = 0$ , ocorrem ganhos na predição de 50% e 35,4% para MMR e AMMR, respectivamente. Em relação ao Teste DM bicausal para o modelo MMR, com nível de significância de 10%, é rejeitada a hipótese nula de que os modelos possuem a mesma capacidade preditiva. Para o teste monocausal, é rejeitada a hipótese nula de que o Benchmark ARMA não seja inferior ao modelo MMR. Em relação ao modelo AMMR no teste bicausal, nos horizontes  $h = 0$  e  $1$ , não é rejeitada a hipótese nula de que os modelos possuam a mesma capacidade preditiva, enquanto que para  $h = 2/3$  e  $1/3$  essa hipótese nula é rejeitada. No teste monocausal, a hipótese nula de que o Benchmark ARMA não seja inferior (em capacidade preditiva) ao modelo AMMR foi rejeitada.

Os modelos M-MIDAS-REC (MMR) e o modelo AR-M-MIDAS-ROLL (AMMRO) apresentam bons resultados de RMSFE, todos menores que os resultados obtidos com os regres-

sores individualmente, com exceção de  $h = 1/3$  para MMR e  $h = 1/3$  e  $0$  para AMMRO. Esses dois modelos não têm bons resultados com *nowcasting*, obtendo aumento no RMSFE conforme o acréscimo de meses do trimestre corrente. Para o Teste DM bicaudal, não é rejeitada a hipótese nula de que os testes possuem a mesma acurácia preditiva para nenhum dos modelos, enquanto, no Teste DM monocaudal, a hipótese nula de que o Benchmark ARMA não seja inferior aos modelos MMR e AMMRO é rejeitada, com exceção dos horizontes  $h = 1/3$  e  $h = 0$  para o modelo AMMRO, em que a hipótese nula não é rejeitada.

Tabela 6 – Resultados - Modelos com múltiplos regressores

Séries	horizonte	RMSFE (%)	Razão Arma	Teste DM (Two-sided)	Teste DM (less)
MEAN-AR-M-MIDAS	h = 1	6,950	0,872	0,100*	0,052*
	h = 2/3	6,663	0,839	0,134	0,068*
	h = 1/3	6,566	0,827	0,135	0,067*
	h = 0	6,310	0,795	0,124	0,061*
M-MIDAS-TRA	horizonte	RMSFE (%)	Razão Arma	Teste DM (Two-sided)	Teste DM (less)
	h = 0	5,900	0,743	0,158	0,079*
	h = 2/3	6,096	0,768	0,185	0,095*
	h = 1/3	6,085	0,766	0,161	0,080*
M-MIDAS-REC	horizonte	RMSFE (%)	Razão Arma	Teste DM (Two-sided)	Teste DM (less)
	h = 1	5,200	0,655*	0,158	0,079*
	h = 2/3	5,320	0,670*	0,185	0,093*
	h = 1/3	6,060	0,763	0,160	0,080*
M-MIDAS-ROLL	horizonte	RMSFE (%)	Razão Arma	Teste DM (Two-sided)	Teste DM (less)
	h = 1	4,580	0,577*	0,088*	0,040*
	h = 2/3	4,250	0,535*	0,097*	0,046*
	h = 1/3	4,110	0,518*	0,086*	0,043*
AR-M-MIDAS-TRA	horizonte	RMSFE (%)	Razão Arma	Teste DM (Two-sided)	Teste DM (less)
	h = 1	6,480	0,816	0,155	0,077*
	h = 2/3	6,380	0,802	0,100*	0,052*
	h = 1/3	6,660	0,839	0,124	0,062*
AR-M-MIDAS-REC	horizonte	RMSFE (%)	Razão Arma	Teste DM (Two-sided)	Teste DM (less)
	h = 1	5,490	0,691	0,119	0,058*
	h = 2/3	5,330	0,671*	0,100*	0,052*
	h = 1/3	5,310	0,669*	0,079*	0,040*
AR-M-MIDAS-ROLL	horizonte	RMSFE (%)	Razão Arma	Teste DM (Two-sided)	Teste DM (less)
	h = 1	4,090	0,515*	0,110	0,052*
	h = 2/3	4,460	0,562*	0,130	0,065*
	h = 1/3	6,470	0,815	0,592	0,296
h = 0	6,820	0,859	0,677	0,339	

Fonte: Elaboração própria

**Nota:** OS RMFSE foram calculados através dos erros de previsão para 1,2/3, 1/3 e 0 passos a frente, estimando novamente os parâmetros para cada nova observação da variação do PIB trimestral do Rio Grande do Sul. Todos os regressores foram utilizadas em nível e trabalhados individualmente. Está destacado em asterisco os testes de Dibold- Mariano significantes ao nível de 10%

Em relação aos demais modelos, não são apresentados grandes ganhos preditivos quando há comparações aos métodos anteriores. Para o modelo MEAN-AR-M-MIDAS (MAMM), o resultado não é superior aos métodos com regressores individuais, embora tenham havido ganhos com o *nowcasting*. Esses resultados em termos de RMFSE não são significativos, visto que, nos testes de DM bicaudal, apenas para  $h=0$  foi rejeitada a hipótese nula de que os modelos possuem a mesma acurácia preditiva e, no teste DM unicaudal (*less*), para todos os horizontes, é rejeitada a hipótese nula de que o Benchmark ARMA não seja inferior ao MAMM.

Os modelos M-MIDAS-TRA (MMT) e AR-M-MIDAS-TRA (AMMT) também não apresentam grandes ganhos nos resultados de RFSME, nem nos testes de DM bicaudal, com exceção em  $h = 1/3$  para AMMO, em que é rejeitada a hipótese nula de que os modelos possuem a mesma capacidade preditiva. Já para o teste monocaudal, a hipótese nula de que o Benchmark ARMA não seja inferior ao MMO e ao AMMO é rejeitada para todos os horizontes.

Como já era esperado para os resultados empíricos com múltiplos regressores, os modelos com parâmetro  $\beta_i$  variante ao longo do tempo são superiores aos modelos em que esses parâmetros são fixos, por isso os modelos que têm melhor desempenho são a Regressão por Janelas e a Regressão Recursiva, respectivamente, quando comparados à regressão MIDAS tradicional ou à média dos modelos com um regressor. Uma possível interpretação econômica para esses resultados é que tal flexibilidade nos parâmetros dão uma melhor resposta ao momento disruptivo que a economia passou durante a crise sanitária causada pela COVID-19, como consequência das restrições impostas pelos governantes. Ademais, assim como nos modelos com regressores individuais, a inclusão do fator autorregressivo aos modelos tem resultados inferiores em relação aos modelos em que não há esse componente.

### 3.4 COMPARATIVO: MODELO COM MÚLTIPLOS REGRESSORES

O modelo que apresenta o melhor resultado entre todos os modelos é o M-MIDAS-ROLL, razão pela qual o utilizamos neste tópico com o Benchmark para compará-lo aos demais modelos com múltiplos regressores e dimensionar sua superioridade nos testes empíricos. Desse modo, os resultados apresentados na tabela 7 são a razão entre os modelos com múltiplos regressores (MAMM, MMT, MMR, AMMT, ARMMR e AMMRO) e o Benchmark M-MIDAS-ROLL.

Tabela 7 – Comparativo - Modelos com múltiplos regressores

Horizonte	MODELOS com múltiplos regressores - COMPARATIVOS COM M-MIDAS-ROLL					
	MEAN-AR-M-MIDAS	M-MIDAS-TRA	M-MIDAS-REC	AR-M-MIDAS-TRA	AR-M-MIDAS-REC	AR-M-MIDAS-ROLL
$h = 1$	1,511	1,288	1,135	1,415	1,199	0,893
$h = 2/3$	1,568	1,434	1,252	1,499	1,254	1,049
$h = 1/3$	1,598	1,481	1,474	1,620	1,292	1,574
$h = 0$	1,589	1,677	1,373	1,637	1,292	1,718

Fonte: Elaboração própria

**Nota:** Os valores representam a razão entre o RMFSE dos demais modelos com o modelo M-MIDAS-ROLL

Sua superioridade nos resultados empíricos dá-se em todos os horizontes e para todos os modelos, com uma única exceção, a para o modelo AMMRO em relação ao horizonte  $h = 1$ .

### 3.5 MODEL CONFIDENCE SET

Segundo Bernardi e Catania (2015), o fato de dentro de uma mesma metodologia ser possível obter-se mais de um modelo com bons resultados na previsão do PIB (variável mais comum utilizada nos trabalhos) levanta uma questão acerca de qual dos modelos seria o melhor. Em Hansen *et al.* (2011), foram desenvolvidas duas estatísticas testes capazes de um construir um conjunto de modelos denominados "Superior Model Set"(Conjunto de Modelos Superiores) após a eliminação dos modelos de menor acurácia preditiva. Nesses testes, a hipótese nula de mesma acurácia preditiva é rejeitada para um determinado nível de confiança: seus cálculos são obtidos a partir de funções perdas que satisfazem uma condição geral fraca de estacionariedade, ou seja, essencialmente esses modelos podem ser testados sobre vários aspectos. Formalmente, estimam as estatísticas  $d_{i_t}$  e  $d_{i_{j_t}}$  que, a grosso modo, representam a diferença existente entre as funções perdas.

As duas estatísticas geradas para os testes são  $v_m$  e  $v_r$ . A estatística  $v_m$  é obtida a partir da diferença entre a função perda e a média da função perda de todos os modelos. A estatística  $v_r$  é determinada pela média das diferenças relativas entre as funções perdas das previsões dos modelos em análise – para maiores detalhes consultar Hansen *et al.* (2003, 2011), Borges (2012) e Bernardi (2015).

As tabelas 8, 9, 10 e 11 apresentam, respectivamente, o teste MCS para  $h = 1, 2/3, 1/3$  e 0 para os modelos com múltiplos regressores discutidos no tópico anterior. O período em análise tem por base as previsões entre março/2019 e janeiro/2022 (período fora da amostra). Na coluna  $rank_m$  temos o ranking dos melhores modelos preditivos (conforme a estatística  $v_m$ ); na coluna  $v_m$  temos os resultados desta estatística; na coluna  $MCR_M$  temos o p-valor do teste. Em  $rank_r$ , temos o ranking melhores modelos preditivos (conforme a estatística  $v_r$ ); na coluna  $v_r$  temos os resultados da estatística  $v_r$ ; na coluna  $MCR_R$  temos os p-valores do teste realizado; por fim, na última coluna, temos a função perda ( $Loss$ ), obtida a partir dos RMSFE das previsões.

Para um nível de significância de 10%, o modelo M-MIDAS-ROLL apresenta o melhor resultado para as duas estatísticas ( $v_r$  e  $v_m$ ). Por outro lado, em todos os horizontes, o modelo com pior desempenho é o M-MIDAS-TRA, tendo em vista o fato de que ele é eliminado para os horizontes  $h = 1$  e  $2/3$  para ambas estatísticas testes.

Ao verificar-se a partir de somente a estatística  $v_m$ , que é determinada pela diferença entre a observação e a média geral das funções perdas, temos melhores resultados nos modelos de regressão por janela (Rolling regression), com exceção dos horizontes  $h = 0$  e  $1/3$  para o método ar-m-midas-roll. Nesses horizontes, os modelos de regressão recursiva têm um bom desempenho quando comparado ao ar-m-midas-roll; uma interpretação econômica para isso, portanto, deve-se

ao fato da regressão recursiva dar um peso muito maior na entrada mais recente de dados.

Ao analisar-se somente a partir das estatísticas  $v_r$ <sup>2</sup>, a qual é determinada pela média das diferenças relativas entre as funções perdas das previsões, temos melhores resultados no modelos M-MIDAS-ROLL. Os demais resultados não demonstram uma predominância de um modelo em relação aos demais, tendo em vista que cada um dos modelos tem um desempenho diferente em relação à estatística utilizada e a cada um dos horizontes em análise.

De maneira geral, o teste MCS demonstra a superioridade nos modelos de regressão em que os parâmetros estimados variam ao logo do tempo, quando comparados aos modelos que estimam seus parâmetros de forma fixa. Entre os modelos de melhor performance, temos o MIDAS com a regressão por janela em relação à regressão recursiva, e esta se destacou em relação à regressão midas tradicional.

Tabela 8 – Model Confidence set (H=1)

Superior Set of Models							
	rank_m	v_m	MCS_M	Rank_R	v_R	MCS_R	Loss
mean-ar-m-midas	5	1,325	0,378	2	1,541	0,318	6,953
m-midas_rec	3	-0,223	1,000	6	1,862	0,153	5,203
m-midas_roll	1	-1,751	1,000	1	-1,541	1,000	4,587
ar-m-midas-tra	6	1,473	0,266	3	1,652	0,230	6,489
ar-m-midas-rec	4	1,078	0,454	5	1,769	0,169	5,494
ar-m-midas-roll	2	-1,520	1,000	4	1,767	0,169	4,910
<b>p-valor = 0,2464</b>							
<b>Number of eliminated models = 1 (M-MIDAS-TRA)</b>							

Fonte: Elaboração própria

**Nota:**  $rank_m$  = Ranking dos melhores modelos,  $v_m$  = estatística teste do Model Confidence Set  $MCR_M$  = o p-valor desta estatística;  $rank_r$  = Ranking dos melhores modelos,  $v_r$  = estatística teste,  $MCR_R$  = p-valor desta estatística, loss = função perda ou o RMFSE de cada um dos modelos.

Tabela 9 – Model Confidence set (H=2/3)

Superior Set of Models							
	rank_m	v_m	MCS_M	Rank_R	v_R	MCS_R	Loss
mean-ar-m-midas	6	1,360	0,318	4	1,594	0,401	6,663
m-midas_rec	3	0,176	0,965	6	2,226	0,113	5,325
m-midas_roll	1	-1,734	1,000	1	-1,526	1,000	4,250
ar-m-midas-tra	5	1,244	0,432	2	1,526	0,450	6,378
ar-m-midas-rec	4	0,287	0,956	3	1,584	0,410	5,335
ar-m-midas-roll	2	-1,004	1,000	5	1,928	0,174	4,465
<b>p-valor = 0,3178</b>							
<b>Number of eliminated models = 1 (M-MIDAS-TRA)</b>							

Fonte: Elaboração própria

**Nota:**  $rank_m$  = Ranking dos melhores modelos,  $v_m$  = estatística teste do Model Confidence Set  $MCR_M$  = o p-valor desta estatística;  $rank_r$  = Ranking dos melhores modelos,  $v_r$  = estatística teste,  $MCR_R$  = p-valor desta estatística, loss = função perda ou o RMFSE de cada um dos modelos.

<sup>2</sup> com nível de significância de 10%

Tabela 10 – Model Confidence set (H=1/3)

Superior Set of Models							
	rank_m	v_m	MCS_M	Rank_R	v_R	MCS_R	Loss
mean-ar-m-midas	7	0,921	0,511	4	1,720	0,390	6,656
midas_tra	5	0,740	0,721	6	1,839	0,253	6,085
midas_rec	4	0,609	0,815	5	1,796	0,317	6,062
midas_roll	1	-2,153	1,000	1	-1,527	1,000	4,110
ar-m-midas-tra	6	0,811	0,660	2	1,527	0,530	6,664
ar-m-midas-rec	2	-1,613	1,000	3	1,545	0,523	5,310
ar-m-midas-roll	3	0,367	0,946	7	2,039	0,161	6,471
<b>p value = 0,5108</b>							
<b>Number of eliminated models = 0</b>							

Fonte: Elaboração própria

**Nota:**  $rank_m$  = Ranking dos melhores modelos,  $v_m$  = estatística teste do Model Confidence Set  $MCR_M$  = o p-valor desta estatística;  $rank_r$  = Ranking dos melhores modelos,  $v_r$  = estatística teste,  $MCR_R$  = p-valor desta estatística, loss = função perda ou o RMFSE de cada um dos modelos.

Tabela 11 – Model Confidence set (H=0)

Superior Set of Models							
	rank_m	v_m	MCS_M	Rank_R	v_R	MCS_R	Loss
mean-ar-m-midas	5	0,715	0,865	2	1,530	0,410	6,317
midas_tra	6	0,820	0,799	3	1,531	0,410	6,659
midas_rec	3	-0,937	1,000	5	1,693	0,310	5,456
midas_roll	1	-1,864	1,000	1	-1,530	1,000	3,974
ar-m-midas-tra	7	0,991	0,591	4	1,586	0,385	6,510
ar-m-midas-rec	2	-1,201	1,000	7	1,885	0,205	5,053
ar-m-midas-roll	4	0,656	0,879	6	1,742	0,295	6,824
<b>p value = 0,5906</b>							
<b>Number of eliminated models = 0</b>							

Fonte: Elaboração própria

**Nota:**  $rank_m$  = Ranking dos melhores modelos,  $v_m$  = estatística teste do Model Confidence Set  $MCR_M$  = o p-valor desta estatística;  $rank_r$  = Ranking dos melhores modelos,  $v_r$  = estatística teste,  $MCR_R$  = p-valor desta estatística, loss = função perda ou o RMFSE de cada um dos modelos.

Uma interpretação para essa diferença de desempenho entre os modelos, que agrega tanto uma explicação algébrica quanto uma explicação econômica para tal fato, é que os modelos MIDAS-TRA têm parâmetros fixos no tempo, o que trouxe piores resultados em períodos de instabilidade econômica. Por conseguinte, os modelos *MIDAS-Recursive Regression* dão um peso excessivo às novas entradas (ou informações recentes) do modelo, o que se mostrou mais satisfatório em relação ao modelo fixo, dado o período disruptivo da economia que tivemos. Já o modelo *MIDAS-Rolling Regression*, tanto não é fixo no tempo quanto não traz um grande peso à nova entrada de informação que o *MIDAS-Recursive Regression* traz, por isso ele acaba por ser um meio termo entre os dois – sendo mais equilibrado, o que ajuda a explicar a sua superioridade nos testes realizados.

## 4 CONCLUSÃO

Este trabalho considera a abordagem MIDAS desenvolvida por Ghysels *et al.* (2002, 2003, 2005) e complementa este estudo desenvolvendo duas novas abordagens, o *MIDAS Rolling Regression* e o *MIDAS Recursive Regression*, que dão maior flexibilidade à adaptação dos parâmetros aos disruptivos momentos da economia, causados pela crise sanitária mundial devido à COVID-19. A utilização de variáveis econômicas mensais para gerar modelos preditivos de dados trimestrais, como o PIB, tem sido amplamente estudada ao longo dos últimos anos. É, portanto, analisado, neste trabalho, se a abordagem MIDAS pode ser utilizada com sucesso para prever o PIB do estado do Rio Grande do Sul, disposto para composição dos modelos as variáveis mensais da atividade econômica do estado, tais como: vendas internas, vendas para outros estados, exportação, compras internas, compras de outros estados, importação, valor adicionado das cadeias produtivas, arrecadação e a taxa de arrecadação. Ademais, analisa-se se os métodos dão respostas satisfatórias para o período da pandemia de COVID-19.

Para trabalhar com os dados em diferentes frequências (mensais e trimestrais), aplicamos os modelos MIDAS, AR-MIDAS, UMIDAS e AR-UMIDAS com um regressor individual. Para os modelos com múltiplos regressores, adotamos os modelos AR-M-MIDAS-MEAN, MIDAS-TRA, MIDAS-REC, MIDAS-ROLL, AR-MIDAS-TRA, AR-MIDAS-REC, AR-MIDAS-ROLL, sendo os resultados das previsões comparados (por RMFSE) com o Benchmark ARMA. Os modelos de previsão têm um horizonte de 3 meses, além do que foram realizados testes de *nowcasting* para os horizontes de previsão de 2/3, 1/3 e 0 passos à frente, com utilização das informações dentro do trimestre corrente.

Para os diversos regressores analisados individualmente, apenas os regressores vendas internas, vendas para outros estados, compras internas e compras para outros estados apresentam bons resultados para fora da amostra – tanto para previsões de 3 meses quanto para os casos de *nowcasting*. As variáveis importação, exportação, valor adicionado e taxas de arrecadação não alcançam bons resultados, embora tenham sido superiores ao Benchmark ARMA. Esses resultados alcançados nesse trabalho estão de acordo com os resultados obtidos por Clement e Galvão (2008), Guerin (2013) e Marsilli (2014), os quais realizaram previsões do PIB para dos Estados Unidos. Para os modelos com regressores múltiplos, todos têm bons resultados, mas destacam-se os resultados obtidos com os modelos M-MIDAS-ROLL e AR-M-MIDAS-REC, que conseguem, respectivamente, as melhores performances de previsão e *nowcasting* entre todos os modelos aqui analisados. Isso demonstra que, para períodos disruptivos da economia, a adoção de parâmetros flexíveis no tempo apresenta os melhores resultados. Esses resultados, por sua vez, também convergem com o trabalho de Li *et al.* (2022), que adota o modelo de abordagem MIDAS com regressores variantes no tempo, durante o período auge da pandemia, a fim de obter bons resultados para prever a volatilidade do índice S&P 500.

Nossas análises concluem que o modelo MIDAS apresenta bons resultados para modelos de previsão e *nowcasting* do PIB do Rio Grande do Sul, além do que a utilização de informações de dentro do trimestre corrente apresentam um ganho de performance aos resultados. Ademais, a flexibilização dos parâmetros leva a uma melhor solução para se adaptar ao período da pandemia.

Uma sugestão para próximos trabalhos seria replicar os modelos analisados neste estudo para um nível nacional, ou seja, seria a previsão do PIB do Brasil, utilizando tanto séries econômicas quanto financeiras. Além disso, para obtenção de melhores resultados de previsão, partindo da mesma metodologia adotada neste trabalho para os métodos *Midas-rolling regression* e *Midas-recursive regression*, pode-se aplicar as técnicas de *learning machine*: como a utilização de redes neurais (COSTA, 2022) ou a utilização das técnicas de *bagged decision trees*, *random forests* e *stochastic gradient tree boosting* (SOYBILGEN; YAZGAN, 2021), entre outros.

## REFERÊNCIAS

- ALMON, S. The distributed lag between capital appropriations and expenditures. *Econometrica: Journal of Econometrics Society*, **Econometrics Review**, Chicago, v. 33, p. 178–196, 1965.
- ALPTEKIN, A.; BROADSTOCK D. C. AND CHEN, X. W. D. **The theory and practice of econometrics**: Energy economics, Energy Economics. Nova York: Academic Press, 2018.
- ANDREOU, E.; GHYSELS, E.; KOURTELLOS, A. Should macroeconomic forecasters use daily financial data and how? **Journal of Business Economic Statistics**, Ciprus, v. 31, n. 2, p. 240–251, 2013.
- BABII, A.; GHYSELS E. AND STRIAUKAS, J. Machine learning time series regressions with an application to nowcasting. **Journal of Business Economic Statistics**, Chapel Hill, p. 1–23, 2021.
- BARSOUM, F. How useful is financial market data in forecasting uk real gdp. **Working Paper**, Konztanz, 2011.
- BERNARDI, M.; CATANIA, L. The model confidence set package for r. **CEIS Tor Vergata**, Roma, n. 362, 2015. Disponível em: <https://arxiv.org/pdf/1410.8504.pdf>. Acesso em: 21.ago.2022.
- BERNARKE, B. The real effects of the financial crisis. Nova York: The Brooking Institution, 2018. (BPEA Conference Drafts) Disponível em: <https://www.brookings.edu/wp-content/uploads/2018/09/BPEAFall2018-The-real-effects-of-the-financial-crisis.pdf>. Acesso em: 11.jun.2022.
- BORGES, B. K. **Avaliação da habilidade preditiva entre modelos GARCH multivariados?** uma análise baseada no critério model confidence. 2012. Dissertação (Mestrado em Economia) - Universidade Federal do Rio Grande do Sul, Porto Alegre, 2012. Disponível em: <<https://www.lume.ufrgs.br/handle/10183/70011>>. Acesso em: 13.ago.2022.
- BORGES, B. K.; PORTUGAL, M. S. **Previsão macroeconômica para o Brasil utilizando o modelo VAR com dados mistos**. Porto Alegre: UFRGS, 2018. Disponível em: <https://www.ufrgs.br/ppge/wp-content/uploads/2021/06/2018-4.pdf>. Acesso em: 11.abr.2022.
- CASTLE, J. L.; FAWCETT, N. W. P.; HENDRY, D. F. Nowcasting is not just contemporaneous forecasting. **National Institute Economic Review**, Oxford, v. 210, n. 01, p. 71–89, 2009.
- CEPNI, O.; GÜNEY, I. E.; SWANSON, N. R. Nowcasting and forecasting gdp in emerging markets using global financial and macroeconomic diffusion indexes. **International Journal of Forecasting**, Copenhagen, v. 35, p. 555–572, 2019.
- COELHO, A. A. R. **Identificação de Sistemas Dinâmicos Lineares**. Florianópolis: Editora da UFSC, 2004.
- COSTA, K. V. S. Redes neurais artificiais para análise e previsão de fenômenos econômicos: Avaliando o grau de monetização da economia brasileira. Rio de Janeiro: **Instituto de Economia**, 2022. Disponível em: <https://www.anpec.org.br/encontro/2022/submissao/files/I/i8-20cde97d3b66c23871ca786a156c6a53.pdf>. Acesso em: 19.jul.2022.

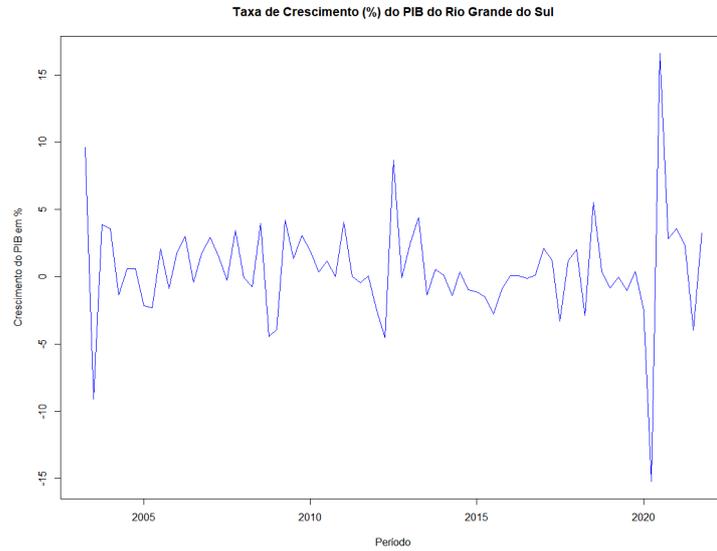
- FERRARA, L.; MARSILLI, C. Financial variables as leading indicators of gdp growth: Evidence from a midas approach during the great recession. **Applied Economics Letters**, Paris, v. 20, n. 3, p. 233–237, 2013.
- FORONI, C.; MARCELLINO, M.; STEVANOVI, D. **Mixed frequency models with MA components**. Rochester: Social Science Research Network, 2018. (SSRN scholarly paper ID 3127429).
- FORONI, C.; MASSIMILIANO, M.; SCHUMACHER, C. U-midas: Midas regressions with unrestricted lag polynomials. **Deutsche Bundesbank - Working Paper**, Berlin, v. 1, n. 35, 2011.
- GALVAO, A. B.; CLEMENTS, M. P. Macroeconomics forecasting with mixed-frequency data: Forecasting output growth in the united stante. **Jornal Business Economics Statistics**, London, v. 26, n. 4, p. 546–554, 2008.
- GALVAO, A. B.; CLEMENTS, M. P. Forecasting us output growth using leading indicators: An appraisal models. **Journal of Applied Econometrics**, London, v. 24, p. 1187–1206, 2009.
- GHYSEL, E.; SANTA-CLARA, P.; VALKANOV, R. The midas touch: mixed data sampling regression. Chapel Hill: UCLA, 2004. (Discussion Paper UCLA and UNC).
- GHYSEL, E.; SANTA-CLARA, P.; VALKANOV, R. Midas regression: Further results and new directions. **Econometrics Review**, Chapel Hill, v. 26, n. 1, p. 53–91, 2007.
- GHYSELS. Mixed data sampling (midas) regression models. **Handbook of Statistics**, Chapel Hill, v. 42, p. 117–153, 2020.
- GHYSELS, E.; SANTA-CLARA, P.; VALKANOV, R. There is a risk-return trade-off after all. **Journal of Financial Economics**, Chapel Hill, v. 76, n. 3, p. 509–548, 2005.
- GROMLING, M. Covid-19 and the growth potential. **Intereconomics**, Colonia, v. 56, n. 1, p. 45–49, 2021.
- GUERIN, P.; MARCELLINO, M. Markov switching midas model. **Journal of Business Economics Estatistics**, Bologna, v. 31, n. 1, p. 45–56, 2013.
- HANSEN, P. R.; LUNDE, A.; NASON, J. M. The model confidence set. **Journal of Economics and Statistics**, Providence, p. 453–497, 2011.
- HANSEN P. R.; LUNDE, A. N. J. M. Choosing the best volatility models: The model confidence set approach. **Oxford Bulletin of Economics and Statistics**, Providence, v. 61, n. 1, p. 839–861, 2003.
- KAGOHARA, D. M. **Avaliando técnicas de nowcasting: uma aplicação do pib brasileiro**. Dissertação (Mestrado em Economia) - Fundação Getúlio Vargas, São Paulo, 2015. Disponível em <https://bibliotecadigital.fgv.br/dspace/handle/handle/10438/13999>. Acesso em: 12.dez.2021.
- KUZIN, V. et al. Midas versus mixed-frequency var: nowcasting gdp in the euro area. **Discussion Paper, Deustsche Budesbank**, Berlin, v. 1, n. 07/2009, 2009.
- KVEDARAS, V.; BALEVICIUS, V. Mixed frequency data sampling regression models: The r package midasr. **Journal of Statistical Software**, Chapel Hill, v. 72, n. 4, p. 1–35, 2021.

- LIAND, X.; LIANG, C.; MA, F. Forecasting stock market volatility with a large number of predictors: New evidence from the ms-midas-lasso model. **Annals of Operations Research**, Chengdo, 2022.
- LUANPING, Z. *et al.*. domestic inflation, exchange rate, and aggregate import demand nexus in nigeria: New evidence from cointegrating regression. **International Journal of Finance & Economics**, Fuzhou, p. 1–12, 2020.
- MARGARIDO, M. A. Aplicacao de testes de raiz unitaria com quebra estrutural em series economicas no brasil na decada de 90. **Informações Econômica**, São Paulo, v. 31, n. 4, 2001.
- MARIANO, R.; OZMUCUR, S. **Predictive Performance of Mixed-Frequency Nowcasting and Forecasting Models**. For Presentation at the 40th PAASE Anniversary – 2020 APMS Annual Scientific Meetings (Online) – US Philippines Parallel Session 4.2 - **Mathematical, Statistical and Data Science**, Preliminary Draf, Philadelphia, 2020.
- MARSILLI, C. **Variable Selection in Predictive MIDAS Models** . Banque de France, Paris, 2014. (Document de Travail n° 520).
- MATTOS, R. S. d. **Tendências e Raízes Unitárias**. Juiz de Fora: Universidade Federal de Juiz de Fora, 2018. (Texto didático). Disponível em <https://www.ufjf.br/rogerio-mattos/files/2018/05/Tendencias-e-Raizes-Unitarias.pdf>. Acesso em: 11.out.2022.
- MORAIS, I. A. C.; PORTUGAL, M. S. Um novo índice coincidente para a atividade industrial do estado do rio grande do sul. **Journal of Econometrics**, v. 37, 2007.
- NAKAJIMA, T. *et al.* capturing changes in asymmetric price transmission : A rolling window tar estimation using bluefin tuna case study. **Journal of International Fisheries**, Tokyo, n. 362, 2011.
- ROCHA, C. D. A. D. **Algoritmo Recursivo dos Mínimos Quadrados para Regressão Linear Local**. Porto: Departamento de Matematica Aplicada , 2001. Disponível em: <https://repositorio-aberto.up.pt/bitstream/10216/9587/3/4802-TM-01-P.pdf>. Acesso em: 29.out.2022.
- RUBIO-HERRERO, Y. W. J. **A Flexible Rolling Regression Framework for Time-Varying SIRD models: Application to covid-19**. Texas: University of North Texas Press, 2021.
- SOYBILGEN BARIS E YAZGAN, E. **Nowcasting US GDP Using Tree-Based Ensemble Models and Dynamic Factors**. Istambul: Spring, 2021.
- TAY, A. S. Financial variables as predictors of real growth. **Research Collection School Of Economics**, Singapura, p. 1–29, 2007. Disponível em: <https://ink.library.smu.edu.sg/cgi/viewcontent.cgi?article=2057context=soe-research>. Acesso em: 11.nov.2021.
- YIH, S.; HWANG, J.-S. A two-phase approach to estimating time-varying parameters in the capital asset pricing model. **Journal of Applied Statistics**, Taiwan, 2009.
- ZIVOT E. E ANDREWS, D. Further evidence on the great crash, the oil price shock, and the unit root hypothesis. **Journal of Business and Economic Statistics**, New Haven, v. 10, 1992.
- ZONGWU, C.; JUHL, T. The distribution of rolling regression estimators. **University of Kansas**, Cidade do Kansas, 2021.

ZUANAZZI, P. T.; ZIEGELMANN, F. A. Previsões para o crescimento do pib brasileiro com séries financeiras e econômicas mensais: uma aplicação de midas. **Revista de Economia Aplicada**, Porto Alegre, v. 18, n. 2, p. 295–318, 2014.

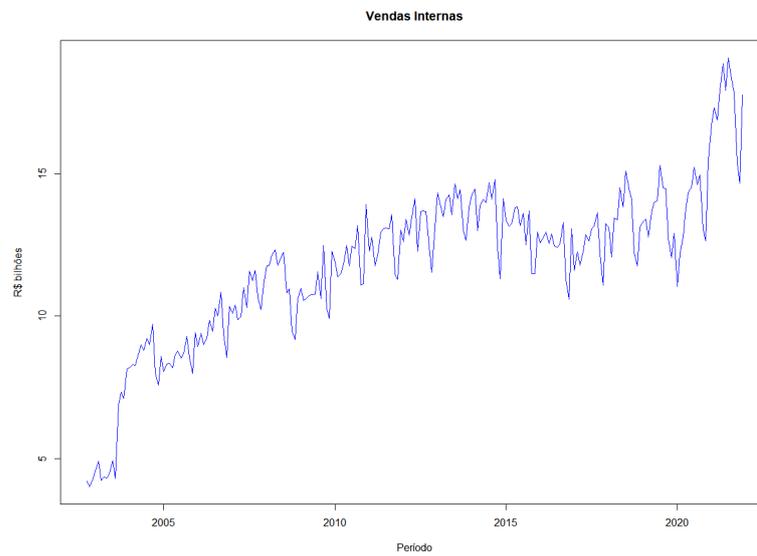
## APÊNDICE A - GRÁFICOS DAS SÉRIES TEMPORAIS

Figura 1 – Taxa de Crescimento do PIB do RS



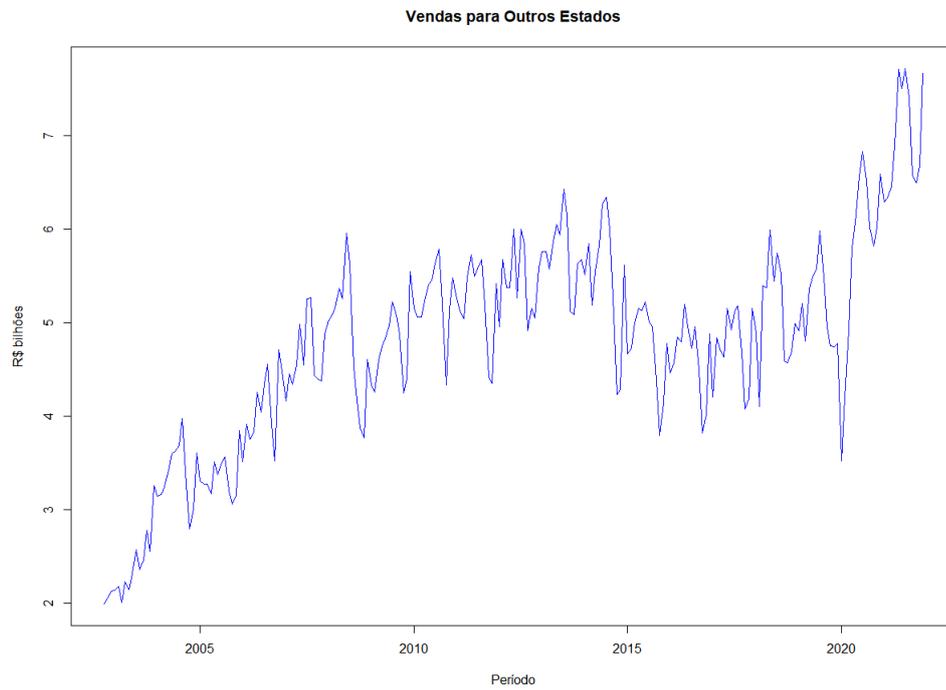
Fonte: Elaboração própria

Figura 2 – Vendas Internas



Fonte: Elaboração própria

Figura 3 – Vendas para outros Estados



Fonte: Elaboração própria

Figura 4 – Exportação

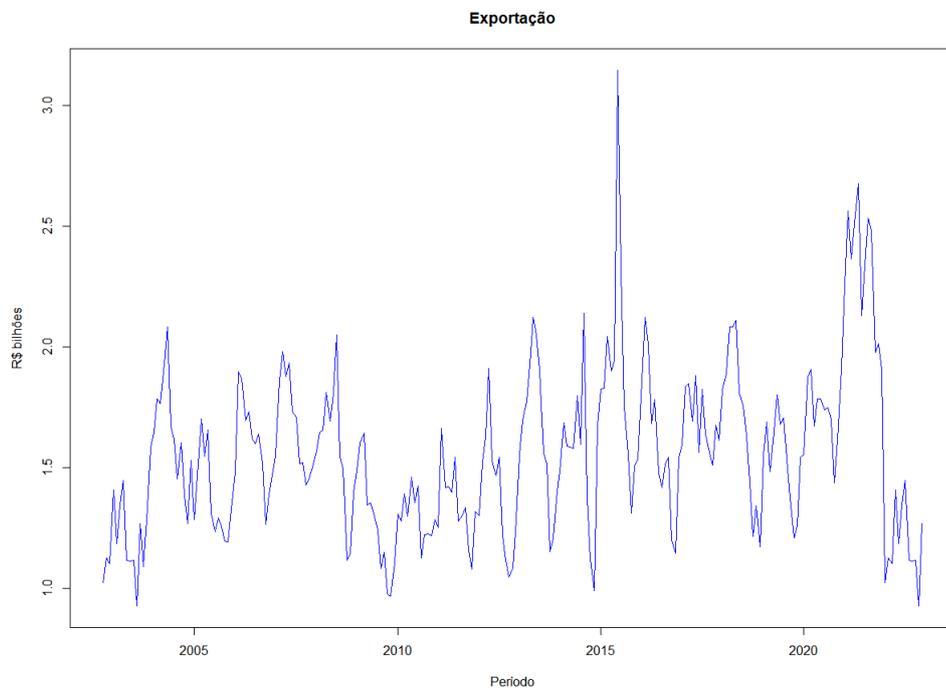
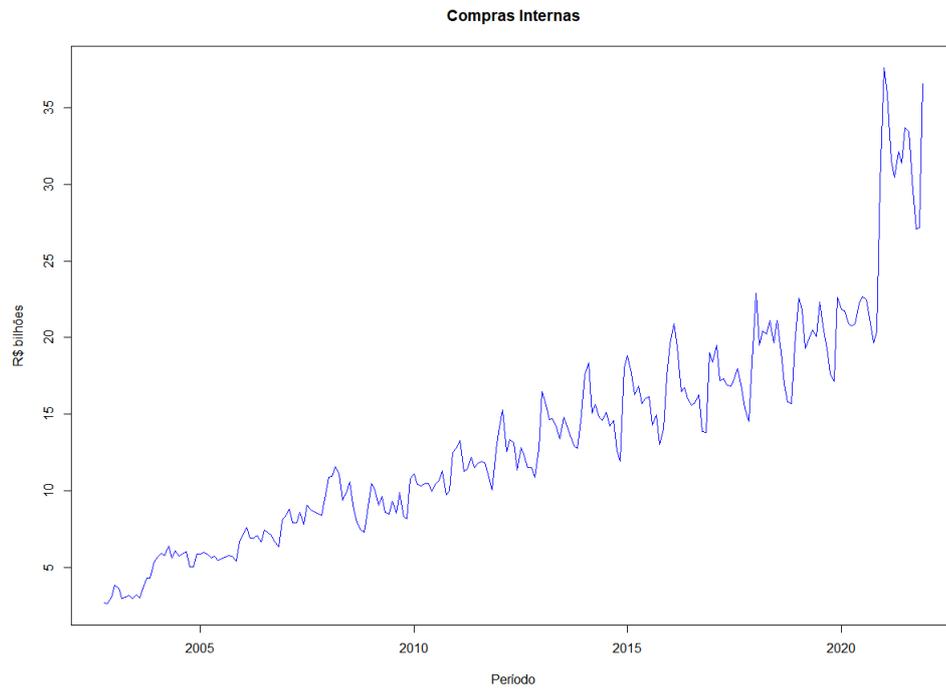
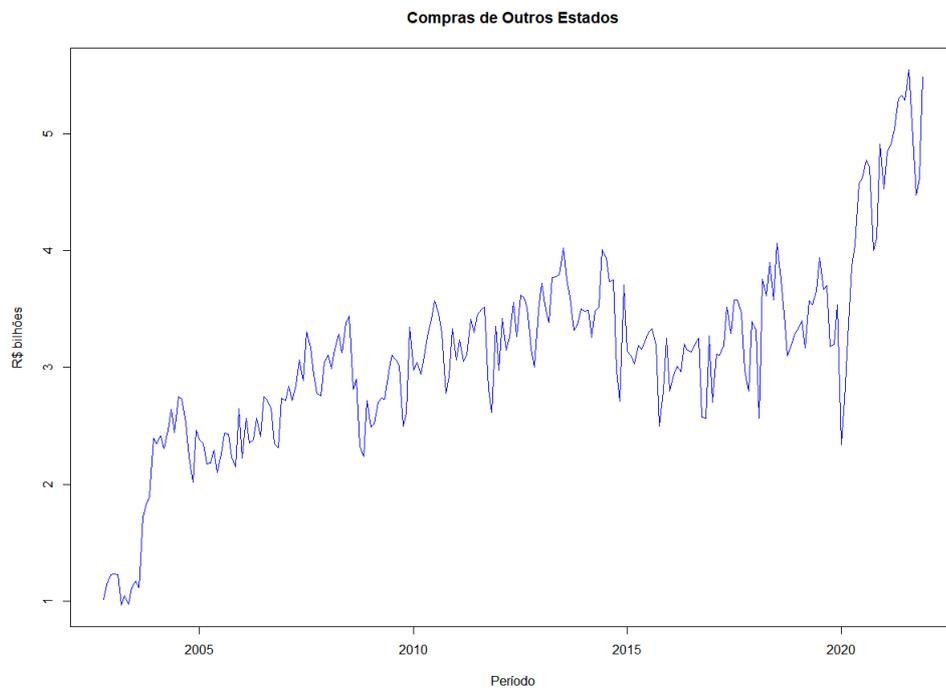


Figura 5 – Compras Internas



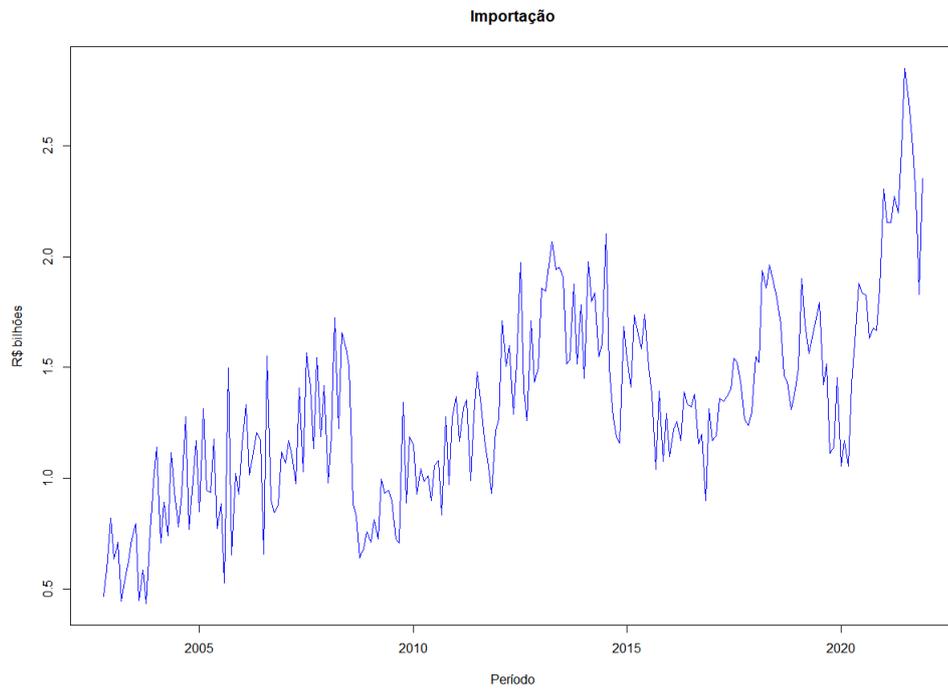
Fonte: Elaboração própria

Figura 6 – Compras de Outros Estados



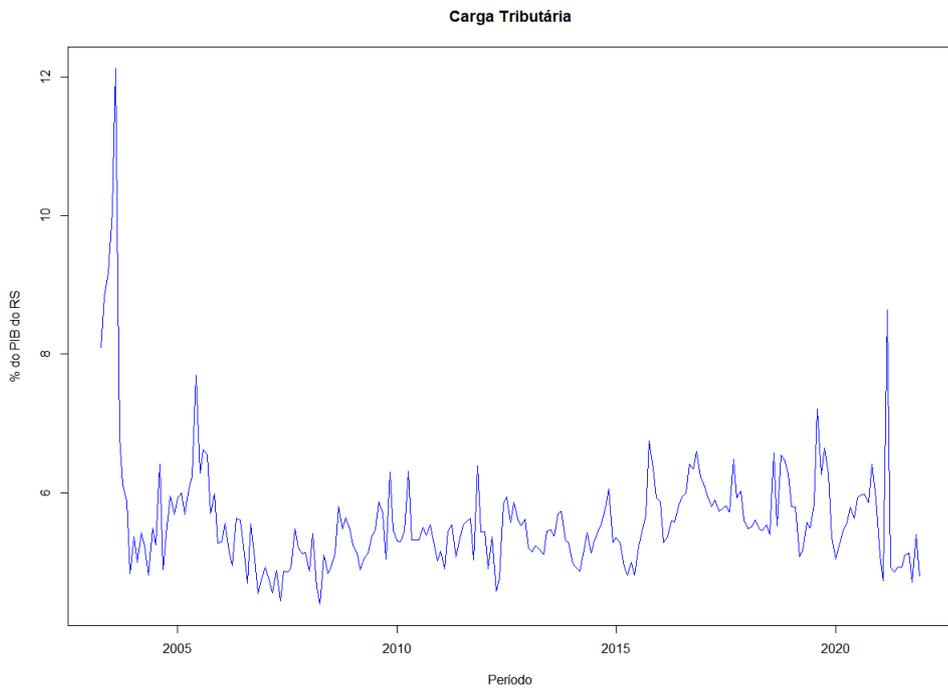
Fonte: Elaboração própria

Figura 7 – Importação



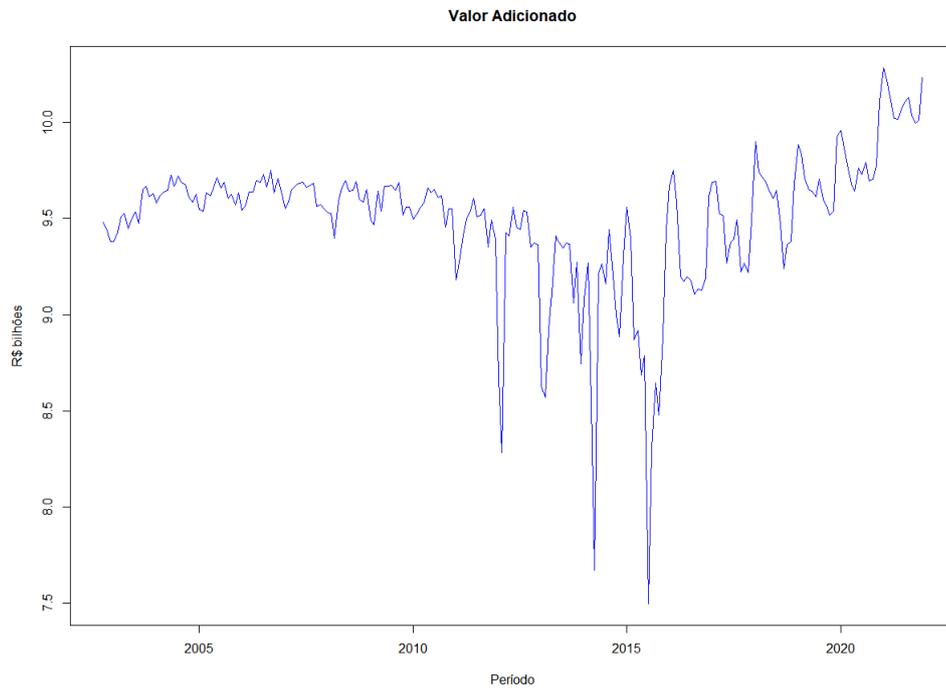
Fonte: Elaboração própria

Figura 8 – Carga Tributária



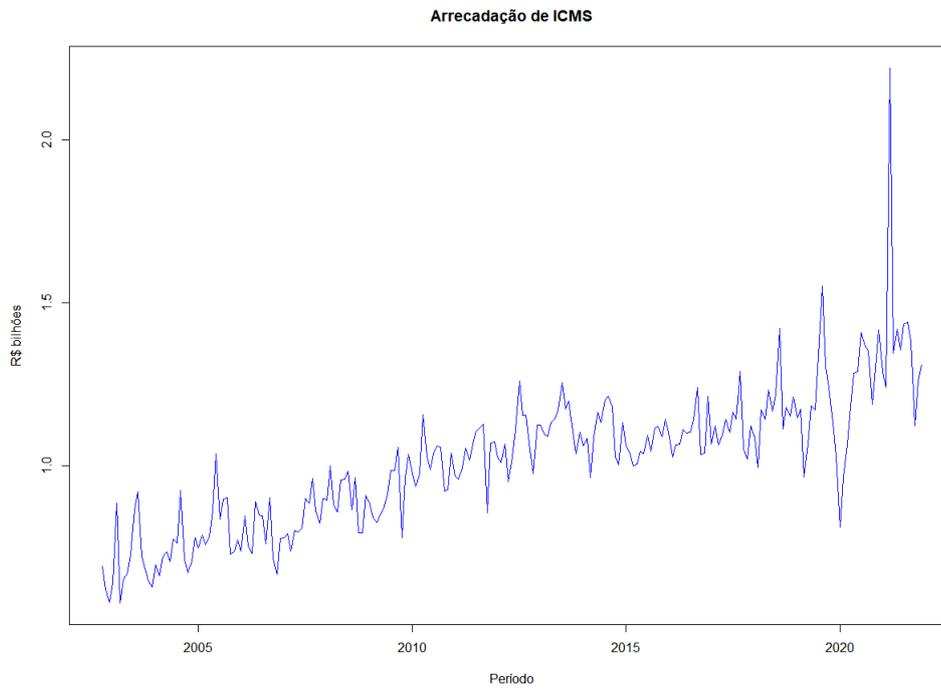
Fonte: Elaboração própria

Figura 9 – Valor Adicionado



Fonte: Elaboração própria

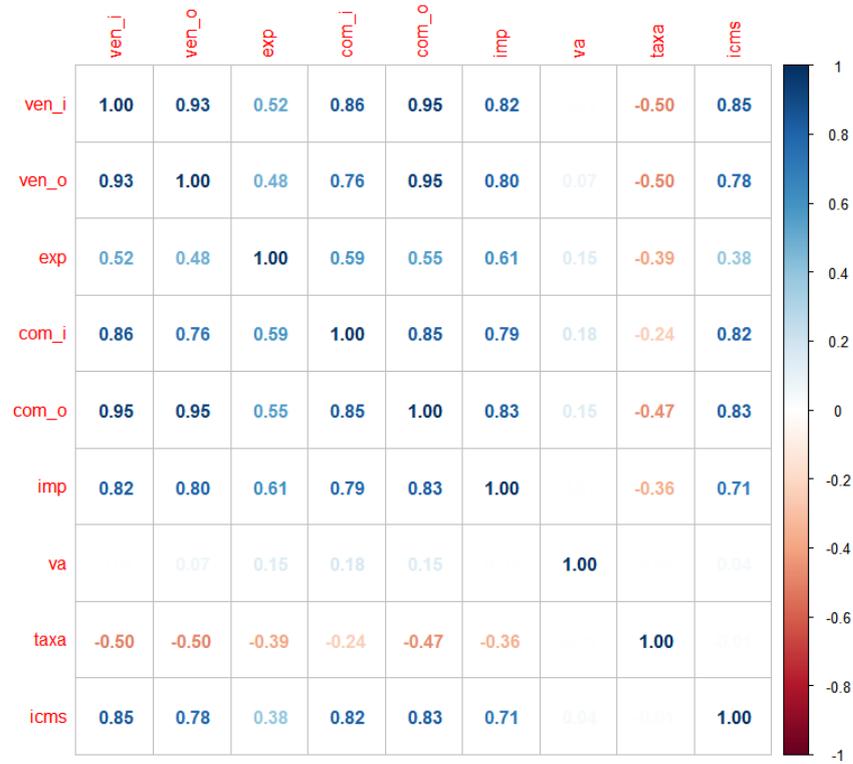
Figura 10 – Arrecadação de ICMS



Fonte: Elaboração própria

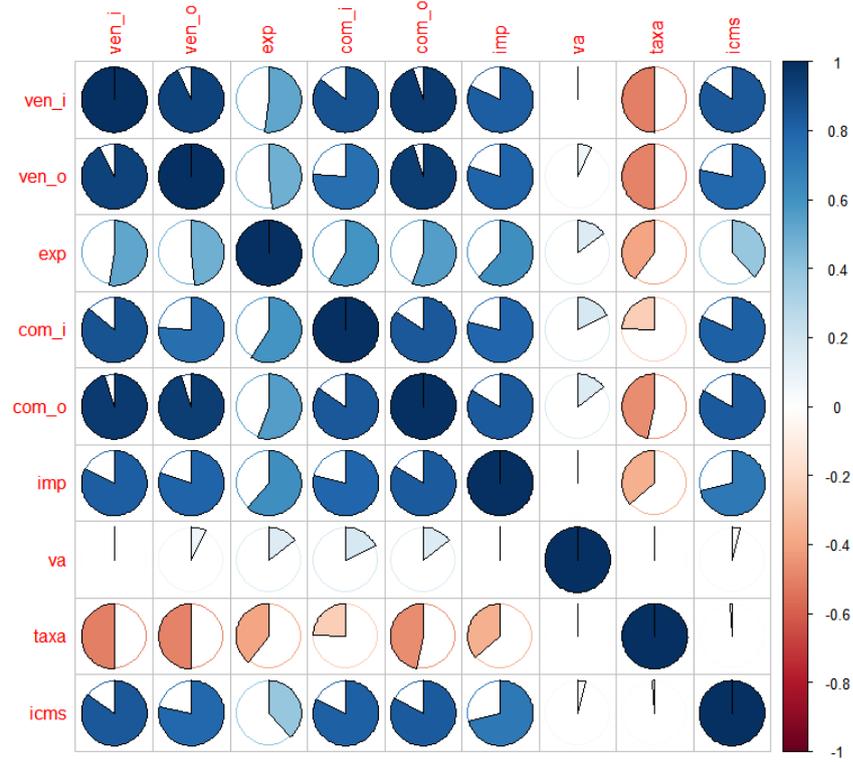
## APÊNDICE B - TABELA E GRÁFICO DA MATRIZ DE CORRELAÇÃO DAS VARIÁVEIS

Figura 11 – Tabela de Correlação



Fonte: Elaboração própria

Figura 12 – Gráfico de Correlação



Fonte: Elaboração própria